Dissertations (2009 -)   Dissertations, Theses, and Professional Projects

Data-Driven Decision-Making: A Case Study of How A School District Uses Data to Inform Reading Instruction

Terese Ann Brecklin
Marquette University

DATA-DRIVEN DECISION-MAKING: A CASE STUDY OF HOW
A SCHOOL DISTRICT USES DATA TO INFORM
READING INSTRUCTION

by

Terese A. Brecklin, B.A., M.A.

A Dissertation submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

Milwaukee, Wisconsin

May 2010

ABSTRACT

DATA-DRIVEN DECISION-MAKING: A CASE STUDY OF HOW
A SCHOOL DISTRICT USES DATA TO INFORM
READING INSTRUCTION

TERESE A. BRECKLIN

MARQUETTE UNIVERSITY, 2010

This case study investigated the organizational structure, assessment philosophy, and instructional practices of one Wisconsin public school district in order to learn about the ways in which the district used assessment data to inform reading instruction. The study was situated within the context of the high-stakes testing environment created by No Child Left Behind legislation.

Data from multiple sources informed the study, including documents, interviews, and classroom observations. Participants included three district administrators, three elementary building principals, three elementary reading specialists and seven elementary classroom teachers. A within and cross-case analysis was conducted to determine the interrelationships between perceptions about assessment and the ways in which this translated into action.

At the time of the study, the district was in the process of transforming itself from a 'district of schools' with strong site autonomy to a 'school district' with uniform expectations for goal-setting, testing, and professional collaboration focused on data. A long range strategic plan established systematic expectations for the top-down change process. These initiatives included the use of test data to inform building goals, the implementation of a new district-wide assessment program, and the implementation of Professional Learning Communities, a staff development structure that fostered collaboration focused on data-driven instruction. These initiatives were implemented within a culture where teachers used data to inform instruction in varying ways. Thus, existing district and classroom assessment practices were taking place at the same time that principals, reading specialists, and classroom teachers were being exposed new ways of using assessment data to inform reading instruction.

A model of Assessment Categories in Public Education accompanies the study. The model was developed to illustrate the ways in which data might be used to inform decisions within a high stakes environment.

ACKNOWLEDGEMENTS

Terese A. Brecklin

TABLE OF CONTENTS

Data Driven Decision-Making:
A Case Study of How a School District Uses Data to Inform Reading Instruction

Chapter 1 - Introduction

We often find ourselves torn between the demands for assessment for
accountability and the need for assessment to improve instruction.
-Winograd, Flores-Duenas, and Arrington (2003)

Assessment for accountability versus assessment to improve instruction. One wonders if the distinction between the two purposes for assessment was on the mind of Congress in 2002. That was the year the members reauthorized the Federal Elementary and Secondary Education Act, commonly known as No Child Left Behind or NCLB. Whether they considered such distinction or not is a moot point. What matters is the reality: NCLB created a high-stakes environment that left educators torn between the demands for assessment for accountability and the need for assessment to improve instruction. NCLB put unprecedented pressure on school districts to meet the demands for accountability as measured by a single yearly assessment. At the same time those knowledgeable about best practices in reading and assessment touted the importance of using data from multiple assessments to inform instruction.

So, which data set holds greater sway? Do data from accountability-driven assessments set a district's course, or, do the results from other assessments guide its work? To put it succinctly, how does a school district use assessment data to inform reading instruction? That is what this case study hopes to uncover.

To set the stage, this chapter begins with an overview of the accountability measures imposed by NCLB. The purpose of this overview is to provide the reader with an understanding of how NCLB mandates might influence the actions of educators within Wisconsin districts. Next, I discuss the perspectives of those who write about best

practices in literacy assessment and their concerns about the kinds of reading

competencies measured by high stakes assessments. Finally, I argue that a problem arises

when educators in a high stakes environment fail to ask probing questions about the ways

in which schools use assessment data to guide reading instruction. This chapter closes

with a rationale for the study and a list of the general guiding questions that shaped the

work.

The NCLB Environment

The bipartisan 2002 reauthorization of the federal Elementary and Secondary

Education Act (ESEA), commonly known as No Child Left Behind (NCLB), calls for

each state to develop an accountability system that describes how the state will take

responsibility for the academic achievement of all students, including subgroups of

students considered most vulnerable to failure. Subgroups include groups of students

defined by race or ethnicity, those eligible for subsided lunch, English language learners,

and students with disabilities. Each state must set achievement standards and develop a

system to measure progress toward meeting those standards. Wisconsin's measure is the

Wisconsin Knowledge and Concept Exam (WKCE). Currently, this assessment is

administered yearly to every public school student in the state. This criterion-referenced

instrument measures proficiency in reading and math for all students in Grades 3 through

8 and in Grade 10. Students in Grades 4, 8 and 10 are also assessed in language arts,

science, social studies and writing.

In response to the results from this yearly assessment, Wisconsin public school

administrators spend considerable time analyzing WKCE data. Inevitably, the data

influence the scope and content of district and school improvement plans—plans which

are designed to increase student achievement on specific concepts or skills measured by the test. When these improvement plans call for program alterations, curricular revisions and/or instructional changes, they directly impact a student's school experiences. Furthermore, improvement plans that track student achievement with additional assessments also affect a student's school experience because these assessments take time away from instruction.

School administrators and school boards take these improvement plans seriously for several reasons. For one, district reputations are either enhanced or damaged by press reports that rank districts based upon the percentage of students who score at proficient and advanced levels. However, public relations concerns are minor compared to the fact that poor WKCE scores can affect a district's operations and its autonomy.

The ultimate goal of NCLB is to ensure that 100 percent of students will demonstrate proficiency in reading and math by 2013-14. The law requires that states measure achievement annually and evaluate that achievement in light of interim achievement goals established by the state. Through this annual process, a state determines if its public school districts and individual public schools have made Adequate Yearly Progress (AYP). Those entities that miss the benchmark face consequences in the form of sanctions that increase in severity the longer the school or district fails to make AYP. These sanctions can ultimately impact a district's operations and its autonomy. For example, schools that do not make AYP for four consecutive years must take Corrective Action and may have to take such actions as replacing relevant staff, adopting a new curricular program, or appointing an outside advisor. Those that miss AYP for five or more consecutive years are subject to restructuring and may have to

face such actions as replacing all relevant staff, reopening as a charter school, or contracting with a private management company.

One can best appreciate a school district's need to continually improve its performance on the WKCE by examining Table 1.1, which shows the interim benchmarks districts must meet in order to demonstrate AYP. Notice how the interim benchmarks for the rate of improvement in reading increase by 6.5 percentage points every three years between 2001-02 and 2009-10. After that, the expectations for the rate of improvement increase dramatically. Beginning in 2010-11, the interim benchmarks increase 6.5 percentage points each year instead of every three years. These accelerated expectations keep the pressure on districts both to perform well and to perform at increasingly higher levels. In examining this table, it is important to remember that the expectations apply to all students, including all the subgroups of at least 40 students. (These groups are defined by race/ethnicity, subsidized lunch, English proficiency and disability.) By considering the potential sanctions of NCLB, along with the inevitable press coverage and increased public expectations, one can appreciate the importance of data generated by the WKCE.

Table 1.1
Expectations for the Percent of Students Scoring at
Proficient or Advanced Levels in Reading as Measured by the
Wisconsin Knowledge and Concepts Examination

Benchmark	Year	Reading
Starting Point	2001-02	61%
	2002-03	61%
	2003-04	61%
Intermediate Goal	2004-05	67.5%
Begin New Test	2005-06	67.5%
	2006-07	67.5%
Intermediate Goal	2007-08	74%
	2008-08	74%
	2009-10	74%
Intermediate Goal*	2010-11	80.5%[a]
Intermediate Goal*	2011-12	87%[a]
Intermediate Goal*	2012-13	93.5%[a]
All Proficient *	2013-14	100%

[a] Notice the rapid rise in expectations beginning in 2010-2011

In summary, Wisconsin has responded to the requirements of NCLB by setting

achievement standards and developing the Wisconsin Knowledge and Concepts

Examination (WKCE) to measure progress toward those standards. In turn, districts

administer the WKCE yearly and respond to WKCE results by developing improvement

plans designed to improve student achievement as measured by the test. The plans may

affect changes in curriculum, instruction, and assessment. Districts are motivated to

develop effective improvement plans because published test scores influence public

perceptions of a district's quality which, in turn, influences enrollment and community

support for the schools. Districts are further motivated to develop effective improvement

plans because those that fail to demonstrate adequate yearly progress as measured by the

WKCE face sanctions which can impact a district's operations and autonomy. In short,

NCLB has created an environment where test results have a strong influence on how

curriculum, instruction, and assessment are comported. A problem arises when a single

test has such widespread influence. This problem becomes more complex when we consider the content of standardized tests in light of beliefs about literacy assessment.

Perspectives on Literacy Assessment

The most common criticism leveled against high-stakes standardized tests is that machine scored tests focus on a narrow range of literate behaviors that are easy to measure (Johnston & Costello, 2005; Afflerbach, 2007a, 2007b; Winograd, Flores-Duenas & Arrington, 2003). These critics want more out of achievement tests than scores that represent grade level success on basic literacy measures such as words lists, tests of subskill knowledge and low level comprehension. Granted, some assessments do examine students' ability to tackle higher level literacy tasks. Wisconsin's WKCE asks students to analyze text, as well as extend and evaluate text but only to the extent these skills can be measured through multiple choice items and the occasional constructed response. Based upon my experiences with these tests, Afflerbach is quite accurate when he asserts that machine scored tests provide little or any information about students' critical reading ability, their ability to evaluate the truthfulness of a statement, detect bias, discern the author's motive, or recognize propaganda (Afflerbach, 2007a). In addition, Johnston (1999) prompts us to expand our thinking about the range of literate behaviors a student might possess. He writes about literacy's social side in which readers engage "in collaborative sense making." As far as I know there are no machine scored assessments that measure one's ability "to use others' literate thinking as a tool to extend both individual and collective thought" (Johnson, 1999, p. 19).

When one considers the higher level thinking embedded in the tasks mentioned above, and when one considers the type of instruction that would foster such a level of

literacy, one can appreciate the concern that high-stakes tests have the power to narrow the classroom curriculum. The argument is that when teachers feel pressured to emphasize tested content over other rigorous content that will not be tested, rigorous content loses. Sadly, when teachers claim they have to prepare students for 'the test' in lieu of other lessons, their comments speak to a concern raised by Amrein and Berliner (2002) who wonder if high stakes testing "really induces teachers to upgrade curricula and instruction or lead students to study harder and better" (p. 21). In such high stakes environment, accountability testing does not serve the purpose NCLB intended it to serve. It does not support better teaching or transform schooling and it does not better the lot of those traditionally underserved. Instead it can serve to reify existing problems and inequalities (Darling-Hammond, 1994; Winograd et al., 2003). The first standard in the Standards for the Assessment of Reading and Writing (2007) eloquently remind us that we need to be aware of the tremendous impact assessments have on students. Standard 1 states:

> Assessment experiences at all levels, whether formative or summative, have consequences for students. Assessments may alter their educational opportunities, increase or decrease their motivation to learn, elicit positive or negative feelings about themselves and others, and influence their understanding of what it means to be literate, educated, or successful. It is not enough for assessment to serve the well-being of students "on average"; we must aim for assessment to serve, not harm, each and every student. (Standard 1, para. 1)

Although school leaders have no choice but to respond to the external pressure of high-stakes tests, they can strive to implement a literacy assessment program that exemplifies best practices, a program "that uses a variety of appropriate indices to address the needs of different audiences" (Winograd et al., p. 208). Appropriate indices in

such a program might also promote the assessment of non-cognitive or affective aspects of reading development, something that Afflerbach (2007b) claims are often neglected. Afflerbach contends that "accomplished readers share important characteristics that are related to but different from their cognitive skills, strategies, and achievement" (p. 156). Consequently, he charges that our assessment agenda is incomplete when we ignore how students develop in terms of motivation, perseverance, and self-esteem because these factors can either be either powerful facilitators or obstacles to a child's reading development. Without this information, he argues that we "limit the inferences we can make about students' reading development, teacher effectiveness, and the value of the reading curriculum" (p. 156).

In summary, group-administered standardized tests, by their nature, measure a limited range of literacy competencies. When data from these tests inform school improvement plans, the plans naturally focus on improving that same limited range of literacy competencies. In turn, those targeted competencies inform curriculum, instruction, and assessment practices—practices that may narrow students' literacy experiences and negatively impact their literacy achievement. A problem arises when, in this high stakes era, we fail to ask about how districts use assessment data to guide reading instruction. That leads to the rationale for this study.

Rationale for the Study

It is easy to become disheartened by the power of high-stakes tests. However, we need not be disheartened if we see the high stakes pressure as a force that prompts us to ask questions about the kinds of data we collect and how we use that data to improve instruction and, as a result, student achievement. Let's leave it to others to debate the

merits of high stakes testing. The challenge for district leaders in this high stakes era is to think more broadly about the concept of data-driven instruction as does McLean (1995) who claims that "the implementation of a complete program of data collection and use can lead to the improvement of education as has no other educational innovation of the last century" (McLean cited in Johnson, 1997, p. 1). In the paragraphs that follow I will cite research that supports a more complete program of data collection and then link that research to the rationale for this study.

Black and Wiliam's (1998) meta-analysis of research from several countries appears to be the contemporary catalyst for discussions surrounding the impact of assessment on student achievement. Their research supports the efforts of those who work to develop complete programs of data collection. The English authors conducted an extensive review of research on classroom practices that included the use of formative assessments as a common feature. The 250 published studies they reviewed encompassed a range of age groups and subject areas. For the purpose of their study Black and Wiliam (1998) defined formative assessment as a practice "encompassing all those activities undertaken by teachers, and/or by their students which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (pp. 7-8).

Black and Wiliam (1998) measured the learning gains across the studies by comparing the average test scores of pupils in formative assessment settings to scores of similar groups of students who were not exposed to such a setting. They found that studies involving formative assessment had an effect size of between .4 and .7. They claim that these effect sizes are larger than those commonly found in studies of

educational interventions. An effect size "expresses the increase or decrease in achievement of the experimental group (the group of students who are exposed to a specific instructional technique) in standard deviation units" (Boston, 2003, para. 7). An effect size of .20 is considered small, an effect size of .50 as medium, and an effect size of .80 as large (Cohen cited in Boston, 2003, para. 10). To put the numbers in another perspective, an effect size of 1 suggests that a particular technique advanced learning by one standard deviation above the mean (Atherton, 2005, para. 5). Marzano discusses how effect size can be interpreted as a change in the percentile ranking of the "average" subject in the experimental group. By way of example, he uses an effect size of .85 and advises that it can be interpreted in the following way:

> …[if]the mean score of subjects in the experimental groups is .85 standard deviations above the mean of subjects in the control group, [then] the average student in the experimental group is at the 80th percentile of the control group, an increase of 30 percentile points. (Marzano, 1998, p.11)

When one considers Marzano's interpretation of effect size into possible percentile point gains, it is easy to appreciate why Black and Wiliam (1998) promote the value of formative assessment in raising the achievement of low pupils in particular and of all pupils in general. Black and Wiliam write:

> While formative assessment can help all pupils, it yields particularly good results with low pupils by concentrating on specific problems with their work and giving them a clear understanding of what is wrong and how to put it right. Pupils can accept and work with such messages, provided they are not clouded by overtones about ability, competition and comparison with others. In summary, the message can be stated as follows: feedback to any pupil should be about the particular qualities of his or her work with advice on what he or she can do to improve, and should avoid comparisons with other pupils. (Black &Wiliam, 1998, p. 142-143).

In light of McLean's (cited in Johnson, 1997) call for a complete program of data collection and Black and Wiliam's (1998) assertions about the value of formative

assessment, a high-stakes testing environment need not be all gloom and doom. In fact, the high-stakes pressure prompts us to ask about the kinds of data we use to guide and improve instruction and the ways in which we use that data. Hence, we need to examine the ways in which both WKCE data and other assessment data inform instruction, specifically reading instruction, in Wisconsin public schools. Through a case study of one district, we might gain insight into how this plays out in practice. To my knowledge, no one has examined reading instruction from that perspective.

WKCE results suggest this is a timely pursuit. That's because the trend data for reading as reported by the Wisconsin Department of Instruction shows relatively flat scores for the period from November 2002 to November 2008. Clearly, student achievement has not skyrocketed as a result of mandated testing. Since it is unlikely that the expectations for public accountability will wane, this dismal trend line prompts us to think about assessment more broadly and ask about ways we might use a full range of assessment data to guide instruction that leads to improved student achievement.

This leads to the broad question guiding this study: How does a school district use data to inform reading instruction? Specifically this study will investigate the organizational structure, assessment philosophy, and instructional practices of a southeastern Wisconsin school district in order to learn about the ways in which the district's administrators, principals, and teachers collect, analyze and use data to inform reading instruction within the current high-stakes climate established by NCLB. It is hoped that the results of this dissertation will be the catalyst for a larger discussion of how schools might use data-driven instruction with the end result being that we become

more thoughtful of how we use data to guide instruction that truly improves reading

achievement.

The research questions guiding this study focus on the assessment of reading

competencies in light of the conceptions about data and assessment that might influence

the ways in which data are used to make decisions. These questions include the

following:

- What reading competencies do districts measure and why?
- Who selects what competencies should be measured, how they should be measured, and the context in which these assessments occur?
- How do conceptions of data and conceptions of assessment shape policies and procedures?
- How do conceptions of data and conceptions of assessment inform programmatic or curricular decisions?
- How do conceptions of data and conceptions of assessment inform instruction at the classroom level?

The purpose of the first two questions is to learn about the range of reading

competencies that might be measured in addition to those stressed in the Wisconsin

Assessment Framework. The framework is a document produced by the Wisconsin

Department of Public Instruction, which provides details about the Wisconsin Standards

that are tested in the WKCE. These particular questions were inspired by the work of

authors who are concerned that standardized tests measure a limited range of literacy

competencies and that, consequently, they narrow the literacy curriculum (Afflerbach,

2007; Amrein & Berliner, 2002; Johnston & Costello, 2005; Winograd, Flores-Duenas, &

Arrington, 2003).

The remaining three questions are designed to explore participants' conceptions

of data and assessment and the ways in which those conceptions appear to influence their

actions.

In Summary

The requirements of No Child Left Behind (NCLB) have led to increased accountability for student achievement with the expectation that all students will achieve proficiency in math and reading by 2013-14. NCLB requires states to measure student proficiency and report the results annually. Public school districts face either publicity problems and/or potential sanctions if incremental proficiency benchmarks have not been achieved. Consequently the data from these tests inform district improvement plans designed to increase student achievement. These improvement plans impact a student's school experience because they generally call for programmatic or curricular changes, instructional changes and additional assessments.

Compounding the issue are the literacy experts who argue that the data from machine scored standardized tests measure a relatively narrow sample of literate behavior. They allege that the current high stakes environment compels teachers to emphasize some elements of the curriculum at the expense of others, a decision that ultimately affects students' learning experiences. With another round of the federal Elementary and Secondary Education Act (ESEA) in the works, and with Race to the Top a regular news feature, it appears that high stakes accountability won't be leaving the political stage any time soon. Consequently, it is worthwhile asking about how we use literacy assessment data to inform reading instruction. We need to examine the ways in which both WKCE data and other assessment data inform instruction, specifically reading instruction, in Wisconsin's public schools. Through a case study of one district, we might gain insight into how this plays out in practice. To my knowledge, no one has examined reading instruction from that perspective.

Data Driven Decision-Making:
A Case Study of How a School District Uses Data to Inform Reading Instruction

Chapter 2 - Review of Literature

The topic of using data to inform reading instruction is a multi-faceted one. Consequently, this chapter addresses the literature in several areas. It begins with a discussion of the terms, data and assessment, and the multiple ways these terms are used in the literature. By delving into the ways in which this terminology is used, we can best appreciate the complexity of data driven decision-making and the range of perspectives that practitioners may bring to the table when they speak about using data to inform instruction. The second part of this chapter focuses on studies that examined the use of data and assessment in various educational contexts. Large scale state studies, district case studies, school level case studies, and studies of effective teachers are included in this discussion.

Multiple Perspectives on Data and Assessments

The descriptions of assessment found in the literature reflect different "philosophical view of reality, knowledge, and learning" as well as philosophical differences over "the levels of student and teacher involvement, the methods used to gather information, the purposes or goals of assessment, and the intended audiences for the results" (Serafini, 2000, p. 208). Serafini argues that these philosophical views affect classroom practice because teachers are influenced by their assessment paradigms when they gather and interpret data about student performance. The same argument could easily be applied at the school or district level; therefore, it makes sense to approach a study of data driven instruction with an overview of the multiple perspectives surrounding the concept of assessment and its purpose. Before examining those multiple

perspectives, however, it is important to define the term, *data,* as used in the phrase, *data-driven instruction*, since every perspective on assessment assumes that some kind of data will be collected as a product of that assessment.

The word data has a precise ring to it. It conjures up images of spreadsheets, bar graphs and certainty. But in the reality of the classroom, data may not find its way to an Excel document, and in fact, it may not even be recorded in a numerical format. Therefore, for the purposes of this study, the term, data, as used in the phrase, data-driven instruction*,* refers to information—either quantitative or qualitative information—that is intentionally captured, recorded, and analyzed at either the individual or group level. For example, norm-referenced reading assessments, informal reading inventories, observational checklists of student behavior, anecdotal notes, or students' self-assessments over time generate data that might inform instruction. Additionally, administrators could capture data describing the number of minutes per day devoted to reading instruction. Or they might tally teacher attributes on an observation checklist of reading instruction. In short, a broad use of the term, data, is essential in this study because any assessment information purposefully gathered by teachers or administrators has the potential to influence the delivery of instruction, and ultimately, the achievement of students.

No matter the type of data one collects, it's one thing to collect data and another thing to use the data to inform instruction. In fact, Mokhtari, Rosemary, and Edwards, (2007) argue that "collecting, organizing, analyzing and using data for instructional and curriculum improvement is a new way of work for many educators" (p. 354). Both Mokhtari et al. (2007) and Fox (2001) contend that educators can begin using data

effectively if they have a framework to guide the work. These authors offer models to help educators conceptualize the data they collect.

Fox's framework promotes a conceptual model of data based on the one used by coaches in California's Reading Success Work. According to Fox, this model distinguishes among three kinds of data—outcome data, demographic data, and process data, each with its own role. When the three are used in combination, the result is "systematic, targeted, and purposeful instruction that results in high levels of student learning" (Fox, 201, p. 11).

Outcome data can be generated from a variety of assessments: state tests, teacher-made tests, observations, and surveys, to name a few. This data is analyzed to learn about individual and group performance and is then reanalyzed in light of demographic data. As Fox sees it, demographic data (i.e., gender, race/ethnicity, socioeconomic status), when used in concert with outcome data, broadens the scope of instructional decision-making because decisions can then be made about the content that is to be learned *and* the students who will be taught that content.

The third category in Fox's (2001) conceptual model is the process data category. He calls process data "change data" because the term refers to the "components, attitudes and practices that comprise the educational program." Furthermore, Fox asserts that process data are the only data that teachers and administrators control. He stresses that outcome and demographic data can be collected and analyzed indefinitely, but that nothing will change until an analysis of process data leads to changes in teacher and/or administrative behavior. Process data are considered after outcome and demographic data are analyzed because process data define factors that might influence the outcome data in

light of the demographics. Process data arise from many sources, including but not limited to curriculum organization, materials purchases, instructional time, instructional strategies formative assessments, and even classroom management.

The Data Analysis Framework for Instructional Decision-Making, designed by Mokhtari et al. (2007) also contains three categories: professional development, classroom data, and reading performance data. The first two categories, professional development data and classroom data are useful because they enrich our understanding of the range of information that might be placed in Fox's (2001)process data category. For example, Mokhtari at al.'s (2007) professional development category includes data from "evaluation or feedback surveys and coaches' logs of how they spend their time and the types of activities they engage in to assist classroom teachers" (p. 355). The classroom data category includes teacher surveys of instructional practices, as well as classroom observational data collected collaboratively by classroom teachers and literacy coaches

Mokhtari et al. (2007) and Fox (2001) help us reflect upon the idea that data serves us in different ways. Data can help us count (outcome data), sort (demographic data), or act (process data). Although one can't argue with Fox's assertion that nothing will happen unless "an analysis of process data leads to changes in teacher and/or administrative behavior" it is unfortunate that neither he nor Mokhtari et al. distinguish between the concepts of summative and formative assessments. Their student data categories – the outcome category for Fox and reading performance category for Mokhtari et al.—lump assessments that have very different purposes into one category. (Fox also lists formative assessment in his process category.) By failing to differentiate among the purpose of assessment both authors miss the connection between the purpose

of the assessment and the data that is generated as a result of it. The following paragraphs examine varying conceptual perspectives of assessment for the purpose of clarifying the fact that different types of assessment have different purposes.

Conceptual Perspectives of Assessment

A number of authors (Black & Wiliam, 1998; Johnston & Costello, 2005; Stiggins, Arter, Chappuis, & Chappuis, 2004; Afflerbach, 2007; Popham, 2008) discuss the concept of assessment. Nuances peculiar to each author aside, the discussion focuses on two main assessment categories: summative assessment and formative assessment.

The term *summative* implies an ending point; that the time for learning is over and it is time to assess the results of that learning. For that reason, Stiggens et al. (2004) refer to summative assessments as Assessments of Learning. Unit tests, semester exams, final performance assessment, and standardized tests all fall under the category of summative assessments. Educators use data from these assessments to judge the quality of a student's achievement or performance in relation to pre-established criteria or in comparison to a peer group. The data serve as the basis for assigning a grade, for certifying proficiency, or for demonstrating accountability. Results are reported to an external audience, be it parents, district stakeholders, or the public in general. Summative data may inform additional instruction for the group of students who took the assessment or, for future groups of students. (This happens with the WKCE because results are available in late spring, too late to impact the current year's instruction.)

In contrast, the data from formative assessments provide evidence about the status of the learning that is in progress, and the degree to which progress is being made toward achieving a pre-identified learning target or learning objective. Because this type of

assessment focuses on the immediate needs of the students, Stiggens (2004) refers to formative assessment as assessment for learning. Popham (2008) emphasizes that formative assessment should be thought of as a multi-step planned process not a test or a particular assessment tool (p.6). Consequently, paper and pencil assessments, observations, rubrics, conversations, can all be part of the formative cycle. As a result of the process, a teacher may adjust a current lesson, decide to refine the next lesson, or decide to plan different experiences for different students. Ideally, the data will prompt action on the part of student, too. Students who receive feedback from a formative assessment (as opposed to a grade that is generated from a summative assessment) can self-assess their progress, adjust their learning tactics and set improvement goals. "Formative assessment is all about decision-making," writes Popham. "It answers the question: Is an adjustment needed, and if so, what should the adjustment be?" (p. 23).

In today's high-stakes environment it is essential to be cognizant of both categories—summative assessment and formative assessment—because each has its own purpose and role in the decision-making process. Data generated from summative assessments will naturally influence decisions in a way that is far different than data emerging from formative assessments, particularly when those decisions have high-stakes consequences. The chart below captures the distinctions between the two types.

Table 2.1 - An Overview of Assessment Categories[a]

	Summative	Formative
Purpose	Appraise student achievement	Promote better learning
Driving Force	Accountability	Improvement of instruction and learning
Timeline	Fixed point in time after completion of learning	On-going process used during learning
Audience	Governmental bodies, school officials, the general public, parents through grades	Teachers, students,
Educators' Roles	Design or select Administer or proctor Interpret results Develop improvement plans Assign grades	Design/select and administer Interpret results Plan next steps in instruction Provide feedback to students Encourage student goal-setting

[a]Based on the work of Stiggens, Arter, Chappuis, and Chappuis (2004); Popham (2008).

Table 2.1 above illustrates that summative and formative assessments have distinct purposes—purposes that are intertwined with other elements in the chart. The purpose meshes with the driving force which in turn influences the timeline for administration, the audience for the results and educator's role in light of the assessment's purpose. In contrast to Fox's model (2001) which uses the type of data (outcome, demographic, process) to drive an improvement process, the purpose of the assessment shapes the kind of action that will take place as a result of the data.

Balanced Assessment

Table 2.1 is arranged so that the physical width of the two assessment categories is equal. This visual arrangement may leave a reader with the impression that the two categories merit equal time in the educational arena, that assessment should be a balance between summative and formative measures. With that in mind, it is interesting to investigate the term *balanced assessment* and the way in which it is discussed in the literature. Balanced assessment generally refers to a

comprehensive approach toward assessment—an approach which utilizes a "rich array of means for assessing students' reading development and achievement" (Afflerbach, 2004, p. 4).

Some who promote the concept of balanced assessment appear to place great value on daily formative-style assessments within the classroom context. For example, Afflerbach (2007) calls for balanced assessment, yet promotes formative assessment by claiming that "formative assessment that informs our efforts to teach students so that they experience success is more valuable than summative assessment that makes after the fact determinations." Furthermore, he touts the importance of student-involvement, asserting that a high stakes environment doesn't foster students' progress toward capable self-assessment, a skill which supports their success as independent readers. Farr (1992) writes about student involvement, too, and gives a nod to the value of formative assessment, contending that tests should be systematic attempts to gather information that helps students understand their own literacy development. In his mind, formative assessment also helps teachers and others in the school setting gather the information needed for curriculum planning. Popham (2008) also advocates for formative assessment. In fact, he labels formative assessment as "a potentially transformative instructional tool that, if clearly understood and adroitly employed, can benefit both educators and their students" (p. 3).

In contrast to the arguments presented by Afflerbach (2004, 2007), Farr (1992), and Popham (2008), Shellard (2003) calls for a balance between on-going informal classroom assessments and more formal standardized measures. Shellard

believes that a carefully planned and comprehensive approach to reading assessment will help us incorporate what we know about effective reading skills into classroom instruction. Running records and authentic assessments that allow a teacher to make instructional decisions and provide feedback on strengths and weaknesses have merit for Shellard, who cites the Learning First Alliance on the matter: "Frequent assessment of developing readers, and the use of information for planning instruction is the most reliable way of preventing children from falling behind and staying behind" (Learning First Alliance, 2000 cited in Shellard, 2003, p. 43). That said, Shellard thinks that data from informal assessments are insufficient for determining the effectiveness of a reading program. She believes standardized measures are useful for comparing data across classrooms and schools, and that the tests prompt us to take steps that address gaps in scores due to race, gender, or socioeconomic level. She also feels standardized results have a place in helping teachers and parents understand how a particular child compares with other students in specific areas.

Shellard's (2003) model appears to capture practices favored by Winograd, Flores-Duenas, and Arrington (2003) who write: "The best practices in literacy assessment, then, are those that use a variety of appropriate indices to address the needs of difference audience. Thus, the choice does not have to be assessment for accountability versus assessment for instruction" (p. 208). In terms of the 'best' indices that might be used, Winograd et al. (2003) bypass the task and name 30 types of assessments that cover the gamut from informal reading inventories to disposition to developmental rubrics. In doing so Winograd et al. caution that looking for the best "in some absolute sense is fraught with perils, because the field of literacy has a

huge array of classroom assessment methods and strategies that can be used either effectively or inappropriately" (p. 213).

Fortunately, that huge array of methods and strategies can be examined in light of a list of 18 principles for using assessment wisely. Winograd et al (2003, pp. 208-09) compiled this list from a number of sources. Below are several from the list several that assist one in thinking deeply about the characteristics of high quality assessments.

The best practices in assessment include the following:

- Involve students in their own learning and enhance their understanding of their own development.
- Start with what students know
- Focus on students strengths rather than just reveal their weaknesses.
- Use criteria and standards that are public, so that students, teacher, parents, and others know what is expected.

Before leaving this discussion of balanced assessment, it is worthwhile to examine the C-A-L-M approach designed by Glazer and Searfoss (Glazer & Brown, 1993). The model developed by these authors lacks the full balance recommended by Shellard (2003), however, the value of the model lies in the fact that C-A-L-M's approach to assessment considers both text and context. The model is based on the "notion that multiple tools, multiple environments, and multiple strategies for collecting and sorting data are essential for describing students' performance." The process (We ac<u>C</u>umulate data, to <u>A</u>ssess <u>L</u>anguage growth over ti<u>M</u>e) embeds assessment in multiple settings: individual, instructional, recreational, intervention, and formal (test-taking) settings. Four components drive the process in these settings. Teachers ask questions about strengths and needs, observe the learner while considering the questions, observe the learner over time, and restate the initial

questions in light of the data collected. The use of assessment results from multiple settings, along with a routine for using the data, appears to create the type of climate needed in order for a balanced assessment philosophy to thrive because it prompts extensive data gathering that will either support or refute high-stakes data.

In comparison to Shellard's (2003) model described above, the Glazer and Brown (1993) C-A-L-M model appears to be focused on the achievement of individual students rather than group achievement or overall program effectiveness. Furthermore, the C-A-L-M model, despite its focus on formative assessment, appears to lack a student involvement component which is a key characteristic of formative assessment. A noteworthy aspect of this model is that it embraces a routine for data collection. The concept of a routine is commendable because it implies systematic attention to the collection and use of data.

The theme that emerges from this discussion of balanced assessment is that the notion of "balanced assessment" can be interpreted in different ways. Those with a passion for formative assessment such as Afflerbach (2004, 2007) and Popham (2008) appear to put formative assessment in the forefront of a balanced assessment program. Glazer and Searfoss (1993) appear to lean in that direction as well. On the other hand, authors like Shellard (2003) and Winograd et al. (2003) appear to approach the topic with an eye toward political reality. They whole-heartedly promote the use of formative assessment but recognize that it must co-exist with high stakes summative measures.

That summative and formative assessment can exist in tandem seems simple enough as long as all parties label and use these assessments for the intended

purpose. However, in a high-stakes environment, we can't turn our backs to the political pressures that exist. Serafini (2000) addresses political pressures that shape how we implement assessments. In fact, he creates an assessment category to address the issue. (His other categories "Assessment as Measurement" and "Assessment as Inquiry", are conceptually similar to the commonly held definitions of summative and formative assessment described above.) Serafini's "Assessment as Procedure" category helps us appreciate how an assessment format might masquerade as formative assessment when, in reality, it is an assignment pressed upon students and teachers. Assessments in this category stem from an external mandate, hence there is limited teacher input and little or no staff development related to the development or interpretation of the assessment results. By way of example, Serafini discusses an instance in one district where portfolios were required. Because the teachers were required to implement portfolios, the portfolio process became a mandatory classroom activity rather than a tool for reflection. In this type of context, "teachers are burdened with another set of procedures given to them by their administration in order to provide scores for an external authority" (Serafini, 2000, p.210).

Serafini (2000) enhances our understanding of how seemingly well-intended assessment practices can backfire in today's political climate of accountability. Those who speak about only two categories fail to mine the nuanced uses of assessment which can exist within a system that is under public scrutiny. In a high-stakes era, it may be helpful to create a hybrid assessment model that captures the assessment challenges we face today. Such a model would honor the concept of balanced assessment, capture the

importance of identifying a purpose for assessment, and acknowledge the fact that political accountability alone will not serve to transform education. The specific elements of such a model will be discussed later in this paper.

Using Data in a School Context

The studies included in this section vary in scope and context from large scale studies conducted at the state level to smaller studies which focused on data use at the district or school level. No matter the study's scope, the use of data to guide instruction often appeared in concert with other factors, factors which were linked to policies or practices at the district or school level. In this review, I will focus on the general contributions of each study or groups of studies and highlight the themes that emerged within a particular group of studies, as well as across studies.

Four studies conducted by teams of researchers under the auspices of the Center for the Study of Reading provide insight into the ways in which a district's decision-making model might influence assessment and instruction. The teams, Stephens et al.,1993a; Shelton et al.,1993; Weinzierl et al, 1993; and Stephens et al.,1993b, examined assessment and instruction in four Illinois districts, which they dubbed Alpha, Beta, Gamma and Delta.

The purpose of the studies was to understand the relationship between assessment and instruction in light of the decision-making process in each district. The authors who conducted these studies defined *instructional decision-making* in a broad sense so that the term included topics like textbook purchasing policies, curriculum development and site management, as well as the monitoring of student progress as measured by both standardized and classroom assessments. What is interesting about their definition is that

topics like textbook purchasing policies, curriculum development and site management exemplify the concept of process data (Fox, 2001) cited earlier. Process data describes the changes made in response to outcome data (standardized test data, for example) and demographics data.

The definition of assessment used by Shelton et al. (1993), Stephens et al. (1993a), Weinzierl et al. (1993), and Stephens et al.(1993b) also merits attention because they categorized assessment types in two ways: "assessment-as-test" (i.e. standardized tests) and "assessment-that-was-not-test" (i.e. teacher observation or informal diagnostic procedures). Although the authors' "assessment-as-test" designation mirrors that of a summative assessment, their use of the term, "assessment-that-was-not-test," can not be fully equated with formative assessment because the critical component of student involvement is missing from the definition. Note that way the authors categorized the assessment was their definition. In the field, they found that participants defined various terms so differently that they were prompted to reexamine their assumptions about school culture. They wrote that they had presumed that there was a homogeneous culture called "school," however, they discovered four different cultures in four different districts. In fact, "the meaning of particular concepts—assessment, curriculum, accountability— varied so significantly across districts that to 'do school' in one district was not the same as to 'do school' in another" (Stephens et al., 1995, p. 9). This finding is not particularly surprising to those who work in schools. The value of the finding is that it was supported by evidence from multiple districts and multiple buildings within those districts.

One theme that emerged as a result of the cross-site analysis conducted by Stephens et al. (1995) is particularly significant. First of all, remember that the purpose of

the study was to understand the relationship between assessment and instruction in light of the decision-making process in each district. In light of this purpose, the authors discovered that "the salient relationship was not the one between assessment and instruction, but rather the relationship of each of these to the decision-making model in the district" (Stephens, et al., 1995, p.19). In other words, standardized tests appeared to influence instruction in districts with top-down decision making structures; conversely, standardized tests had less influence over instruction in districts with collaborative structures or in those that valued teacher autonomy.

It is important to note that these studies were initiated in 1988 and published in 1993. Stephens et al. published the cross-site analysis was published in 1995. That means the fieldwork preceded the 1994 reauthorization of ESEA which contained some federal accountability and testing provisions. The 2002 reauthorization of ESEA— the NCLB legislation which moved accountability and testing to new heights—was far in the future. Nevertheless, the theme of public accountability is still evident when one examines these studies because each district published the results of district-mandated tests. Some of the districts also published their standing on the Illinois state test. As will be seen below, the influence of test scores on the actions of administrators and teachers varied from district to district.

Shelton et al. (1993) examined practices in the Gamma district, a predominantly white (93.1 %) suburban district, with a low income rate of only .5 %, and a mobility rate of 12.4 %. Gamma students achieved consistently high test scores as measured by required formal assessments including the ITBS, the CAT, and the state tests in reading and math. Students also were required to take a middle and end of year basal test. In

terms of accountability, the results of the district-mandated assessments were reported to the community. Gamma's building principals were responsible for interpreting these district test results and for developing strategies to address any dips in performance. Test results were used to evaluate and modify the curriculum. No information was provided about the strategies the principals developed or whether they felt pressure to demonstrate continuous improvement in test scores.

Despite the building principals' responsibility, the term, assessment, was most commonly used in Gamma to refer to the information gathered by teachers. According to the authors, the principals valued the teachers' formal or informal assessments and considered them "more valuable than district-required assessments" (Shelton et al., 1993, p. 25). That's because the culture in the district supported teacher autonomy. It was one that respected the expertise of classroom teachers who were "considered to be the most knowledgeable individuals regarding what assessment measures should be used and the degree to which they should be valued and communicated" (p. 25).

Classroom assessments in Gamma exemplified those that appear to fit in the category of formative assessment. The assessments included: portfolios, work samples, student self-assessment, teacher judgment, informal notes, student conferences, observation, class participation, anecdotal records, and checklists. Assessments of the summative type included teacher-made tests, basal tests, and other published tests. Other than the summative type of district-mandated comprehension and skills tests from the basal series, it appears that teachers based their assessments on professional judgment rather than a set of grade level standards or benchmarks. They were free to assess and to

use assessment results as they saw fit. The authors do not report any teacher collaboration focused on data analysis.

Some assessments used in Gamma were matched to students needs. Assessment "did not always mean a group test in which individual were graded on a scale in relation to the performance of others" (Shelton et al., 1993, p. 29).Teachers might select various items for individual students that reflected the student's particular learning needs. Feedback, which is one of the hallmarks of sound formative assessment, was also part of the assessment process. Students received sticky notes on their writing that told "Things I Need to Work On" and "Things I Do Well" (p. 29). Furthermore, self-assessment was woven into cooperative learning activities. Students assessed themselves along a continuum to gauge both their social and academic progress. We do not know if students used their self-assessments to set goals for improvement.

Informal assessments were valued. One particular teacher walked around her class to listen to her students read. She reported that she "often took that information as a sort of inner record of students strengths and weaknesses, adding [that she] probably did more things inwardly than outwardly" (p. 6).

Surprisingly, despite the varied assessments reported by teachers, their primary instructional approach centered on basal materials and whole group instruction followed by cooperative group activities. One team even used ability grouping. Students who exhibited specific needs were handled one-on-one. This instruction occurred during a 20-minute Sustained Silent Reading time or in the Learning Resource Center (LRC). Parent volunteers who were supervised by a learning specialist provided some instruction in the LRC. We know that teachers sent students to the LRC with a prescribed program,

however, it is unclear how assessments informed a student's prescription or if assessments were used to determine if the student made progress in the LRC.

The Gamma district's initiatives during the year of the study included grade level teaming, thinking skills, and cooperative learning—typical initiatives in the early 1990s. "Hands-on" workshops provided teachers with practical way to implement these initiatives. It is disappointing that the study does not speak about staff development that might have prepared teachers to use the type of formative assessment tools listed above: portfolios, work samples, student self-assessment, observation, class participation, anecdotal records, or checklists.

In contrast to Gamma's teachers who appeared to enjoy considerable autonomy, the teachers in the Beta school district (Weinzierl et al. 1993) practiced their craft in an environment that valued standardized test scores and a textbook-driven curriculum. The Beta School District was located in city of about 59,000 people and was home to a "large university, a community college, a few manufacturers and a number of service industries" (p. 7). The student body was 65.7 % white, 30.5 % black, 8.5 % Hispanic, and 2.4 % Asian. It had a 23.3 % low income rate compared to Gamma's 0.5 percent and a mobility rate of 26.2 percent compared to Gamma's 12.4 %. According to the Weinzierl et al., Beta administrators, principals and teachers perceived standardized test data as valuable. Furthermore, all stakeholders were an audience for the results. Data from the SRA and the state math and reading assessment were reported to the public on the state mandated School Report Card. This Report Card was sent to every home in the district.

When Weinzierl et al. (1993) sought to understand literacy instruction in the context of assessment and decision making across Beta, they discovered that the

"curriculum was synonymous with the textbooks" and teachers made sure they covered

the district-selected textbook in order to prepare students for the tests (p. 1) Furthermore,

test data guided instruction because teachers modified the curriculum to address topics in

which students received low scores. In other words, data was linked directly to test

content which, in turn, guided instruction. Unlike the Gamma educators who placed a

higher value on classroom assessments, Beta educators believed that test data "were the

most viable way to understand how students were doing." The educators also "seemed to

feel that the tests measured what mattered" (Weinzierl et al.,1993, p. 1). Despite the

emphasis placed on test scores, the authors report one Beta teacher who used informal

assessments to pinpoint a need for skills instruction not covered in the basal—quite a

contrast to another teacher who used questions about the basal stories to move students to

the next reader or a different reading group.

The student population in the Delta district differed from both Gamma and Beta.

White students comprised 82 % of the student body. Hispanics represented 12.3 %,

Asians, 5 % and blacks .5 %. The student mobility rate of 22.3 % was slightly lower than

Beta's 26.2 rate; however, only 7.4 % of Delta's students were reported as low income,

compared to Beta's 23.2 % of low income students.

Compared to Beta, the instructional environment in Delta (Stephens et al., 1993b)

was even more prescribed. That is not surprising because Delta's superintendent wanted

low test scores to rise and had instituted several practices linked to that goal. Teachers

were given an item analysis based upon standardized test data and were expected to

"modify their teaching so that scores on the test would improve" (Weinzierl et al., 1993,

p. 19). In addition to the emphasis on test scores, textbooks were also synonymous with

curriculum in Delta—so much so that Delta administrators collected basal end-of-unit test scores in math and reading each week. Administrators used the data to monitor the pace of instruction across the district with the expectation that all teachers would cover material at the same pace. This set the tone for instruction in the district so that "instruction was predominantly but not exclusively, a matter of covering the materials in the district adopted books and ensuring that students had passed mastery tests at a level of 80% accuracy before progressing to the next book in the series. Responsibility was predominantly, but not exclusively, to the tests and to the material [in the district adopted books]" (p. 19).

In contrast to Delta's explicit expectations of teachers, Alpha had the reputation of allowing for considerable teacher autonomy. This factor made Alpha unique in comparison to other districts the researchers involved in this series of studies had known or read about (Stephens et al, 1993a). With 71.3 % white students, 20.3 % black students and 7.0 % Asian students, Alpha had the highest percentage of low income students (32.2 %). Its student mobility rate of 23.6 % was similar to Beta's and Delta's mobility rate.

Alpha's teachers were free from allegiance to specific textbooks. Given a basic curriculum guide, the teachers determined lesson design, materials, and grouping patterns. Apparently, the guidelines were so broad that they did not even influence a teacher's expectations for the quality of work much less the grade assigned to it. Decisions were based on the perceived needs of the individual student to the point where one teacher "would accept a child's best effort whether or not it represented what otherwise might be considered 'good' work" (Stephens et al, 1993a). As one might expect, standardized tests had little impact on instruction in Alpha. In fact, the researchers

note that teachers' opinions were valued more than data from the standardized tests. This observation exemplifies the Alpha district culture, a culture in which teachers had control and autonomy because they were viewed as professionals and decision-makers. What we don't know is how teachers made their decisions.

For example, we don't know why a teacher assigned her first/second grade multiage students to one of three homogeneous groups and why she chose specific activities for each group. It would be interesting to know why one group was assigned two worksheets focused on the /s/ sound while the other groups read from basal texts. There is no reference to data that might have informed these decisions, nor any evidence that the teacher's tacit knowledge informed the decision.

One can infer how assessment data might be used to guide instruction from the description of multi-age third and fourth grade class arrangement that included weekly one-on-one reading and writing conferences, which likely served as a form of informal assessment given the way the teacher prepared for the conference. Before the conference or "conversation," the teacher reviewed the student's book conference notebook which contained summaries written by the student. The teacher then wrote questions for the student to answer and expected the student to have prepared a passage to read aloud. The conference itself focused on individual student needs (for example, summarization or characterization), however, we don't know if the data collected as a result of these conversations guided specific changes in instruction at the individual, small group or whole group level, or, if the teacher tracked student improvement related to the topics discussed at the conference.

Similar one-on-one conferences were conducted in a multiage 5th and 6th grade classroom along with teacher-led daily literature group discussions. Students chose their literature group book from among three or four pre-selected by the teacher, however, again we don't know what influenced the teacher's selection of the books. Because this teacher had a considerable amount of interaction with students, one could assume she was influenced by her analysis of the students' conference notebooks, the worksheets she graded, the content of the individual conferences or small group discussion. Yet, we don't know for certain because we don't know how the teacher monitored progress in the classroom. Without that answer, one wonders if instruction in Alpha is activity driven rather than data driven.

In summary, the case studies of Alpha, Beta, Delta and Gamma provide insight into the relationship between assessment and instruction in the context of a district's decision making model. (As noted earlier, the studies suggest that standardized tests appeared to influence instruction in districts with top-down decision making structures; conversely, standardized tests had less influence over instruction in districts with collaborative structures or in those that valued teacher autonomy.) That said, one leaves this series of studies craving more insight into the content of the formal and informal assessments used by the teachers, how (or if) they analyzed and used the data to inform instruction how they monitored students' progress, and how their practices of data usage compared to other teachers.

Improving Status on External Measures

The studies discussed in this section focus on schools or districts that improved their standing on state accountability measures. These studies are discussed together because it was their improved performance that made them the objects of investigation. The overarching goal of each of these studies was to understand how these schools improved their performance and to identify strategies that might be useful to other schools. The authors hoped to discern patterns or procedures that might have contributed to achievement gains.

Data-driven instruction was one of the themes that emerged in a study funded by the Ohio Department of Education. In that study, Kercheval and Newbill (2002) examined practices in fifty Ohio school districts that had improved their rating on Ohio's Local Report Card, an annual report of students' performance on standardized statewide tests which documented rates of improvement on those tests, student attendance and graduation rates. The authors linked improvement to "key effective practices," practices which they described as consistent with those linked to effective schools research: curriculum alignment, professional development, an emphasis on literacy, data analysis and tracking, intervention and remediation strategies and test preparation strategies. Improvements were attributed to the fact that the schools used the key effective practices together in such a way that the practices were dubbed as "linked and mutually reinforcing" (p. 4).

Although the practices cited are described as mutually reinforcing, it is important to remember that data from externally imposed assessments drove the process. This external influence is a key factor because the externally imposed assessments shaped the

curriculum in the improved schools. One could also argue that the external assessments influenced the organizational structure because all districts in the study had individuals assigned to track progress toward state, district and school performances goals—goals which were apparently driven by strand item analysis of the state proficiency test. Perhaps then, it is no wonder that schools and classroom teachers made frequent use of assessments to track progress toward these competency goals and that "every school had intervention practices in place to monitor and identify students needing extra help to pass the proficiency tests" (Kercheval & Newbill, 2002, p. 22). Unfortunately, the Ohio study is silent on whether the schools used data to inform instruction that was not directly related to the state competency test. One wonders if the improved schools used other data—perhaps data generated at the classroom level—to plan instruction to meet individual needs and to foster reading competence in areas other than the tested goals.

A North Carolina Department of Education (Cobb, 1997) study examined "typical high poverty schools that were showing good and steady growth across achievement levels" over time (p. 5). Investigators hoped to discover what these schools did to enhance student learning—information that might help other high poverty schools improve student achievement. Again external accountability is a key factor in this study. The North Carolina study was driven by a new state accountability model which measured success on cohort growth from year to year and placed "new demands on individual schools to improve student performance on state End-of Grade (EOG) Tests 3-8" (p. 1). Despite the author's disclaimer that tests are not the only way to measure successful learning, and that test are one indicator of student learning and a key ingredient for school accountability, the North Carolina model exemplifies the pressure

of high-stakes testing, with money being offered to schools that improved and technical

assistance being offered to those that fell short of the mark.

Common characteristics among the improved schools in North Carolina are

similar to some of the characteristics identified in the Ohio study cited above (Kerchval

and Newbill, 2002). These characteristics include the use of assessment results to inform

practice, a focus on a standard course of study, and the use of intervention for struggling

students. The North Carolina study (Cobb, 1997) also identified universal high

expectations, an atmosphere of care and respect, a safe and orderly environment, teacher-

driven reform, principal leadership and support, didactic and hands-on instruction, and a

willingness to experiment as other characteristics of the improved schools.

Cobb (1997) also reported various forms of assessment were used in the North

Carolina schools. These included "commercial and locally developed Informal Reading

Inventories and computerized test item banks in addition to state-developed diagnostic

materials" (p. 11). Although we know that "teachers assessed students along the way, not

just at the end of the year or prior to the EOG [End of Grade] Testing," no information is

provided detailing the ways data from various assessments were balanced with EOG data

to drive instruction on a daily basis. However, the report suggests that data from the EOG

test carried substantial weight. Like the Ohio schools, the North Carolina schools used

state test data to develop intervention plans. Staff studied the EOG Test results, used

testlets and other strategies, and linked their classroom work to the test achievement

levels. Although staff development was provided to help teachers understand and work

with state standards, we do not know if the focus was on all state standards or only some

of those standards—the ones measured on the test. Furthermore, no mention was made as

to whether staff development specifically focused on ways to use data to guide instruction. This is of interest in light of the fact that researchers reported that some teachers "may have over-interpreted results, or used them in ways that stretched the reliability for individual students" (Cobb, 1997, p. 11), however they do not provide concrete examples of ways in which this happened. One wonders about the degree to which such practices may have invalidated the gains these schools made on the state test.

It appears that North Carolina data influenced classroom instruction in ways not directly related to test preparation. The authors reported that teachers in the improving schools reacted to their achievement results by also experimenting with strategies such as portfolio assessment, double reading periods, and common planning time. Most teachers indicated that "they had moved to more manipulative in mathematics and used a combination of phonics with teaching reading in context and whole language" (Cobb, 1997, p. 12). These changes in practice exemplify the concept of process data (Fox, 2001). Process data describes the changes made in response to outcome data (i.e., standardized tests, formal and informal classroom assessments, etc.) and demographic data.

A study by the Charles A. Dana Center (2001) examined the characteristics of five high-poverty Texas high schools that demonstrated marked improvement on the high stakes Texas Assessment of Academic Skills (TAAS). Students in these schools also performed better than the state average on at least one of 3 other academic indicators: the Texas Learning Index score in reading and math, the Texas Algebra 1 End-of-Course Examination, and/or enrollment and course offerings in Advanced Placement Programs. As with the studies cited above, the goal of this study was to understand how these

schools improved their performance and to identify strategies that might be useful to other schools.

Dana Center Researchers, working in teams of three, made two-day visits to each of the schools. In addition to conducting individual and focus group interviews with school personnel, parents, and students, they observed classes and other activities, and examined various documents in order to discern the "ideas and practices that seemed to support student achievement" (Dana Center, 2001, p. 2). The practices they found included: setting clear goals and establishing high expectations; using data to guide instruction; focusing on instruction and individual learning; supporting teachers and enhancing collaboration; and fostering an environment of respect and affection for students. The practices of goal-setting and enhanced collaboration are noteworthy because these characteristics did not surface in the studies cited earlier.

Although the Dana Center authors recognized that local circumstances influenced the ways the five high schools used these practices, they highlighted some essential elements that appear to have merit, no matter the circumstance. One was the desirability of training teachers in how to use assessment data without burdening them with the task of compiling and disaggregating the data. Relieving teachers of this burden freed them to concentrate on their main task: improving instruction based upon the data. Professional development time could then be allocated to developing the capacity of teachers to deliver high quality instruction.

Dana Center authors emphasized the interconnectedness and interdependence of these practices, yet, they also highlighted the way in which the use of data influenced the other practices, noting that "the staff at the schools used student assessment data to set

clear goals for student achievement, to determine areas where instruction could be improved, to focus on individual learning by identifying students who needed additional support, and to enhance collaboration around the academic goals of the campus" (Dana Center, 2001,p. 15).

Although the authors do not discuss the degree to which schools in this study used a top-down decision-making model, one can infer that a top-down structure existed. For example, the practice of using data to set academic goals was fostered by campus and district administrators who ensured data was collected, analyzed and disseminated in a timely way. Furthermore, administrators not only assisted teachers in accessing and interpreting the data, they also disaggregated and presented data in a way that allowed teachers to see student progress at the student, teacher, and grade levels (p. 18). In general, the data consisted of performance on the TAAS, as well as performance on the end-of-year course and AP exams, however, individual schools gathered other data including school-level assessments of student process at six or nine-week intervals, a step which promoted continuous monitoring of progress. In one school, "the staff developed their own assessment tool for a baseline measurement of the reading abilities of all freshman and developed similar assessments in writing, biology, and U.S. History" (Dana Center, 2001, p.17).

All schools provided staff development designed to show teachers how to use data to make instructional decisions. Of interest is an Eight-Step Process developed in the Brazoport Independent School District. (One can infer a top-down structure from some of the words or phrases used in the model: establish a timeline/calendar; use calendar to prioritize; establish an instructional focus as a school wide priority; report progress to

principal at the local and district level.) This continuous improvement process includes

the following steps[a]:

- Disaggregate the data by individual student and classroom and identify strengths and weaknesses.
- Establish an instructional timeline which states what will be taught, when it will be taught and when it will be assessed
- Create an instructional calendar based on the timeline. Use the calendar to prioritize deficient skills, and target skills to be repeated throughout the year
- Establish an instructional focus as a school wide priority (i.e., if summarization is the skill to be taught, then discuss it in reading, science, and even summarize how to play a game in PE.)
- Assess students for mastery and analyze the results to determine patterns in their responses. Log results and use them to make instructional decisions over a three-week period. Collaboration is an important part of this step. Progress is reported by the principal at the local and district level.
- Provide tutoring for struggling students and administer assessments to track the students response to instruction
- Provide enrichment for students who have demonstrated mastery.
- Provide maintenance activities so that skills are retained.

[a] Retrieved and adapted from http://brookview.warren.k12.in.us/ instructional/8step.asp. Retrieved July 17, 2007

Teacher collaboration also played a role in a study conducted by Rumery (2000).

The author described a data-driven approach used in Maine that combines the use of data

generated by state tests with data generated at the local level. This approach, which

appears to have a strong professional development component, was initiated after the

Maine Educational Assessment was aligned with Maine's Learning Results, a curriculum

framework. Rumery says that because of this alignment, "the resulting data on student

achievement can be used more comprehensively—along with data generated at the local

level—to plan classroom, school, and district improvement strategies" (p. 3). According

to Rumery, the state conducted scoring sessions through a process which includes

benchmarking activities using a scoring rubric and live student papers. (This type of

professional development has a strong collaborative component.) Rumery reported that

teacher surveys indicate this professional development experience assists them in the

classroom and within their schools and districts. "Schools from throughout the state adopt

the scoring rubric and students understand what they need to do to improve their writing"

(p. 4). The author does not mention if Maine classroom teachers promote writing that is

not influenced by the demands of the Maine writing assessment. Consequently, we don't

know if this state-wide practice limits students' opportunities to write for a variety of

audiences for a variety of purposes, nor do we know how or if teachers use the scoring

rubric to monitor progress or guide instruction.

Standardized Tests in Low Stakes Contexts

Brozo and Hargis (2003) report on how assessment data was used to drive

instruction at a Tennessee high school. The work of these authors is an interesting

contrast to that reported in the Dana Center research cited above. Brozo and Hargis are

critical of the overuse and abuse of standardized testing and believe the "best assessment

for learning occurs within the context of daily instruction and situated learning activities"

(p. 61). However, they believe standardized reading assessments tests can be used in a

low-stakes fashion as tools to improve teaching and learning. The authors explained how

test results can be used in this way by recounting their experiences providing consultative

services to a high school in a small Tennessee school district that had received a

Tennessee Goals 2000 Grant. According to Brozo and Hargis, "the overarching goal of

the grant was to determine the reading abilities of all students and the effectiveness of

initiatives to improve them" (p. 61). The authors use a computerized reading test to

determine a general instructional level, followed by either the Gates-MacGinitie Reading

Test or the Nelson Denny Reading Test. After the tests were scored, the authors

assembled the data and shared the each student's results with the teachers. In this respect,

the role of the researchers mirrors that of administrators in the Texas study who

disaggregated and organized the data in order to make it accessible to teachers.

Prior to receiving the test results, the teachers had been unaware of the reading

abilities of their students. The results, which were surprising to teachers, revealed that

35 % of the students (n=346) were reading one or more levels below grade level

placement, while 18 % were reading one or more levels above grade level placement.

Faced with the results, "many teachers found themselves forced to adjust their

assumptions about individual student performance in their classes" (p. 62). The teachers

realized that there may be underlying reading issues related to what they had perceived as

laziness or disengagement. For example, they found that a high school student perceived

as lazy was reading at a fifth grade level while a perceived "disengaged" student was

reading at a Nelson Denny grade equivalent of 17.9.

Year-long, reform initiatives at the Tennessee high school emerged in response to

the test data. The three initiatives included the implementation of sustained silent reading,

the use of young adult novels in content area classes, and providing alternative reading

assignments for both struggling students and superior readers. Reading achievement tests

were re-administered at the end of the year and the data demonstrated a jump in reading

performance. Sixty-five percent of students were reading at or above level in November

compared to 73 % of students reading at or above level in mid-May. Is this significant?

Brozo and Hargis (2003) make no such claims because the project did not involve a

controlled experiment nor was previous reading achievement data available for

comparisons. Nevertheless, they assert that the test results were the impetus for the reforms that were implemented: "Low-stakes reading achievement testing helped the teachers become much more sensitive to the importance of finding ways to accommodate the diverse needs of each of their students" (p. 64). There are some compelling attributes of Brozo and Hargis's work that connect this particular study to the studies conducted in high stakes atmospheres. For one, data from an externally created assessment provided the impetus for instructional changes. In addition, the changes occurred in an environment that was structured to support the use of data. Like district or building administrators in other studies, Brozo and Hargis, created a context in which teachers could examine and discuss the data.

Data and the Collaborative Process

A salient feature of the next study is that, like Tennessee study discussed above, the data from standardized but not high-stakes tests influence the change process. Furthermore, the study describes another type of supportive context—a context that is an attribute of the CIERA School Change Framework (Center for the Improvement of Early Reading Achievement). The Framework was designed to bridge the gap between our knowledge of school reform and best practices and the ability of educators to translate that knowledge into action at the school and classroom levels (Taylor, Pearson, Peterson, & Rodriguez, 2005; Taylor & Pearson, 2005). Its design was influenced by the characteristics of effective high poverty schools noted by Taylor and Pearson (2005) in their review of such schools. These characteristics, which mirror many of the effective schools characteristics in the studies cited earlier in this review, include a focus on improved student learning, strong building leadership, strong staff collaboration, on-

going professional development, systematic sharing of data, the alignment of instruction to standards, and parent involvement. This school reform effort promoted reflective professional development which occurred at two levels: the faculty level and study group level. Within both settings, participants learned about effective school improvement approaches and about effective high poverty schools (Taylor, Pearson, Peterson & Rodriguez, 2005).

Multiple data sources guided the discussions at both levels. School level data described fluency, as measured by words per minute (wpm), and reading comprehension, as measured by a Gates McGinitie Reading Comprehension subtest. Teachers and schools also received reports "benchmarking their instruction activities and their students' performance against other teachers and students in a national data base of 13 schools, all which were part of the project" (Taylor & Pearson, 2005, p.238). In addition, researchers observed classroom reading practices of randomly selected teachers (two per grade in K-5). Each teacher was observed for one hour, three times per year. These teachers received copies of the observation and information on research behind the practices that were coded by the observer. Furthermore, they were encouraged to get help from the school literacy coordinator or project facilitator on how to interpret the observations.

Monthly large group sessions engaged the entire faculty in discussions about school change, professional development and the school-wide reading program. The faculty heard reports from the study groups, had the opportunity to discuss reading instruction and curriculum, and set goals based on data. In addition, collaboration among teachers is evident. Teachers worked in within-grade and cross-grade small study groups which met at least three times per month. The study groups developed actions plans

"specifying what they were focusing on, what would be done, when the activities would be completed, and how the successes of the study group would be measured" (Taylor et al., 2005, p.47). In addition to action plans, the small study groups read and discussed articles about reading practices supported by research, watched videos of effective practice, as well as their own practice, and engaged in problem solving.

Taylor et al. (2005) were interested in learning if the implementation of the CIERA School Change Framework had a positive impact on reading and writing growth across schools. Schools were rated as high, medium, or low based on a scale built to determine how well the study groups implemented the Framework. The authors found that a school reform-effort rating was positively related to increases in students standardized comprehension scores, as well as their fluency scores. Furthermore, the reform effort accounted for 65% of the variance between schools over a two-year period. A cluster of influential practices seemed to contribute to the successes of the high reform effort schools. These schools had study groups that engaged in a study group topic on research-based reading practices for three to four months and met monthly as a large group to discuss the study group topics. Schools in the high reform effort category also made more of an effort to examine the data on their teaching practices and to make changes suggested by relevant reading research. The CIERA study appears to exemplify a situation in which educators used *process data* (Fox, 2001). He called process data "change data" because it refers to the "components, attitudes and practices that comprise the educational program" (p.13). Fox contends that process or change data are the only data that teachers and administrators control and that nothing will change until process

data are addressed and there are changes in teacher and/or administrative behavior (Fox, 2001, p.13).

Taylor, Pressley and Pearson (2000) compared five large-scale studies on effective moderate to high poverty schools—schools that showed greater than expected reading achievement despite the poverty level of the students. They concluded that that both classroom level factors (instructional) and school level factors (organizational) are key and that "effective schools systematically used student assessment data, usually on curriculum embedded measures to improve performance" (p. 5). The use of data was often part of a collaborative model in which teams of teachers examined the data and then planned instruction for a student.

We can appreciate how collaboration and assessment work in tandem thanks to the work of Taylor, Pearson, Clark and Walpole (1999, 2000) who examined the instructional and organizational factors in 14 high poverty schools that achieved greater than expected gains in reading achievement for primary grade students. The researchers embarked upon this study in hopes of discovering patterns that might inform future CIERA projects related to improving reading achievement in schools. By gathering and analyzing data from multiple sources—observations, teacher logs, teacher questionnaires and interviews— Taylor et al., 1999, 2000) discerned a number of variables that could explain the difference between more effective and less effective schools. One factor, small group instruction, is of particular interest for the purposes of this review of literature because small group instruction was informed by the collaboration between the classroom teacher and other building specialists (Title 1, Special Education, ELL) and was also guided by assessment data. According to the authors, the small groups "tended

to be based upon ability" however group membership was fluid and "movement was common because of the school's commitment to regular systematic assessment and early intervention."

Assessments were used at least three times per year to monitor progress. The results provided benchmarks for each student's progress and served as a basis for moving students to a different instructional group, thus preventing the small instructional groups from becoming "rigid and inflexible." Assuming students were actually moved if their scores indicated as such, these assessments clearly guided instructional decision making. Additionally, assessments also appeared to be a factor in driving the collaborative way in which classroom teachers, specialists, and paraprofessional organized the delivery of small group instruction—the type of instruction which predominated in the effective schools. In addition to guiding instruction, the internal monitoring systems also served as a form of internal accountability because the results were shared among colleagues.

It is important to note that the use of data was embedded in a larger system of reform. Of the four schools judged as most effective, three had already implemented a formal reading intervention program while another had implemented an all school reform model. Furthermore, the effective schools had a strong commitment to reading, spending an average of 134 minutes per day on reading compared to the 113 minutes spent with less effective schools. Notable instructional practices in the effective schools included explicit phonics instruction coupled with coaching children how to apply their knowledge when reading, higher level questioning, and more time spent on independent reading.

The instructional and organizational factors identified above also surface in a study of urban elementary schools discussed by Johnson (2002) who identified a number

of school improvement strategies linked to improved achievement. Achievement data played a key role in shaping the instruction in these schools. Principals and instructional coaches helped teachers learn how to disaggregate data in order to identify students in need of additional support. These schools also focused on a curriculum alignment process which ensured that students were exposed to the knowledge, concepts, and skills needed in order for them to be able to perform well on assessments. In addition, teachers had time to review and analyze student work, as well as time to create and score practice performance assessments.

Teaching to the Test

Gove (2006) investigated the perceptions of 15 teachers in a Reading First district regarding the usefulness of the DIBELS assessment and how the results influenced their classroom instruction over a year and a half period. DIBELS, which is the acronym for the Dynamic Indicators of Basic Early Literacy Skills, is a series of timed subtests which measure a child's ability to recognize and segment sounds and decode nonsense words, as well as the child's reading rate and retelling fluency. Teachers entered children's responses into Palm Pilots which streamlined scoring and recording of the data.

Gove (2006) reported that prior to the use of DIBELS as an assessment, many teachers in the district used whole group instruction. It appears that the data encouraged the greater use of either small group or individualized instruction matched to student needs. The data also assisted teachers in selecting suitably leveled materials. Unfortunately, the use of the DIBELS also prompted teaching to the test. The authors reported that some teachers had children practice reading nonsense words or had them practice reading quickly without emphasizing the importance of comprehension. Sadly,

these practices are counterproductive to developing competent readers, as well as counterproductive to the concept of using data to guide instruction.

Gove (2006) used a structured interview format with questions focusing on general perceptions of the Reading Program and specific questions about the use of DIBELS and the management system associated with it. Thus, the scope of the study does not provide insight into contextual factors surrounding the use of the assessment results. We can assume that professional development was lacking as one teacher commented that teachers did not receive training on what to do differently in light of the assessment data. We might also infer a top-down structure given the district's Reading First status. We don't know if other assessments were used systematically in addition to DIBELS, if collaboration took place among the teaching staff, or if students grew in their reading achievement.

Timed assessments consisting of a series of subtests like the DIBELS are a stark contrast to assessments such as those used in conjunction with New Zealand's National Educational Monitoring Project. Johnston and Costello (2005) cite one such assessment described by Flockton and Crooks (1996) in which children take the role of a class library committee. The children are given a set of books and are required to select the best books. Working within a time limit, each child, as well as the group as a whole has to justify the selections. The activity is videotaped so that teachers can later analyze students interactions with the texts. Johnston and Costello (2005) praise this type of assessment as one that examines the "independent and interdependent practices central to a democratic classroom and society."

Studies of Exemplary Teachers

Pressley et al. (1997) used a survey to examine the instructional practices of fifth grade teachers considered by their supervisors to be effective teachers of reading and writing. The study was driven by the belief that the voices of practitioners were missing from debates surrounding the topic of quality reading instruction. Survey data from 62 teachers indicated the teachers were committed to eight instructional components, one of which was an "extended evaluation of literacy competencies using diverse assessments" (Pressley et al.,1997, p. 1). Assessment was not a stand-alone, but was part of literacy instruction which balanced many components, including whole language experiences and skills instruction (p. 1). The additional components included: extensive reading; varied grouping patterns; instruction in word-level and comprehension level skills and process; background knowledge development; the writing process and mechanics; literacy in the content areas; and student motivation for reading and writing.

Because this study was structured as a survey, we only know that the exemplary teachers used varied assessments including commercial tests, standardized tests, writing portfolios, responses to daily reading (i.e. both written and oral responses to literature), and curriculum driven tests. Furthermore, a study designed this way can only capture what teachers report. It cannot provide information about actual practices related to assessment or to any of the other eight identified components. Consequently, we have no details about the reported diverse range of assessments or how the data from these assessments were evaluated and used to inform instruction. We also don't know about the ways in which pressure or lack of pressure from external assessments influenced

teachers' practices or the expectations the district might have placed on teachers. Practices related to goal-setting or professional collaboration were not mentioned.

Pressley et al. (1998) conducted an observational study to determine the practices that distinguish outstanding primary-level reading literacy teachers from those who were more typical or representative of the teachers in their district. These 30 teachers were observed repeatedly with observers attending "especially to teaching processes, the types of materials used in the class, and student reading and writing outcomes" (Pressley et al., 1998, p. 10). A formal interview was used to enhance the investigators' understanding of what he or she had observed in the classroom and included a member check typical of qualitative studies.

In general, the students of the effective teachers were engaged in literacy most of the time; were generally reading texts at the end of first grade that were at least end-of first grade level; and were writing compositions of three to four sentences with good spelling and use of conventions (Pressley et al., 1998, p.7). Reading test scores were descriptively better, with word analysis scores being statistically better. A stunning finding was that "the standardized test performance of lower achieving students in the most-effective-for-locale classrooms equaled or surpassed the achievement of the "average student in the least effective for locale classrooms" (p. 14). Characteristics and features consistently associated only with the most effective classrooms included: high academic engagement, effective classroom management; a positive classroom environment; a balance of skills instruction and immersion in literature and writing, matching demands to student competence; encouraging self-regulation; and curricular integration (p. 16). As with the study of exemplary fifth grade teachers, we don't know

about the ways in which pressures or lack of pressure from external assessments influenced teachers' practices or the expectations the district might have placed on teachers. Practices related to goal-setting or professional collaboration were not mentioned.

In terms of assessment guiding instruction, Pressley et al (1998) reported that the most effective teachers "consistently monitored students as they read and wrote and offered mini lessons on an as-needed basis." In addition, the effective teachers consciously matched children with books appropriate to their reading levels. Unfortunately, the authors do not provide detailed information about the ways in which these teachers consistently monitored progress and how closely they reassessed students following a mini-lesson. Nor do we know how they used assessment data to match children with the right books. One leaves these studies of effective teachers wondering how the monitoring of student progress, assessment and data usage of effective teachers compare to that of teachers judged as typical.

One might infer that the effective teachers in this study exemplified the implementation of the best literacy assessment practices promoted by Winograd, Peter, Flores-Duenas, and Arrington (2003). In such an environment, teachers gather multiple kinds of evidence and "engage in rich discussion about how to help children become better readers, writers, listeners and speakers" (p. 208). Furthermore, an environment like this promotes the "most effective literacy assessment"—that which occurs "when a competent teacher and a confident student can work side by side in a trusting relationship that focuses on growth, nurturance, and self-evaluation" (p. 206).

Ensuring that this kind of environment exists for all students comes with challenges. Winograd et al. (2003) write about the systemic nature of these challenges, the first being the need to educate teachers who are "well-prepared in the instruction and assessment of literacy." The authors also highlight the need for schools to be structured to support the teachers' use of assessment, noting that it is unrealistic and unfair to expect more of teachers who already have a heavy workload and too little time to do more. They write, "If we truly believe that teachers should engage in the assessment of literacy, then we must strengthen the ways teachers are prepared and improve the way schools are organized" (p. 206).

Themes Emerging from the Literature

One cannot escape the fact that the data from summative assessments developed by an external source are an element in data driven decision-making. This appears to be the case whether the data are used in a high stakes or low stakes context. The following paragraphs will examine this argument from high-stakes contexts first.

I define a high stakes context as one in which data from an externally imposed assessment galvanizes a district or school into action for the express purpose of improving student achievement on the externally imposed assessment. Examples include the Delta district in Illinois (Weinzierl et al., 1993); improved schools in Ohio (Kercheval and Newbill (2002), in North Carolina (Cobb, 1997), and in Texas (Dana Center, 2001). In contexts like these we can see how the organizational structure of a district or school fostered the use of data. Stephens et al. (1995) in a cross-site analysis of four districts, including Delta, contend that standardized tests appeared to influence instruction in districts with top-down decision making structures; conversely, standardized tests had

less influence over instruction in districts with collaborative structures or in those that valued teacher autonomy. What Stephens et al. (1995) fail to do is to fully examine the nature of external pressure and the degree of influence it has on whether a district will impose a top-down structure or support a culture that values teacher autonomy.

By way of example, we can infer external pressure had a great deal of influence on the decision-making structures in places like Ohio, North Carolina and Texas because the studies describe the ways in which leaders structured the data-driven environment for the purpose of improving student achievement. In a top-down structure, the system is organized to support the use of data. Leaders put systems in place so that data are organized, analyzed, and disseminated in order to allow teachers to spend their energies on changing instruction in response to the data. The leaders also provide professional time for data-driven work and they dedicate resources practices materials and/or curriculum materials aligned with the test content.

What appears to vary is the degree to which leaders fostered collaboration in response to data. Collaboration emerges as a specific characteristic of the improvement efforts in Texas (Dana Center, 2001), and we can infer that collaboration was embedded in the analysis of student work that occurred in Maine (Rumery, 2000). What is important to notice is that the collaboration in a high-stakes context is employed as a strategy for using data from an externally imposed summative assessment. The value collaboration in a high stakes environment is that it supports a degree of teacher autonomy even within a high-stakes climate. For example, teachers in one Texas school "developed their own assessment tool for a baseline measurement of the reading abilities of all freshman and

developed similar assessments in writing, biology, and U.S. History" (Dana Center, 2001, p. 17)."

The question we have to ask about collaboration in a high stakes environment is whether collaboration systematically narrows the curriculum matched to the test or whether it fosters an environment that enables teachers to meet the expectations of externally imposed assessments while exposing students to content and experiences that go beyond that which is tested.

Now let's turn our attention to the use of summative assessments as a driving force for data driven decision-making within a low stakes context. In such a context, summative data from standardized measures may have an influence on the actions of staff because they provided an external perspective on student achievement without the external pressure that can accompany high-stakes results. In these contexts the strong top-down administrative structure present in high stakes schools is absent, however, there is a palpable structure that accompanies the use of the data. For example, when Brozo and Hargis (2003) conducted their study of a Tennessee high school, the researchers served an administrative function because they analyzed and disaggregated test results for the staff. This freed staff from that burden and enabled them to focus on what they might do in response to the data. It wasn't a top-down structure that supported the use of data but a researcher-driven structure that inserted data from an externally designed instrument into the system.

A similar context appears to exist in the studies conducted by Taylor et al. (2005) because participating teachers and schools received reports "benchmarking their instructional activities and their students' performance against other teachers and students

in a national data base" of other schools that were participating in the study (p. 238). Researchers also provided teachers with observational data of their classroom practices and information on research coded by the observer. One can assume that this kind of systemic support from the researchers freed teachers to fully attend to discussions of reading instruction and curriculum and to the work of setting goals based upon data.

No matter the setting, a key theme that emerges from the literature is that systems are organized in particular ways. Some systems have a decidedly top-down structure; others have a more autonomous structure. However, what we notice is that even in those schools with more autonomous structures, data from some type of external summative measure informs decisions even when the assessment is not a high-stakes instrument. Even in cases where formative assessments are used to inform instruction and monitor progress, it seems as if schools are still tied in some way to the content imposed by summative assessments. This is not to say that the content imposed by summative assessment is irrelevant or unnecessary, or that formative measures should trump summative ones. It is simply recognition that externally developed summative measures influence the way schools do business.

The studies in this review of literature also demonstrate that the interrelationships between multiple factors influence student achievement, with assessment being one of those factors, rather than a distinct entity. Key factors include strong leadership, professional development, an emphasis on literacy, data analysis, and tracking intervention and remediation strategies and test prep strategies (i.e. Kercheval & Newbill, 2002; Cobb, 1997; Dana Center, 2001; Taylor, et al., 2005). It seems that the challenge for researchers and practitioners is to devise a way to balance these interrelated factors in

a way that serves accountability purposes without losing sight of the responsibility to

develops students who are life-long readers and deep thinkers.

Data Driven Decision-Making:
A Case Study of How a School District Uses Data to Inform Reading Instruction

Chapter 3 – Methodology

Introduction

Chapter 1 examined the range of perspectives surrounding the concepts of data and assessment—perspectives that reflect philosophical differences regarding the methods and purposes of assessment, the intended audiences for the results, and the level of student and teacher involvement in the process (Serafini, 2000, p. 208). Serafini argues that philosophical views affect classroom practice because teachers are influenced by their assessment paradigms when they gather and interpret data about student performance. Likewise, one could argue that leaders at the school and district levels are also influenced by their assessment paradigms, paradigms which in turn influence administrative expectations regarding the use of data. These expectations may then shape procedures and activities at the classroom level.

The studies reviewed in Chapter 2 illustrate the interaction between those systemic expectations—either tacit or explicit—and district, school and teacher actions. The studies also demonstrate the interconnectedness among the varying factors that influence student achievement, with assessment being one of those factors, rather than a distinct entity. The current study focused on understanding how educators in a Wisconsin district use assessment data to inform reading instruction, recognizing that data-guided instruction cannot be examined in isolation but must be studied within a real-life context. Questions designed to uncover the "how" of assessment in practice included the following:

- What reading competencies do districts measure and why?
- Who selects what competencies should be measured, how they should be measured, and the context in which these assessments occur?
- How do conceptions of data and conceptions of assessment shape policies and procedures?
- How do conceptions of data and conceptions of assessment inform programmatic or curricular decisions?
- How do conceptions of data and conceptions of assessment inform instruction at the classroom level?
- What evidence demonstrates that the use of assessment data improves student learning?

The remaining part of this chapter is organized into four sections. The first section explains the research model. It describes the characteristics of qualitative methodology, the appropriateness of the case study as a research strategy, and the rationale for the research design. The second section describes how the case district, schools within the district, and participants within those schools were selected. It also provides a brief profile of the participants. The third section explains the data collection process and the fourth section explains the data analysis process.

The Research Model

<u>Qualitative Methodology</u>

Qualitative research is an umbrella term for research strategies used to describe and interpret complex topics in a natural context (Bogdan & Biklin, 1998; Glesne, 1999, Merriam, 1998). According to Glesne, qualitative researchers approach their work from an interpretivist's perspective and operate under the assumption that, as researchers, "they deal with multiple, socially constructed realities or 'qualities' that are complex and indivisible into discrete variables" (Glesne, 1999, p. 5). Qualitative researchers also recognize that historical and cultural settings shape human interactions and the meaning

that is constructed (Cresswell, 2003). Those subjective meanings, in turn, influence behavior in a setting.

Given these intricacies, qualitative researchers are systematic observers who enter the field intentionally, all the while recognizing that their personal biases and their very presence in the research setting may influence what they ultimately observe. Nevertheless, qualitative researchers seek to understand by embedding themselves in the context, gathering information personally, and generating meaning from data collected in the field (Creswell, 2003). The use of multiple methods of data collection is often a hallmark of qualitative research with participant observation, interviewing, and document collection being the most dominant (Glesne, 1999). Once the data gathering is underway, qualitative researchers favor "an inductive orientation to analysis and findings that are richly descriptive" (Merriam, 1998, p. 11).

Unlike quantitative studies which follow a prescribed procedure, the design of a qualitative study, although systematic, may evolve over the course of a study because factors that surface in the field may influence the design in ways unforeseen before the study began. As Lincoln and Guba remind us, "What will be learned at the site is always dependent on the interaction between the investigator and the context, and the interaction is not fully predictable" (Lincoln & Guba, 1985, p. 208). This unpredictability means that that research questions may change and be refined and that "data collection processes might change as doors open and close" (Cresswell, 2003, p. 181).

The Case Study

According to Yin (2003),"Case studies are the preferred strategy when 'how' or 'why' questions are being posed, when the investigator has little control over events, and

when the focus is on a contemporary phenomenon within some real-life context"(p. 1).

Furthermore, Merriam (1998) defines "a case [as] a unit around which there are

boundaries"—one can 'fence in' what will be studied" (Merriam, p. 27). For the purposes

of this dissertation, the bounded contemporary phenomenon under investigation is a

public school district in Wisconsin situated in the current context of mandatory high-

stakes testing. In order to understand how a district's administrators and teachers use data

to guide reading instruction in such a high stakes context, I designed the study with the

intention of gathering a full variety of evidence, including documents, artifacts,

interviews, and observations (Yin, p. 8). Using multiple sources of evidence is critical to

case study research because the practice supports triangulation or converging lines of

evidence so that "the events or facts of the case have been supported by more than a

single line of evidence" (Yin, p. 99). The use of multiple sources supports creation of the

thick, rich description that is so often a hallmark of qualitative research.

As mentioned earlier, the design of qualitative studies may be emergent rather

than preordained. Case studies are no exception. In addressing this factor, Yin (2003)

writes, "Case study plans can change as a result of the initial data collection, and you are

encouraged to consider these flexibilities—if used properly and without bias—to be an

advantage of the case study strategy" (p. 77). Although the overall plan for this study did

not change over the course of the investigation, as one might expect, information that

surfaced early in the study shaped some of the questions that were asked in follow-up

interviews with a participant or in subsequent interviews with other participants. For

example, I learned about the district's reading curriculum document and a new electronic

assessment in my first interview. These documents became talking points when I interviewed other participants.

Research Design

Unique to qualitative research is that the researcher is considered the primary instrument for data collection and analysis; consequently, data are mediated through this human instrument. Merriam notes that, in contrast to instruments such as surveys or questionnaires, humans can be responsive to contexts, adapt techniques to circumstances, be sensitive to nonverbal aspects of a context, process data, clarify and summarize, and explore anomalous responses (Merriam, p. 7, citing Lincoln and Guba, 1981). At the same time, qualitative researchers bring their own biases to the field because "all observations and analyses are filtered through that human being's worldview, values, and perspective … which interacts with other people's constructions or interpretations of the phenomenon being studied"(Merriam, 1998, p. 23). Although researchers may be aware of particular biases they bring to the field, mere awareness does not negate a researcher's particular set of filters. Consequently, the best researchers can do is to openly acknowledge their worldview and then insert elements into the research design that might minimize the way in which their biases or experiences might interact with participants' potential responses. For that reason, I decided to use a semi-structured interview format in the field.

The worldview or perspective that I bring to this study is that of a part-time graduate student who is also a full-time school district administrator. I am in my 26th year in education and in my fourth year as an administrator. I first suspected that my administrator's "hat" was a permanent part of my worldview when Merriam's citation of

Macdonald and Walker (1977) resonated with me. They write about the political nature

of educational case studies funded by those who have direct or indirect power over those

who are studied and portrayed. Although this dissertation lacks the power of a funded

venture, I recognized that some of those I would interview are fellow administrators who

might feel a need to engage in face-saving or reputation building. Macdonald and Walker

say it well:

> At all levels of the system what people *think* they're doing, what they *say* they are
> doing, what they *appear* to others to be doing, and what in fact they *are* doing,
> may be sources of considerable discrepancy … Any research which threatens to
> reveal these discrepancies threatens to create dissonance, both personal and
> political. (Merriam, 1998, p. 43, citing Macdonald and Walker, 1977, p. 186,
> emphasis in the original.)

In order to avoid making participants feel uncomfortable, at least to the extent

possible in an interview situation, I decided to use a semi-structured interview format in

the field. By preparing opening questions ahead of time, I hoped to avoid language that

might appear to be judgmental or threatening. In fact, when writing the questions, I

thought about Spradley (1979) who writes that that an "ethnography starts with a

conscious attitude of almost complete ignorance" with the intent of capturing the *emic*

perspective or insider's viewpoint. Interview questions containing stems such as, *What*

can you tell me about; *Can you describe*; *What can you tell me so that I can best*

understand, exemplify my attempt to capture that *emic* perspective. Appendix A provides

a complete set of interview questions.

The intentional use of documents as a data source also assisted me in capturing

the emic or insider's perspective. Documents—whether they are created for a district

website or a classroom—are designed for a public audience other than the researcher.

Because documents are created with a specific purpose in mind, and are created

independent of the research study, they serve the purpose of capturing the emic

perspective in a context apart from what the participant says in an interview. The

documents also served as talking points within the context of the interview. For example,

one of the interview questions was*, I found this document on your website. What can you

tell me about the district vision or goal that influenced its creation?*

The documents were also invaluable for triangulating the data. For example, one

of the principals spoke about the district's early release days which were designed to

provide time for teachers to work in their Professional Learning Community (PLC)

teams. A district brochure about the purpose of the early release days confirmed that the

principal's interpretation of the PLC concept was accurate as defined by the district. The

purpose described in the brochure also aligned with the superintendent's explanation of

what the district expected to happen during that time. Of course, triangulation can only

take one so far. There was no way to confirm that the principal accurately reported what

was happening in the particular building on early release days. Although I have no reason

to doubt the principal's statement, this does illustrate a limitation of this study: the design

did not permit me to investigate events such as PLC team meetings which would have

given me insight into another aspect of data driven decision making in the district.

Case and Participant Selection

Defining a Pool

Purposeful sampling was used to identify a pool of districts from which a case

was selected. Potential cases were drawn from one region in Wisconsin, a region defined

by the boundaries of its Cooperative Educational Services Agency. The public school

districts in this region, like all public school districts in the state, face the same

expectations imposed by NCLB: they must meet interim benchmarks demonstrating adequate yearly progress and they must meet the 2013-14 expectation that all students will score at proficient or advanced levels in reading and math. Despite identical expectations, the level of urgency for meeting the proficiency requirements may vary among these districts due to differences in demographics. As discussed in Chapter 1, NCLB demands that states take responsibility for the academic achievement of all students, particularly for subgroups of students considered most vulnerable to failure— i.e., students of color, those with low socioeconomic status, English language learners, and students with disabilities. In real terms, this means that school districts with NCLB subpopulation groups of 40 students or more face the greatest pressure from high stakes testing because those subgroups are less likely to meet adequate yearly progress benchmarks. For that reason, I decided to narrow my pool of cases to districts in the region whose demographic makeup included some subpopulations of at least 40 students. A district of this type must make considerable gains if it is to meet the adequate yearly progress benchmark in 2011-2012, namely that 87% of all students will score at proficient or advanced levels. In other words, a district with those demographics must accelerate growth within four years.

Three practical considerations further influenced my selection of potential cases for the pool. Large urban districts were eliminated simply because, as a lone investigator, I felt I could not adequately study any large district. Similarly, I felt I could not adequately study all schools within a district so I limited the scope of my study to three elementary schools within a district. Time was a final consideration. The pool was narrowed to those districts within a reasonable driving distance. This allowed me the

freedom to conduct the necessary classroom visits during the school day without spending long periods traveling from my home base to the sites. Once I narrowed the pool to districts within a reasonable drive, I further narrowed the pool by examining data posted on the Wisconsin Department of Public Instruction website. From that pool of districts, I identified those that contained at least three elementary schools. Both practical and methodological considerations influenced this decision. I felt I could manage the logistics of studying three sites within a district while still tapping into the multiple perspectives that might exist across a district. Finally, I narrowed the target pool to four districts that had subpopulations likely to be at risk of missing the Adequate Yearly Progress benchmark of 87% proficiency in reading by 2011-2012. It is important to disclose that the subpopulation statistics in my pool of districts were similar to those of my own district. Certainly this similarity piqued my interest in these districts and influenced their placement in my pool. I contacted Superintendents by email and phone calls. Of the four superintendents, one, the Superintendent of the Valley View School District, was willing to meet with me and consented to participation after I explained the purpose of my study.

Selection of Participants

Purposeful sampling was also used to define the types of participants best suited to this study. Given the topic—that of how a district uses assessment to inform reading instruction—it was important to have participants in the study who had influence over the assessment process at their respective levels within the district. Consequently, district level administrators, building principals, building reading specialists and classroom teachers were recruited for the study.

The recruitment process began after I met with the district Superintendent and gained permission to conduct my study within the district. The Superintendent then introduced me to the elementary principals via email. Once the email was sent, I contacted principals individually to solicit their school's participation in the study. I was strategic in contacting principals. I needed participation from three Valley View elementary schools. In addition, I wanted the three schools to be located in different parts of Valley View for the simple reason that demographics and achievement levels varied somewhat from school to school. Although Valley View schools were relatively high achieving in comparison to the state average, the first three schools I solicited for participation still varied somewhat in their achievement levels as measured by the WKCE, well as in the level of diversity reported by the state. Once a principal consented to have his or her school participate, I recruited participants in a way that suited the principal. Two principals allowed me to present the intent of my study at a faculty meeting. Following the presentation, I contacted each teacher individually via email to solicit their participation. Another principal agreed to participate in the study, communicated the intent to teachers and gave given me permission to contact them. I solicited the participation of these teachers via email without having met them beforehand.

It is important to note that the three participating schools entered the study at different points during the school year. Although I had secured an informal consent to participate from three principals in August 2008, only one of those schools ultimately participated in the study. One principal later decided not to participate in the study. The other principal supported participation and allowed me to present information about the

study to school's faculty. Unfortunately, there were no teachers in this building who were interested in participating in the study. Fortunately, I was able to secure the participation of two other schools, both of which met my criteria for varying from the other school in achievement and/or diversity.

This unexpected delay in securing schools meant that participants were interviewed at different times of the year. Some were interviewed earlier in the school year, others later in the year. The time window may have influenced the way participants responded to some of the interview questions because new initiatives were being implemented at the time of the study. Consequently, participants who were interviewed early in the year had a different set of experiences with the new initiative compared to those who were interviewed later in the year. This is a limitation of the study.

Before beginning each initial interview, I reviewed the consent form with each participant and gained his or her written permission to participate in the study. Copies of the consent form are located in Appendix B.

Data Collection and Analysis

Data for this study were collected through the examination of documents, through interviews with the participants, and through classroom observations. The following sections explain how the data were collected and analyzed.

<u>Document Collection</u>

I collected documents from a variety of sources. For example, I downloaded two key documents from the district website. One described the elements of the district's long-range strategic plan; the other explained the purpose of early release days and the purpose of Professional Learning Communities (PLCs). I found the two documents after

doing a systematic search for anything that might provide insight into the district's philosophy and/or goals. I also searched for documents that referenced any or all of the following terms: *reading, reading assessment, assessment,* and *data-driven instruction.* This systematic search involved visiting each tab on the website and scanning the page for links to the topics I was interested in.

I gathered other documents during the interview process. If a participant mentioned a document, I asked to see it or I asked the participant to show me a document as an example. No one volunteered to share their materials, so whenever a document appeared to be of potential value to my research, I asked for a copy. Permission was granted in all instances but one. This will be discussed later. Participants provided me with a wide range of documents. For example, the first reading specialist I interviewed allowed me to photocopy the document that described the district's balanced literacy curriculum. The specialist also provided copies of the protocol for the district level reading assessment and sample reports from a new electronic assessment that was being implemented. A district administrator emailed me copies of the principals' building goals. Principals shared copies of documents like team goals and PLC team logs. Teachers allowed me to copy documents such as running records, informal assessments, discussion rubrics and observation logs. Selected documents are featured in the Appendix section.

Interviews

As mentioned earlier, participants in the study were purposefully selected to include individuals who had influence over the assessment process at their respective

levels within the district—district administrators, principals, reading specialists and classroom teachers. This aspect of the study was challenging in two ways.

The first challenge was the reality of voluntary participation. As I mentioned earlier, securing three schools and participants from each school proved to be more difficult than I had anticipated. The second challenge was that of scheduling interviews at a time in which both the researcher and the participant were available. It was especially hard to schedule interviews with administrators. Some interviews had to be rescheduled several times. The weather also forced rescheduling. A snow day and two cold days created a rescheduling nightmare. All 17 participants were interviewed face-to-face for the first interview. I promised that interviews would last between 60 and 90 minutes and adhered to that promise as closely as possible.

Follow-up interviews were conducted at a time that suited the participant. Some follow-up interviews occurred in person, immediately following a classroom observation; others occurred on-site at another time. Some follow-up interviews were conducted by phone. Although I allowed 30-60 minutes follow-up interviews in my protocol, I found that the follow-up interviews or phone calls lasted about 20-30 minutes. I think that was because the follow-ups were questions directly related to what I had observed or were clarifications of material covered extensively in the initial interview. Most participants appeared to share their thoughts, ideas and materials freely. Others were more guarded. Although they might have opened up if I had been able to spend more time with them, the study was not designed to allow for extended interactions with the participants. This factor is a limitation of the study.

Before beginning each initial interview, I reviewed the consent form (See Appendix B) with each participant and gained written consent to participate. Interviews were guided by the questions listed in Appendix A. Although some questions were tailored to the participant's role, all questions probed for information about the purpose of assessment, the kinds of assessments that were used, the reading competencies the assessments measured, and the actions related to policy practices or instruction that were informed by the data. Interviews were tape recorded and transcribed by me.

Observations

Classroom observations provided another opportunity for data collection. Once I completed an initial interview with a teacher, I scheduled one appointment to visit the teacher's classroom and observe the reading period in action. The length of the observation varied depending upon the length of the teacher's reading block on the day I visited. Observations ran from 50 minutes to 60 minutes. The overall purpose of the observation was to establish a shared context between the classroom teacher and myself so that I could use my observations as a starting point to discuss how the teacher used assessments to inform the lesson or how the teacher might use data collected during the lesson to inform future lessons. This purpose guided my classroom observations. During the lesson I listened for comments teachers made about student reading achievement that seemed to inform the lesson that was taking place. I watched for evidence of assessment—either informal assessment through questions or the ways in which a teacher used a checklist or observation log or real time student work to make decisions. I looked over students' shoulders to see their self-assessments—the logs or journals they were

keeping or the rubrics they scored. I did not interview students; I asked the teacher about work I observed as part of the follow-up interview.

In general, the teacher introduced me to the class with the explanation that I was conducting research for my dissertation and that I was visiting to see the kinds of things students did during their reading period. This set the stage so that it was easy for me to pull a chair around to shadow the teacher working with an individual child or small groups of children. As a result of these observations, I gathered additional documents for my collection, including blank samples of student logs, as well as samples of teacher notebooks or record sheets. Observation notes were transcribed for later reference.

Data Analysis

Documents

The documents gathered during this study varied in purpose and complexity and were suited to different types of analysis. The most complex documents were created at the district level and described either planning or programming. When working with these lengthy documents, I analyzed each document holistically to get a sense of the overall content and purpose of the document. This initial analysis was followed by a closer reading in which I looked for content that was connected—even in tangential ways—to the topics of reading, reading assessment, assessment, and data-driven instruction. By purposely casting a wide net for themes, I ensured that I didn't prematurely narrow the data set; I did not want to compromise the potential for triangulation that might occur when other data emerged. Next, with post-it notes in hand, I employed open coding in which I systematically examined a document sentence by sentence. When I noticed a salient topic, I labeled it with a post-it attached to the margin.

I then created summaries of the larger documents. This process helped me focus on the

critical attributes of the documents that were related to the topic of data-driven decision

making. As a way to keep track of my discoveries, I created lists of the topics I had

coded. The lists, which were similar to grocery lists, served as reminders of all the codes

I had discovered. I used the lists as a reference for typing my codes into textboxes in

Inspiration, a software program which creates graphic organizers. The full process I used

is explained in the following section.

Interviews and Observations

I transcribed all my own tapes. Although transcription is time-consuming, the

process enabled me to perform an informal holistic analysis as I typed. After printing a

participant's transcripts, I followed the same general open coding procedure I used with

my document analysis. I skimmed the document, highlighted key passages and wrote

codes in the margins. This process enabled me to focus on related parts of the transcripts.

For example, when the transcript showed that that a participant spoke about the Measures

of Academic Progress (MAP) assessment, I wrote MAP the margin. After coding MAP a

few times, I began to distinguish among the different ways that particular participant

referred to MAP. For example, MAP and district implementation, MAP and building

goals, MAP as formative assessment, MAP as summative assessment, MAP and staff

development. This process enabled me to gain a nuanced understanding of a topic. In the

case of the MAP assessment, MAP became more than a computerized adaptive test. It

was a computerized adaptive test inextricably linked to a particular participant's

perspective.

Once I analyzed data from documents and interviews, I began the process of comparing one segment of data with another. For example, the transcripts of the three principals contained references to the MAP assessments. In order to understand how all principals spoke about MAP, I copied the MAP-related portions of each principal's transcript into a new document. This enabled me to cluster similar comments together. Within the MAP section, comments about MAP assessment in light of a principal's building goals were placed together; comments about MAP as formative assessment were placed together, and so on. I repeated that same process with the other participant groups. Then I examined the documents from each group side-by-side in order to understand more about the ways in which participants in the four groups spoke about one topic. This comparative analysis was done in order to determine similarities and differences. I grouped data together on a similar dimension, named categories and looked for patterns (Merriam, 1998, p.18). Comparative analysis was a necessary part of the data analysis process because my key unit of analysis was at the district level. I needed to understand the parts before I could formulate an understanding of the whole.

Following the same approach used with the document analysis, I created lists of the topics I had coded and used the lists as a reference when I typed codes into Inspiration® software. Once the codes were typed, I organized the textboxes into categories, deleted redundant ones, and arranged the boxes to show how one aspect of the data was related to another one. I also began to apply selective coding in which I created labels for similar concepts. Some of these categories were inspired by the themes in the literature; however, my overarching categories became the strands in my Assessment Categories in Public Education model. It is important to note that this arrangement of

textboxes was an iterative process; I arranged and rearranged the textboxes multiple times before settling upon a final arrangement for each participant group (Appendices E-H). This process illustrates Miles and Huberman's (1994) comment about displaying data: "The creation and use of displays is not separate from analysis , it is part of analysis. Designing a display is an analytical activity …you know what you display" (p.25).

Overall, the process of analyzing, clustering, and re-analyzing, served the purpose of "triangulating" the data or using the data from one source to support themes discerned from another source. I found that in the process of reviewing, questions arose that called for additional evidence or clarification (Yin, 2003, p. 61). When clarification was needed, I would go back and review the original document or transcript to see where I might have missed a connection. Through this process I gained a true understanding of what Merriam described as "checking, verifying, testing, probing, and confirming collected data as you go." She said the process would follow "in a funnel-like design resulting in less data gathering in later phases of the study along with a congruent increase in analysis checking, verifying, and confirming (Merriam, 1998)."

Participant Profiles

As a researcher, I have an obligation to guard the identity of those who agreed to participate in this study and still provide enough details for readers so that they can understand the context of the setting. In order to protect confidentiality, I have used pseudonyms, for the district and the participating schools. I refer to the district as the *Valley View School District* and the elementary schools as *Woodland, Oakdale* and *Maple Grove*. Participants are identified either by role or by school affiliation and, where appropriate, references to gender were eliminated. Grade level designations were

generalized as either primary or upper elementary. Furthermore, I substituted general labels for specific labels that might give clues to the district's identify. For example, I referred to one assessment as the *district level reading assessment* instead of using the district's name for this assessment. I also used the terms, *Director* and *Assistant Director* for two district administrators in place of their actual titles. Finally, I summarized WKCE test data and demographic data rather than link the data to specific schools.

The Valley View School District

The Valley View School District is a public school district located in suburban Valley View, a city of approximately 50,000 residents. Although the city is primarily a residential community, it also has a thriving commercial area and is home to numerous businesses. Residents have easy access to cultural and sporting events, museums and universities. Valley View has two high schools, two middle schools and multiple elementary schools. At the time of the study, the district's student body was approximately 74% white and 26% students of color, with African American students representing the largest percentage of this group. Approximately 12% of the students qualified for subsidized lunch and 3% were considered as Limited English Proficient.

Overall, Valley View students score quite well on the Wisconsin Knowledge and Concepts Exam (WKCE), the state's criterion-referenced examination which is administered every fall. The percentage of elementary students in Grades 3-5 who scored at proficient or advanced levels ranged from 89 to 94%, well above the state's 74% benchmark. However scores for some subpopulations,—Economically Disadvantaged, African-American, and Students with Disabilities—are less robust. Between 58 and 79% of the students in these groups scored at proficient or advanced levels.

<u>Valley View District Administrators</u>

The Valley View district level administrators were relatively new to the district. Before becoming the Superintendent in Valley View, the Superintendent had served as an administrator in other districts. Prior to the Superintendent's arrival in Valley View, there had been some contentious issues in the community related to enrollment and school closings, as well as concerns about the quality of the elementary reading curriculum. In response, the Superintendent involved a variety of stakeholders in the task of creating a long-range strategic plan for the district. The Superintendent believed the plan would foster a positive climate focused on student learning. Changes that arose as a result of the plan will be discussed in Chapter 4.

In addition to the Superintendent, two other district level administrators participated in this study. These administrators, the Director and the Assistant Director, were invited to participate because each had responsibilities related to curriculum, instruction, and assessment. Like the Superintendent, both of these administrators were relatively new to their positions within the district. The Director had worked extensively with assessment before coming to Valley View and had experience working with the Measures of Academic Progress, an electronic assessment that was being implementing in conjunction with the strategic plan. The Assistant Director had been a school administrator before accepting a district-level administrative role in Valley View. This individual had been actively involved in the strategic planning process.

Oakdale Elementary School

Oakdale Elementary School is located in one of the more affluent sections of Valley View. At the time of the study, the student body at Oakdale Elementary School was approximately 85% white. Fewer than 10% of the students were eligible for subsidized lunch or were classified as Limited English Proficient. Approximately the same percentage of students was classified as students with disabilities.

Although Oakdale School had some of the higher proficiency scores in the district, the Oakdale principal aimed for continuous improvement. The principal was open to the school's participation in this study and voluntarily offered to cover classes for teachers who participated in the study so that they could meet with the researcher during the school day.

Oakdale's reading specialist had extensive experience as a reading specialist in other schools within the district as well as at Oakdale. The specialist was committed to high quality literacy instruction and collaborated with other reading specialists and teachers across the district to promote best practices. The Oakdale principal and the Oakdale teachers who participated in the study spoke enthusiastically about the reading specialist and valued the support she provided to teachers and students.

The three Oakdale teachers who participated in the study were experienced teachers with 9 to 20 plus years of experience. One teacher taught at the primary level; the other two teachers taught at the upper elementary level. All three teachers said they collaborated with the reading specialist on a regular basis. In fact, when I observed the teachers' classrooms, the reading specialist was in the rooms some of the time either modeling a lesson, collaboratively teaching a lesson, or working with students.

<u>Woodland Elementary School</u>

In contrast to Oakdale Elementary, Woodland Elementary School had more diverse student body and a larger percentage of students who were eligible for subsidized lunch, classified as disabled, or classified as Limited English Proficient. Like the Oakdale principal, Woodland's principal also spoke about continuous improvement and wanted to see Woodland's proficiency scores rise. And, like the Oakdale principal, the Woodland principal was open to having the school participate in this study. This principal also voluntarily offered to cover classes for teachers who participated in the study so that they could meet with the researcher during the school day. In addition to the principal, Woodland's reading specialist and two classroom teachers volunteered to participate.

The Woodland reading specialist was eager to participate in the study and contribute to the research base. She was relatively new to the school and was eager to improve the school's literacy program. Before becoming a reading specialist, this participant had been a classroom teacher in the district. This prior experience as a classroom teacher gave her a unique view on the challenges teachers face when learning how to implement a balanced literacy model like the one used in Valley View.

Two upper elementary grade teachers from Woodland also participated in the study. Both were relatively new to the profession. The young teachers referred to each other as teaching partners and said they spent quite a bit of time planning together and trying to figure out the finer points of implementing a balanced literacy model within their own classrooms. Both teachers looked to the Woodland reading specialist for support on a regular basis. Furthermore, the pair actively sought to hone their teaching skills by participating in a workshop on teaching reading that was held in the district.

The Maple Grove student body was approximately 77% white. Thirteen percent were eligible for subsidized lunch, 10.3% were classified as students with disabilities, and 2.6% were classified as Limited English Proficient.

Maple Grove's students scored relatively well on the state test, but not well enough for Maple Grove's principal who wanted students to perform better. The principal made it a practice to track the school's achievement data and used it to promote conversations among the Maple Grove staff. The principal was eager to participate in the study and believed teachers in the building would learn something new through their participation. In addition to the principal, two teachers and the school's reading specialist participated in the study. Both teachers were primary grade teachers with between 7 and 10 years of experience. Both spoke about pursuing additional coursework or advanced degrees in reading and appeared to enjoy talking about their professional practice.

The reading specialist had served as a specialist at Maple Grove for a number of years. Although the Maple Grove teachers and the principal spoke enthusiastically about the reading specialist and valued the support she provided to teachers and students, this participant seemed reluctant to participate in the study, even after she had agreed to do so. She claimed she was too busy for an interview and was only willing to meet with the researcher after the school year was over. When we did meet, her comments were terse and it was somewhat challenging to get her to elaborate upon her answers. When asked to share a copy of a document she spoke about, her typical response was, "I don't have a sample." This demeanor set her apart from the other participants who were enthusiastic in

discussing their perspectives practices and were more than willing to share samples of documents related to the topics discussed during the interviews.

The next chapter will provide details about the perspectives and actions of the administrators and teachers who agreed to participate in this study. The chapter will begin with a discussion of the district administrators, followed by sections which are devoted to each of the other participant groups—the principals, reading specialist and classroom teachers.

Data-Driven Decision-Making:
A Case Study of How a School District Uses Data to Inform Reading Instruction

Chapter 4 Results

How does a school district use data to inform reading instruction? When I posed

this question in Chapter 1, I explained that this study would investigate the organizational

structure, assessment philosophy, and instructional practices of a school district in order

to learn about ways in which the district's administrators and teachers collect, analyze,

and use data to inform reading instruction in a high-stakes era. The general questions that

guided the investigation included:

- What reading competencies do districts measure and why?
- Who selects what competencies should be measured, how they should be measured, and the context in which these assessments occur?
- How do conceptions of data and conceptions of assessment shape policies and procedures?
- How do conceptions of data and conceptions of assessment inform programmatic or curricular decisions?
- How do conceptions of data and conceptions of assessment inform instruction at the classroom level?

I have attempted to answer these questions by gathering a wide variety of

evidence from documents, artifacts, interviews, and observations. Using multiple sources

of evidence is critical to case study research because the practice supports triangulation,

or converging lines of evidence, so that "the events or facts of the case have been

supported by more than a single line of evidence" (Yin, p. 99).

This chapter is organized so that the study's primary unit of analysis—the district

level—permeates the entire chapter. The first part of the chapter focuses on the content of

several district documents, including the balanced literacy curriculum, the long-range

strategic plan, and a brochure that described the district's focus on data-driven decision

making and assessment. These particular documents were selected for analysis because

they addressed one or more topics central to this study: reading, reading assessment, and the use of data. Furthermore, the documents provide a context for understanding some of the actions and practices of the Valley View staff members who participated in the study.

The second part of this chapter discusses the perspectives of the participants. Each group is addressed in turn. I begin with a discussion of the district administrators, the change initiatives they brought forth, and the ways in which those initiatives influenced reading instruction, reading assessment and data-driven decision making. Next, I focus on the perspectives and actions of building principals and how they implemented the district's change initiatives within their respective buildings. This is followed by a discussion of the building reading specialists and the ways in which the district's balanced literacy curriculum and the district's change initiatives influenced their work. The discussion ends with a discussion of the classroom teachers' perspectives on assessment and how it informs their instruction. The chapter closes with a summary of the key themes that emerged across participant groups. It also highlights themes that are unique to specific groups. I created a summary table of to capture the response of the different participant groups. The table is located in Appendix I.

District Documents

As mentioned above, three district documents— the balanced literacy curriculum document, the long-range strategic plan, and a district-produced brochure—were selected for analysis because they addressed one or more of the topics central to this study: reading, reading assessment, and data-driven decision-making. These documents were also selected because they provide a context for understanding some of the actions and practices of the Valley View staff members who participated in the study.

It is important to note that the three documents were created at different times, by different authors, and for different audiences and purposes. In succeeding sections, one will see that the participants' connections to these documents depended on their respective roles, their level of experience, and their exposure to professional development.

The Literacy Curriculum

The balanced literacy curriculum document, which had a revision date of 1998, was the oldest of the three documents. For some participants, this document represented the status quo because it articulated, as two reading specialists told me, "the way we do reading" in Valley View. It contained information about the philosophy, instructional practices, and assessments that have (at least officially) guided the district's elementary reading program for some time. From an outsider's perspective it seemed likely that this document would be obsolete by the end of the year because at the time of the study the district standards and benchmarks were being revised and a reading textbook adoption was in the early stages of the process. However, these district level actions were not yet impacting reading practices at the building level. Principals, reading specialists and classroom teachers who participated in the study knew that reading was a district focus and that some changes would be taking place. However, they knew little about the details. Those participants who were reading specialists or classroom teachers appeared to be taking a wait and see attitude. Basically, their day-to-day instructional approach was shaped by the balanced literacy philosophy described in the curriculum document. For that reason the core elements of this document merit closer examination.

As might be expected from any balanced literacy reading curriculum, this document emphasized the importance of students becoming effective communicators when reading, writing, speaking, listening, and viewing. It also stressed the developmental nature of reading as well as the importance of engaging students in a literacy environment that immersed them in reading, writing and speaking about a variety of genres. In terms of content, the document, which contained several sections, was a repository for all things reading. It included a scope and sequence, the protocols for the district level reading assessments, descriptions and teaching directions for various strategies, the Wisconsin and IRA/NCTE Standards, district reading goals, articles on struggling readers, and more. If that sounds like a lot of information, it is. In fact, two district administrators commented on the all-encompassing nature of the document and viewed it as too cumbersome to be useful to classroom teachers. Based on the comments of participating teachers, the administrators' analysis appeared to be accurate.

Participants who had more than three years of teaching experience made general comments about the literacy curriculum document when I showed them a copy during the interview. Although the balanced literacy philosophy guided their practice, I did not sense that any of the experienced teachers consulted the curriculum document on a regular basis. For example, when asked, one teacher said she referred to the document when selecting a genre for a new unit of study. "There are certain genres we know we need to touch on but we don't always do it in the same order," the teacher said. Overall, it appeared that the participating teachers collaborated with each other and/or their building reading specialist when it came to classroom teaching techniques or meeting the needs of struggling readers. The curriculum document did not appear to guide the daily instruction

or the on-going assessment practices of the experienced teachers. In fact, no experienced teacher had the curriculum document readily available when I asked about it. However, these experienced teachers spoke confidently about the balanced literacy model. In fact, all the experienced teachers spoke about the professional development they had received on balanced literacy when they came to the district. This professional development enabled them to understand the material in the curriculum document and enabled them to implement the balanced literacy model in their classrooms. Their professional development included instruction on how to do running records and how to conduct the district level reading assessment. One could infer that this training influenced the ability of these participants to use assessment data to inform their instructional decisions. That is because all experienced teachers spoke confidently about their ability to use the district level reading assessment to guide their instructional decisions. Their comments will be discussed later in this chapter.

Teacher participants who were relatively new to the profession (experience of three years or less) freely acknowledged that they were overwhelmed by the range of information in the document and had either barely looked at it and/or rarely used it. The curriculum document did not appear to guide their instruction or their on-going assessment practices either. Instead, they relied on their pre-service experience, limited training from the district, or guidance from the building reading specialist. Unlike their experienced counterparts, the two less experienced teachers were not provided with the extensive professional development that had been offered in the past. Although these young teachers looked to the reading curriculum document for district level reading

assessment protocols and the rubrics used for grading, they both expressed uncertainty in how to use the data to inform their instruction. This will be discussed in more detail later.

Because the assessment portion of the curriculum document guided the formal assessment practices of novice and veteran teachers alike, it's important to understand this specific component of the document in greater detail. The curriculum document stated that the district level reading assessments were part of a balanced assessment system and had multiple purposes: to provide a link between district standards and instructional practices; to document individual growth; to inform individual, small and whole group instruction; and to evaluate the K-12 program. Based upon my interviews, it appeared that the district level reading assessments mainly served two of those purposes—documenting individual growth, and informing individual, small and whole group instruction. Each teacher spoke about the students' reading folders as the place where they documented students' scores on the district level assessment. They also spoke about their practice of consulting the students' folders at the beginning of the year when they were in the processing of placing students in groups. Reading specialists spoke about consulting with teachers in light of the assessment results to determine appropriate group placement or intervention needs.

I did not come across any evidence to suggest that the assessment results were used to examine any links between district standards and instructional practices. This is evidenced by a comment about the district level assessment that was made by one of the principals:

> I don't have anything against [the assessment]. I think it's a good half-way judge because we do them in January. I don't have any problem with them. I don't know how we use the data. Maybe it's my fault but the … years that I have been principal, I haven't really looked at that data. It gets compiled and

it gets on a shared drive but we haven't really used here so my thinking is that if we have really been doing this for a long time and we've had this data, and I am a data [person] and I haven't used it and now we have other things. I just don't know how important it is. You know when we get together and our Superintendent asks us for our data, it's never [the district level reading assessments] we haven't had those conversations at our principal meetings.

Administering and scoring the district level reading assessments was primarily the classroom teachers' responsibility. These assessments were administered in January to all students in Grades 1-5. Students who performed below grade level on the January assessment were reassessed in May, and again the following September. New students in Grades 2-5 were assessed in September, or upon arrival. Teachers recorded the results on each student's reading record and used the results to guide the decisions they made about group placement and intervention needs. Reading specialists were consulted if a teacher had questions about a child's performance, and sometimes the reading specialists did additional testing to confirm the teacher's results.

Teachers administered the Rigby PM Benchmark (Nelly & Smith, 2000) individually to students in Grades 1-3. (Rigby was the tool selected by the district for the district level assessment at these grade levels.) Each student read passages orally and responded orally through retelling and answering comprehension questions. The scoring protocol included benchmarks for accuracy (percentage), fluency rate (WPM), fluency quality (rubric), retelling (rubric), and comprehension (percentage). The protocol also identified a benchmark proficiency level for each grade level. In addition to the scoring protocol, the curriculum document also described the competencies that were linked to what students would need to know and be able to do to succeed on the assessment. The knowledge and skills included: concepts about print, phonemic awareness, letter-sound

relationships, fluency, and story elements, operations on print, self-monitoring, searching for cues, cross-checking, self-correcting, word analysis strategies, story mapping, retelling, and making simple inferences. The primary grade teachers who participated in this study assessed these same strategies through on-going classroom assessments. Appendices P and Q provide examples of these assessments.

The district selected a different district level reading assessment tool for use with older students. The district's upper elementary teachers used passages from the Qualitative Reading Inventory III (Leslie and Caldwell, 2001) to assess their 4th and 5th grade students. The authors' protocol for an individually administered test was modified to accommodate group testing in an untimed setting. As a result, students responded in writing to pre-reading questions. They then read a passage silently, and responded in writing to literal and inferential comprehension questions. The critical knowledge and skills that were matched to this assessment included applying knowledge of text structure, such as story mapping, and finding the topic, main idea and detail, and inferring. The scoring protocol called for a comprehension score of 75% or better on grade level narrative and expository passages. Students who tested below the grade level expectation were reassessed with progressively less difficult passages. As with the primary grade protocol, results were entered on each student's individual record and teachers used the results to guide the decisions they made about group placement and intervention needs. Based upon what I observed, the upper elementary teachers who participated in this study assessed some of these same strategies through on-going classroom assessments which required students to respond either orally or in writing. For example, I observed one teacher circulating around the room for the purpose of listening

in on students' discussion about the main idea and details in a non-fiction article about a science-related topic. The teacher used the information to reinforce the concept of main idea during the whole group instruction that followed. In another upper elementary classroom, one of the rubric descriptors that assessed the quality of written responses said: *Post-it notes and responses show deep thinking (inferences) about the characters and important events in the story*.

Over the course of the interview process, I learned that some participating teachers administered the district level assessment more than was required. For example, two of the teachers administered the district level assessment to all their students in September. Another teacher participant assessed individual students as needed if the teacher needed more information about a student's reading progress. Most adhered to the district schedule for administering the test.

I also learned that participants who were primary grade teachers regarded the results from the district level assessments as an accurate measure of their students' reading abilities and used the results as a guide in determining next steps for a particular child. Teachers said the results of the district level reading assessment confirmed their hunches about children that they were already watching. Commenting on one student, a teacher said, "I knew she was making really good progress, and then, when I did the [district level reading assessment] it was like WOW she just kept going and going [from one level to the next on the assessment] and it was just wonderful and she was so happy and feeling good about her reading."

The primary grade teachers also spoke about cases in which they were surprised by the results on the district level reading assessment. For example, one teacher spoke

about a child whose poor comprehension skills were revealed by the district reading assessment: "In the small group she will talk a little bit, and in the whole group she will talk a little bit, but just reading on her own, there was no comprehension. She had the fluency part down and the decoding, and the word ID for the known words, but then the comprehension, no."

Results from the district level reading assessment, along with data gathered through observations and other classroom assessments, also influenced a child's reading grade. Teachers reviewed all the data they had collected on each student in data light of the district grading rubric for reading achievement. Then they assigned a grade defined by a 4-point scale (Appendix 0). The rubric was used during parent-teacher conferences to explain the child's progress and grade. "This helps me justify the grade I give a student," said one upper elementary grade teacher.

Overall, reading specialists and teacher participants who taught upper elementary students were beginning to wonder if the district level reading assessments prescribed in the curriculum document would be abandoned in favor of the Measures of Academic Progress, a new computer adaptive assessment which was in its first full year of implementation in the district. As will be discussed later, the results of this assessment were beginning to influence the culture of professional practice. MAP implementation evolved from action steps identified in the district's long-range strategic plan, an all-encompassing plan that had data-driven decision making and student achievement as a primary focus. This plan and the changes it was intended to foster are explained below. An explanation of the MAP assessment is also included.

Change Documents

Unlike the reading curriculum, which was a long-standing program document, two relatively new documents were selected for analysis because they embodied a vision for systemic change. Both documents—the long range strategic plan and the brochure that described the district's focus on data-driven decision making and student learning— touted the district's focus on student learning and its expectations for continuous improvement, data-driven instruction, and collaborative professional practices.

The Long-Range Strategic Plan

As soon as I received the Valley View Superintendent's permission to conduct research in his district, I began mining the district website to see what I might learn before I conducted any interviews. I looked for general information about the district's philosophy and goals, as well as for specific information about the topics central to my study: reading instruction, assessment and data-driven decision-making. My ultimate goal was to find documents or portions of documents that would serve as the focus of one of the pre-determined interview questions: *I found this document on your district website. What can you tell me about the district vision or goal that influenced its creation?* Fortunately, portions of Valley View's massive long-range strategic plan met my needs.

The plan fills eighty-six pages and addresses six major areas: curriculum, post high school preparation, community service, student conduct, communication, and employment practices. In general, the plan promotes continuous improvement with the goal that all Valley View students are able to demonstrate proficiency in the state academic standards.

Although I reviewed the entire document in preparation for this study, only the action plan designed to address state standards and curriculum is relevant to this discussion. This portion of the plan makes numerous references to reading, assessment, and data analysis. It not only acknowledged the "critical importance of reading and writing" and dubbed "reading and writing K-12 as foundation skills for success in all areas" but also identified reading as a area of focus in the district.

Unlike the reading curriculum document, which focused on a program model, the standards and curriculum action plan was more global in scope. Beyond specific references about the importance of reading and the mention of examples of specific assessments that could be used to assess reading achievement, it contains general directives. I have selected one objective and one related strategy which I believe capture the intent of the standards and curriculum action plan.

The measurable objective I selected states that "all students will demonstrate proficiency in the Wisconsin Model Academic Standards." The strategy that complements the objective states that the district will "perform a comprehensive, in-depth, data-driven analysis of student performance and current academic curricula, and will formulate an improved scope and sequence for each academic core area that will maximize achievement for all students." Broad statements like these are supported by a series of action steps covering many pages.

The supporting action steps are so tightly focused that they are easily summarized. Basically, the standards and curriculum action plan takes the stance that it is important for all students to demonstrate proficiency in the Wisconsin Model Academic

Standards and outlines steps to be taken to promote that overall level of proficiency. The

steps are summarized as follows:

1. Gather data continuously and systematically from multiple sources.
2. Train district personnel to use data to make instructional decisions.
3. Analyze data to determine strengths and needs.
4. Let data guide instructional decisions, including intervention and enrichment.
5. Assess and monitor short-term progress in foundational skills.
6. Compare student data annually and track long-term achievement gains in core areas across schools and among subgroups.

Even after distilling the standards and curriculum action plan into six steps above,

I wondered how the district might accomplish the ambitious objectives related to using

data continuously to guide instructional decisions. The "how-to" becomes clear when one

reads the brochure posted on the district website. In contrast to the long-range strategic

plan, which created a structure to foster change, the brochure specified the actions which

would underpin the focus on student learning—actions which included data-driven

decision-making and tracking growth in student achievement.

The Brochure

The Valley View homepage contained a prominent link to a page dedicated to the

topic of student learning. That page contained a downloadable two-sided brochure and a

link to a video of the document. The Superintendent, Director, and Assistant Director

recorded the the voice-over for the video which contained two still frames, one for each

side of the brochure.

One side of the brochure contained a diagram consisting of a circular core or

bulls-eye surrounded by concentric rings. This diagram demonstrated how various district

initiatives aligned with one another, as well as how they aligned with the concept of

student learning. The core of the diagram was labeled student learning. The concentric

rings identified the specific actions aligned with the core focus. These actions include: identifying the district standards and benchmarks that are aligned to the state standards; assessing student achievement of the standards and benchmarks through a variety of assessments; and collaboratively examining student assessment data followed by making adjustments to instruction based on the data.

In concrete terms, if one considers the diagram in the context of a curriculum area such as Reading/Language Arts, the district would first identify grade level standards and benchmarks that are aligned with the Wisconsin Model Academic Standards for English Language Arts. Then, student achievement related to those standards and benchmarks would be assessed through multiple measures, including the WKCE, the Measures of Academic Progress (MAP) and other tools. Finally teachers would collaboratively examine the assessment data and determine the appropriate focus for reading instruction as it applies to whole group instruction and to individual needs.

These actions were in various stages of implementation when this study was conducted. A district committee was finishing the work on identifying district standards and benchmarks for Reading Language Arts. District level assessments that had been previously developed to assess reading achievement were still in place and a standardized electronic achievement test called the Measures of Academic Progress Test (MAP) was in its first full year of implementation in Grades 2 -10. Principals, reading specialists and teachers were learning about the Professional Learning Community concept, including how to collaboratively examine student assessment data and make instructional decisions based upon the data. Since MAP data was often mentioned as one of the topics addressed during PLC meetings, it is important to understand more about this assessment.

Measures of Academic Progress

In brief, Valley View piloted the Measures of Academic Progress Test (MAP) in 2007-08 and implemented the test across the district for Grades 2-10 during the 2008-09 school year. MAP is a computer-based test that is aligned to state standards. A unique aspect of the test is that it is an "adaptive" test. This means that the computer adjusts the difficulty of the item each student will receive based upon his or her answer to the previous item. In other words, a difficult item will be offered when a child answers a question correctly. Conversely, a student who answers a question incorrectly will receive a less difficult item. A district newsletter posted on the district website explained MAP as an "assessment program that provides educators with the information they need to improve teaching and learning."

From reading various on-line newsletters that mentioned MAP, I learned that the district was evaluating the test to determine how it fit within its overall assessment plan. That determination would likely take some time because, when this study was conducted, all participants were at various stages in their understanding of MAP reports and how to use them. Specific comments will be discussed later in this chapter. For example, out of all the classroom teachers who were interviewed, one upper elementary teacher spoke with confidence about understanding the various reports and the way in which the report data could be used to inform instructional decisions. In contrast, another teacher commented, "We haven't looked at this all that much and to me it seems really overwhelming…depending on what your student's score is you might have 20 different pages [of MAP reports] that you are trying to plan instruction from."

Undoubtedly, district educators would become more comfortable with these data as time went on, especially since the district had allocated time during the contract day for data-related work. Much of this work would be accomplished through the district's Professional Learning Communities (PLC) initiative which was described on the second side of the brochure.

Professional Learning Communities

The district calendar ensured that teachers would have time each month to do the work of a professional learning community: create common assessments based on standards, analyze data and make instructional decisions based upon the data. This PLC work was slated to happen during three of the four district-wide early release days that were scheduled each month. The brochure listed the essential questions and focus activities that were to guide the PLC work. These appear in Table 4.1 below.

Table 4.1 Essential Questions and Focus Activities Identified in the Brochure

Four Essential Questions	Focus Activities
1. What is it we want our students to learn? 2. How will we know each student has acquired targeted standards and benchmarks? 3. How will we respond when students do not learn? 4. How will we extend and enrich learning for students who have mastered the standards and benchmarks?	1. Establish targets and benchmarks. 2. Clarify essential outcomes by grade or course 3. Develop common assessments 4. Analyze assessment results 5. Plan for intervention and instructional strategies

The PLC concept will be discussed again in other sections in this chapter. Before continuing, it may be helpful to summarize the discussion thus far. I selected these documents for analysis because they provided a context for understanding some of the

key structures and practices that were either in place or were in the process of being implemented at the time of the study. These documents existed for different purposes. The reading curriculum document established the philosophy for the elementary balanced literacy program and established the protocols for the district level reading assessments. Even though participants used the document sporadically, they spoke of the district's "balanced literacy curriculum" when they explained the ways in which they taught reading in their classrooms. In contrast, the long-range plan and brochure on student learning provided evidence of the planning behind the systemic change that was beginning to take place at the time of the study. The change involved a system-wide focus on student learning and emphasized the use of data-driven decision-making and teacher collaboration as key components of the process. Although the curriculum document may have shaped system-wide actions at one point in time, this was not the case when this study took place.

Despite their different purposes, the documents have some attributes in common. For example, all three emphasized the importance of reading, the importance of monitoring student growth in reading, and the importance of using assessment results to make instructional decisions. The documents also acknowledge the importance of standards and benchmarks. However, the curriculum document contained standards from several sources. This made it impossible to see a one-to-one connection between standards, instruction, and assessment. On the other hand, the plan and brochure explicitly connected student proficiency on the Wisconsin Model Academic Standards to a tight alignment between the standards, assessment, and instruction. How this would evolve into a new curriculum document remained to be seen. What's important to note is

the emphasis on the tight connection. The plan and brochure also called for *collaborative* data analysis and decision-making within the context of a Professional Learning Community. This notion of professional collaboration around student achievement was not evident in the balanced literacy curriculum document.

At the time of the study, all three documents influenced instructional practices in varying degrees. The existing curriculum document set the overarching tone for the literacy practices and the formal assessment practices of the reading specialists and classroom teachers who participated in the study. However, it was not used uniformly across the district, and was considered too cumbersome to be useful.

The concepts embedded in the long-range strategic plan and the brochure on student learning were just beginning to impact professional practices at the time of the study. Most notably, teachers were learning how to use student assessment data within the context of a Professional Learning Community structure. Additionally, the data generated through the MAP assessment was beginning to foster discussions about how this new data set could be used to guide instruction.

In general the reading specialists and teachers who participated in this study had little knowledge of the long-range strategic plan. They knew reading was a district focus and that some changes would be occurring, however, they were unaware of the plan's details. Principals were more familiar with the long-range strategic plan. Two principals could summarize the plan's intent and had been involved in either the planning process or in subcommittee work. On the other hand, district administrators lived and breathed the plan. The next section provides a glimpse into their world[1].

---

[1] Appendix I contains a summary table which features sample responses of participant groups organized by category. Readers may find it helpful to review the table before proceeding.

District Administrators

Valley View's district level administrators were relatively new to the district, and as sometimes happens, the board and community were looking to the Superintendent and other district administrators to usher in change. Prior to the Superintendent's arrival there had been some contentious issues in the community related to enrollment and school closings. There was a also small but vocal group of parents strongly opposed to the district's balanced literacy curriculum. Additionally, there was a concern among some community members that the literacy curriculum was not uniform across the district. In short, factions were fracturing the school community.

Consequently, the Superintendent envisioned fostering harmony among stakeholders and wanted to redirect everyone's energies around a common purpose. The process of creating a long-range strategic plan served to redirect those energies. "I thought it was time to shift the focus to what we are supposed to be about and to bring some people together. We brought together some of the people that were the loudest and the angriest in the mix to put a long range plan together to help us move our focus in a different direction," the Superintendent said. That "different direction" was a data-driven focus on student learning. At the time of this study the Superintendent and two other district administrators, a Director and an Assistant Director, were all working together to move the district in the direction established by the plan.

The Purpose of Assessment

When it is the Superintendent who is leading the way on data-driven decision making, it's critical to understand the views this individual holds about assessment because these views ultimately establish the tone for the entire district. When asked, the

Valley View Superintendent initially identified two purposes for assessment: to guide instruction and to document student growth. "If you are using assessment the way it should be used, it should be to guide what we do in instruction," the Superintendent said, adding that there was also a third purpose for assessment: accountability.

The Superintendent characterized the accountability that accompanied the WKCE as a reality that schools have to accommodate. "You just can't ignore it or pretend it's not important, because it is. And, as a district, we have to score well," the Superintendent said. The Superintendent believed the accountability purpose of assessment was generated by those outside of the educational system and that the results had limited usefulness. Nevertheless, the Superintendent also believed that it was important to have the right conditions in place so students would do well on the state test. This included, in the Superintendent's words, "teaching to the test."

While some might recoil at the thought, the Superintendent explained that both the district standards that were currently in place, as well as the Wisconsin Knowledge and Concepts Examination (WKCE) were based upon state standards. In that respect, the Superintendent considered teaching to the test appropriate—not in the sense of teaching questions verbatim—but in the sense of teaching to the standards underlying the test. "Teaching to the test, the way it is set up structurally, is the way we are trying to move our organization," the Superintendent said.

Moving an organization forward is a challenging task. In the case of Valley View, it appeared to call for work in three areas: staff development related to assessment, an alignment of standards and benchmarks, and the supervision of principals. The paragraphs that follow are intended to provide an understanding of how the

Superintendent and other district administrators perceived the challenges ahead and the steps they were taking to move the district forward.

<u>Gaps in Assessment Practices</u>

As mentioned in the beginning of this section, the Superintendent viewed the primary purpose of assessment as guiding instruction and tracking student progress. Unfortunately, there were gaps between what district administrators believed were sound assessment practices and some of what they had observed at work within the district. Closing the knowledge gap on best assessment practices was part of the work that had to be done and the Superintendent was upfront in addressing the issue. "I think we have a lot of misunderstanding about what assessment is and what it should be used for," the Superintendent said, citing the concept of final exams to make the point. "We have review week and final exam week. Then we rest up for a week and teach six weeks." The Superintendent criticized the final exam concept because it appeared that the results were never reviewed, and thus were never used to guide any instructional changes based on student results**.** Furthermore, the collaborative practices the Superintendent valued were generally absent when it came to analyzing these kinds of assessments. Frequency of assessment was another issue to be addressed. The Superintendent believed the existing assessments weren't administered often enough to determine how students were progressing. The superintendent commented:

> We have got to make sure that if we are going to remediate, we are going to remediate. We don't take a child and put them in a separate group for a year and just leave them there. And the only way you are going to know if they are making those gains is through assessing.

The scarcity of common assessments across the district also had this individual's attention. "We don't have a lot of common assessments and we have people who just don't believe we should be doing common assessments," the Superintendent said. In some cases, assessments that had been created had never been fully implemented and had then been dropped. In other cases, there was resistance to the concept of common assessments. The Superintendent attributed resistance to the tendency of some teachers to prefer isolation coupled with an attitude the Superintendent described as, "I want to be left alone, just let me do my own thing. I know what I am teaching." The Superintendent hoped that the implementation of MAP would help create an environment that fostered collaborative practices focused on analyzing MAP data and subsequently using the data to guide instructional decisions and the content of common assessments.

As far as common assessments that were in place, the long-standing district level reading assessment was one that, on the surface, appeared to be a common assessment. Yet, even this assessment concerned the Superintendent. That was because the Superintendent had evidence that the administration of the assessment, as well as the way in which results were used, varied across the district. In addition to inconsistencies in assessment practices, district administrators were concerned about the lack of uniformity in instruction and believed a core text would stem variations in practice across buildings—practices they believed contributed to some of the gaps in student achievement between buildings.

The Assistant Director, also had concerns about the existing district level reading assessments. The Assistant Director said that this assessment had "muddied the waters" when it came to an understanding of the term *common assessment*. The Assistant Director

believed teachers heard the words *common assessment* and immediately thought of the

district level reading assessment which was administered twice a year to all students. In

contrast, this administrator believed a common assessment could also be simply one

question administered at the end of the week. The Assistant Director explained that when

a team of teachers got together to analyze such an assessment, the results of the analysis

could lead to a discussion of instructional strategies that could be used to help all students

achieve a particular benchmark.

It was too early to say if the long standing district level reading assessment would

continue in its present form, albeit with more consistency, or if it would be abandoned.

According to the Superintendent, that decision would eventually come about through the

reading curriculum review process which was being undertaken as a result of the long-

range strategic plan. The Superintendent explained:

> We really have to start looking at how we are doing, not necessarily how
> we are doing as teachers, but how well our kids learning. That's the shift we
> have to get to. It's what are they learning, not what am I teaching.

Standards and Benchmarks

Some of the tasks embedded in the curriculum review process would help foster

that shift from thinking about the act of teaching to what students are learning. This

would be facilitated by a realignment of standards and benchmarks across all grade

levels. The Director said that content area teams, under the guidance of a consultant hired

by the district, would revisit the district's existing standards and benchmarks in light of

the Wisconsin standards, revise the district standards and benchmarks where needed, and

place them in a standard format. This work was underway at the time of the study.

In the case of Reading, revised standards and benchmarks would replace the lengthy and detailed scope and sequence that was part of the existing reading curriculum. Briefly, the process called for identifying 10-12 standards for each content area. Once the standards were established, three to five benchmarks per grade level would be identified for each standard. A final step in the process would be to identify the assessments tied to the benchmarks. Ultimately, the results of these assessments would enable teachers and administrators to analyze students' strengths and weaknesses in light of the established benchmarks.

The Director, looked forward to having standards and benchmarks documents that were "living" documents used to foster discussion about student learning. The Director characterized the Reading/Language Arts content area team as leading the way with this task, and during our interview, reflected upon a meeting earlier in the day with the reading language arts team. The Director believed that those on the team clearly understood the importance of the newly established benchmarks. "They know each benchmark is something we are going to assess," the Director said. The Director believed that establishing common assessments linked to common benchmarks would create a structure that fostered both accountability and consistency across the district.

The Assistant Director also looked forward to the documents created from the benchmarking process and believed this work would make the district a "truly standards-based district" in which the benchmarks are delineated and students are assessed according to the benchmarks. Like the Superintendent, the Assistant Director was conscious of the variance across the district, a variance that the Assistant Director attributed to a lack of clear standards and benchmarks. "We have a lot of variance in all

of our curricular areas because we really haven't had clearly defined benchmarks that we all utilize as a way to unify and organize what our kids learn from one side of town to another, or from one elementary building to another," the Assistant Director said.

The Assistant Director further predicted that a system of standards and benchmarks would promote a better system for feedback. With such a system, the teacher would receive feedback on where students are in relation to the standards and the students would receive feedback on their achievement of the benchmarks as well. Previously, the district had not assessed students to standards. "We said we did, but we didn't have the assessments in place to do that," the Assistant Director said.

Some of the variance across the district was attributed to past practice of site-based management which led to an undesirable level of decentralization. Now, it was time to go from "a district of schools to a school district," said the Assistant Director. The MAP appeared to play a key role in orchestrating that shift.

Measures of Academic Progress

As discussed earlier, MAP was a computer adaptive test that was administered to students in Grades 2-10 every Fall and Spring. It had been piloted the previous year and was in its first full year of implementation at the time of this study. Both the Superintendent and the Director had experience with this assessment when they worked in other districts. To illustrate the point, the Director showed the researcher a set of documents from the Director's previous district as an example of how MAP scores had been used to identify achievement patterns within that district. For example, the Director had been able to compare actual growth to expected growth and analyze the gap between the two data points for each grade level across the district, as well as between grades

within a single school. These patterns were used to foster collaborative conversations about what might have been done in one school that hadn't been done in another.

Naturally, the Director was planning to foster those types of conversations in Valley View but recognized that not everyone was ready for in-depth data analysis and data-driven decision-making. The Director was aware that some elementary staff had expressed concerns that students as young as second grade were not ready to take a computer-based assessment like MAP. However, the Director believed their argument didn't have much merit since the MAP already had nationally-normed scores for students of that age. "Philosophically, it's hard for elementary people to adjust because they are not used to getting that kind of data; they're not used to using standardized data like MAP," the Director said. A shift in the culture would be needed and the Director was working to make this happen.

Steps to cultivate a data-driven culture, started with the building principals. "I don't think the principals here have been asked to be instructional principals much in the past," the Director said. The Director had the impression that principals were given their data and then they would pull out of it what was the most meaningful to them. As a result, Valley View's principals' goals "were all over the place and not necessarily tied to reading or math." The Director noted that some of the principals' goals were operational goals such as adding playground equipment. Consequently, the Director established new expectations for goal-setting which were explained as follows:

> We ask them to have the over-riding goal be linked to WKCE. The MAP testing is the progress-monitoring goal, and then they have an action plan. It's like we need to say, this is what we are tight on. You can be loose about how you get there. And if somebody has some good success, then of course we will share the wealth.

Of course, it's one thing to expect principals to establish achievement goals and another thing to hold them accountable for achieving those goals. That's where supervision of principals became a key factor.

Supervision of Principals

A significant part of the responsibility for moving Valley View in the direction established by the strategic plan fell on the shoulders of the building principals who were responsible for implementing the Professional Learning Community concept within their respective buildings. (Recall that the purpose of the PLCs was to provide time for teachers to work collaboratively to create common assessments based on standards, analyze data, and make instructional decisions based upon the data.) The Superintendent characterized the principals as a group of men and women having a wide variety of strengths in many different areas. Although the Superintendent believed that having a variety of strengths was a powerful thing, the Superintendent also recognized that principals would need a core set of competencies in order to move the organization forward. This included the need for all principals to be on board with the PLC concept. which would take some time—perhaps five years for a district the size of Valley View, the Superintendent estimated. "It's awesome in some places already; others are really struggling and trying to grow. Administrators are different in how they try to operate through the system, so it's not a linear path, that's for sure," the Superintendent said.

While acknowledging the differences among principals, the Superintendent's expectations for change were clear because they came with increased expectations for accountability from the principals. First of all, general expectations were changing because the district was shifting from a system with a site-based management focus

where principals had been left more on their own to a system that was more centralized in focus. "We are a district of schools, we want to be a school district," the Superintendent said. This comment echoes a comment made by the Assistant Director cited earlier.

Specific expectations for principals where shaped by the goal-setting that the principals were required to do. Although principals had been required to set goals in the past, some of their goals had been unrelated to student achievement. With the increased focus on student learning, goals had to have an achievement focus and they were addressed as part of each principal's annual evaluation. These goals appeared to be quite important to the principals because, without being prompted, the principals who participated in this study referred their goals during their respective interviews. The principals had already presented their goals to the Valley View Board of Education in Fall and knew they would be required to report on the achievement of their goals at a June board meeting.

In summary, the three district administrators who participated in this study had similar views on the purpose of assessment. For them, the primary purpose was to guide instruction and to track academic growth. However, all acknowledged the public accountability aspect of assessment as measured by the WKCE and looked forward to closing some achievement gaps between schools and between groups of students. They believed the district-wide steps that were outlined in the long-range strategic plan and operationalized through the implementation of PLCs would serve to promote proficiency for all students. Prior experience with MAP led them to believe that the results from this assessment would be valuable for guiding PLC conversations about using data to inform instruction.

Reading had been identified as the district's curriculum focus for the year. This appeared to be in reaction to concerns about the district reading program that had been expressed by some parents, as well as by the fact that district administrators had recognized that there was inconsistencies in reading instruction and assessment across the elementary schools. Reading-related changes at the district level focused on revising the district standards and benchmarks for reading/language arts in light of the state standards. Eventually, assessments would be developed to measure the achievement of the identified benchmarks. Plans were also underway to select a core text for use across the elementary schools within the district. Although the planned changes challenged some long-standing practices in reading instruction, the district level administrative team was determined to bring about change.

The district administrators recognized that change would take time and that the district teachers varied in their understanding of the purpose of assessment and the collaborative practices inherent in Professional Learning Communities. They also recognized that building principals had varying levels of understanding in the same areas. However, they had high expectations that the building principals would lead their respective schools forward. In fact, Superintendent had refined the supervision plan to ensure greater accountability from the building principals in regard to their building goals.

As we leave this section, it's important to remember that the initiatives launched at the district level established a structure for moving the district forward. To recap: the district established reading as a focus for the year, formally implemented MAP, and established a calendar that accommodated weekly early release days. District

administrators facilitated the reading curriculum alignment and the reading textbook

adoption process. The district also formalized expectations for goal-setting by requiring

principals to use a standard format for their building goals. Although district level factors

were critical in establishing a framework for change, none of these factors ensured

implementation of the district vision across the buildings. That work fell to the building

principals. The following section highlights their work and their challenges.

The Building Principals

As noted earlier, district administrators were aware that principals varied in their

understanding of the PLC concept as well as in their skills as managers and/or

instructional leaders. Those variables aside, it appears there were other factors at work

that influenced the way each building principal operated within his or her unique context.

These factors include: the principal's experiences as a teacher and principal; the overall

achievement level of the students; the role of the building reading specialist in the goal-

setting process; and the staff's skill level in terms of data analysis, goal setting, and

instructional decision-making. These factors varied from building to building. One

consistent factor was the building principal's perspectives on the purpose of assessment.

The Purpose of Assessment

When it came to the purpose of assessment, the building principals' perspective

was similar to that of the district administrators. All principals spoke about assessment as

an on-going practice that accompanies instruction. They viewed assessment as the

process of gathering information about students' mastery of a concept and subsequently

using the information to determine the pace of instruction. Like the district level

administrators, the principals differentiated the term *assessment* from the concept of a

final exam. As the Oakdale principal remarked, "In the past, assessment was final tests but it's not that way anymore. It's the daily assessments you give to see if they [the students] are catching on. It's so much more important that the teacher knows where the students are and if they are learning what the teacher is teaching," the principal said.

This sentiment was echoed by the Woodland principal: "I have never been on the same page with teachers who do assessment for the sake of giving a test. How are you using this to make this kid smarter or better or learn to use it in another context? That's how I see it. I can't see it any other way", the Woodland principal said.

Perhaps, the Maple Grove principal summed up the principals' attitude best, "What you see in the classroom everyday with your assessments, is to me the most important and authentic, second is MAP testing, and third is the WKCE," the Maple Grove principal said.

Despite the principals' view of assessments as a key part of daily instruction, all three paid close attention to the results from more formal measures of student achievement. These measures included the Wisconsin Knowledge and Concepts Exam (WKCE), the Measures of Academic Progress (MAP) and, to a lesser degree, the mandatory district level reading assessment. The results of the WKCE and MAP assessments influenced building goals, PLC agendas, and some staffing and purchasing decisions.

The WKCE

WKCE data mattered to the building principals. When results were released, the principals compared their current building data to data from previous years. Like the district level administrators, the principals looked for gains in the number of students who

were proficient and advanced, and they were conscious of where their respective schools stood in comparison to the other district elementary schools. Naturally, all the administrators wanted to see growth, not decline; however, whereas the district level administrators looked at the relative gains or declines from a district perspective, the building principals took their numbers personally.

For example, when asked about the WKCE, the Maple Grove principal had the current set of WKCE data readily available. "You are going to think I am a freak but before Turnleaf [on-line proficiency reports] came out, our district gave us these printouts so I went home and did it myself. I stayed up until 2 in the morning on the day I got them so that I could get [the results] to the staff."

In reflecting upon the results, the Maple Grove principal remarked, "I will be honest with you; I am a little bit sad at our test scores." The principal was pleased to see that math scores had gained over the last three years, but was unhappy with a 3 to 6 point dip in the reading scores. Although the scores were well above the state benchmark, they were still lower than what that principal had hoped for. The Maple Grove principal attributed the gains in math to the fact that the building had focused on that area. In response to the scores, the principal vowed to find more time for focusing on reading. Overall, the principal was pleased with the way the staff responded to the data they were given on the morning of an early release day. "And guess what, by the time I was walking around to the meetings, they had the data out and were going through it and were saying, *what are we going to do for these kids*."

The Oakdale principal spoke about the WKCE results in the context of what the results meant for students. This principal viewed WKCE scores from the perspective of

knowing that Oakdale was one of the higher performing elementary schools in the district. However, despite that standing, the principal felt that WKCE scores represented a bare minimum standard, noting that a student might score in the proficient range but rank in the 25th percentile on the test. For that reason, scores were not especially meaningful. "It's not very high to be proficient," the principal said. I'd rather have them all advanced or in the upper 90 percent instead of proficient or advanced." Despite these high expectations, the Oakdale principal also recognized that some of Oakdale's students struggled academically and would continue to struggle on the state assessment. The principal paid close attention to the scores of these students and looked to see if they grew from minimal to basic, or if they increased their scores at the basic level. "They didn't get to proficient but they went up so that's one thing I always look at," the principal said.

In terms of analyzing and working with their data, the Maple Grove and Oakdale principals had several advantages. First, each had been principal in their respective buildings for at least five years. They had a history with their students and their staff. Furthermore, each had a building reading specialist who was experienced and who had been assigned to the building for a number of years. The specialists' experience level and history with the staff paid off when it came to assisting the principals at data retreats, goal-setting, and putting improvement plans into action. By understanding how the data retreat process worked in each building, one can appreciate the mix of factors that influenced the process.

As mentioned earlier, teams from each building participated in a data retreat each year. These retreats, which were held in the summer, had been held annually for more

than five years. Teachers participated on a volunteer basis. The only new aspect of the process was that the Director had standardized the goal-setting format. This standardization ensured that building goals would be focused solely on student achievement factors.

The Oakdale principal spoke about the data retreat process as something that was a well-established routine at the school. According to the principal, the school team saw its role as one of identifying strengths and weaknesses and then bringing that information back to the entire staff for review. In reflecting upon the results of the process, the principal said that over the years the Oakdale staff always agreed with the data team's analysis and then worked together to identify strategies to incorporate into their building goal. This teacher ownership of the data and the goals was important to the Oakdale principal. "If they come up with the whole plan because they saw the data and then came up with the strategies, then it becomes their idea about what they are going to teach in their class," the principal said.

The Maple Grove principal used a similar approach to summer data analysis and goal-setting. Likewise, it seemed the Maple Grove teachers had ownership of the data and the building goals. In fact, without prompting, the teachers from Oakdale and Maple Grove who participated in the study spoke about specific work they were doing with students as related to their buildings' goals. This will be discussed in detail later.

It is worth noting that both the reading specialists from Oakdale and Maple Grove were actively involved in the data retreat process and took leadership roles in analyzing the data. This is an important factor because neither the Oakdale principal nor the Maple Grove principal had taught elementary reading before becoming elementary principals.

Consequently, as evidenced by the comments they made about their respective reading specialists, both trusted and relied upon the expertise these reading specialists brought to the task of data analysis and goal setting. Furthermore, the two principals counted on their reading specialists to help teachers design and implement strategies related to the building's reading goals.

The Woodland principal was also eager to see the school's achievement scores improve. Like the other principals, the Woodland principal took ownership of the data and took personal responsibility for understanding it. Understanding the data was especially critical for the Woodland principal for two reasons. First of all, the principal could not rely on the Woodland reading specialist as a full partner in the task because the specialist was relatively new to the building. In addition, most of the members on the school's data retreat team were not classroom teachers and did not teach reading. Consequently, the team members were less sure of how to interpret scores and make recommendations based upon their analysis. The team members needed coaching from the principal on how to get through the process effectively. For this reason, the Woodland building principal recognized the need to be one step ahead of the data team. "I have to know what it [the data] is first, otherwise I don't know if they are on the right track or not. Eventually, I hope that they get to the point where I don't have to sit down ahead of time; we do it together with everybody coming out with the same understanding," the principal said.

Fortunately, the Woodland principal had a background that included experience with teaching reading at the elementary level, as well as previous experience with data-

analysis and goal-setting. The principal had routinely coached teachers in the process in other schools.

Like the Maple Grove and Oakdale principals, the Woodland principal linked scores to individual students, paying close attention to the lowest performing students, listing them by their WKCE proficiency level and color-coding them to see patterns at a glance. "The kids that I am worried about here are the ones who also came up on MAP testing and any other assessments they are doing in their classrooms," the principal said. The Woodland principal also analyzed the data to see how long the low performing students had been in the district. Upon discovering that most of the students had been in the district since four-year old kindergarten (4K), the principal reported telling the staff, "We have an educational debt to these students. We have had them since 4K. We don't have an excuse."

Measures of Academic Progress

MAP test data provided principals with another way to measure student achievement in their buildings. The test was administered to students in Grades 2 through 10. As mentioned earlier, the MAP is a computer adaptive test that was in its first year of full implementation when this study was conducted. Although the district required schools to test students in September and then again in Spring, schools also had the option to test either all students or targeted students in January. Both the Maple Grove and Woodland principals opted to test all of their students three times while the Oakdale principal decided to stay with the Fall/Spring plan for all and test selected students in January.

The Maple Grove principal reflected on life before MAP, noting that without the MAP, WKCE data was all that was available. "MAP testing, I think, has single-handedly changed the way teachers set goals and go about it. They can't wait for data now thanks to MAP because it is real; it is in real time," the principal said. The Maple Grove principal recalled teachers' response during September testing:

> We tested them in September and you should have seen [it]. Teacher's were walking around with their little notepads seeing what the kids' scores were because they now realize that when they go in there they can figure out by strand [students strengths and weaknesses]. What was great for the teachers is that they could go, ok, my student is at 223 so he or she has most of this down over here. Now I want to go over here and see what I can do to extend the children.

The Maple Grove principal also commented on teachers' reactions to the January results: "Our building retested in reading in January and they could not wait to see which kids made their gains and which kids didn't. With our MAP testing, the strands we worked on we improved on, which was great for teachers to see. Our MAP scores were phenomenal," the principal said proudly.

The Maple Grove principal explained how the MAP scores were used to inform classroom instruction, particularly for the students who didn't make gains. For example, the principal offered the scenario of a teacher whose students had failed to gain in vocabulary. The teacher might start concentrating on developing vocabulary on a daily basis even when students are studying a subject area other than reading. "Even when they are doing social studies the teachers are thinking vocabulary. So the assessment practices have really helped our building," the Maple Grove principal concluded.

The Woodland principal also spoke positively about MAP testing. "With the MAP testing, we can look at the different strands where kids are falling down and then

we can make sure that we look at lessons and decide how we can best meet the needs of

groups of kids through flexible grouping." The Woodland principal had used MAP

results in a previous assignment and had a clear vision of how the results should be used.

For example, the principal said a teacher might look at MAP results generated at the

beginning of the year, view it as a baseline, and notice that five students are low in all

four strands. The teacher might provide additional instructional time to those students and

some remediation. On the other hand, when the teacher looked at the top scores, she

might see three students who are high and needed differentiated instruction to meet their

needs. As for students in the middle range, the teacher might see the need for flexible

grouping because one group might do well with analyzing text, but might need help with

determining meaning in context. This situation would call for differentiated learning, as

well, the principal said.

The Woodland principal viewed the MAP assessment as both formative and

summative, depending on when it was administered and how the results were used. The

principal explained the difference:

> In Fall we see it as our formative assessment, and we work, work,work,
> work, work. And after we test in December or January, then we say, ok,
> this is where we saw our deficits, we really gotta kick it up and see what
> happens in May.

Although the Woodland principal considered the May testing to be a summative

measure because it was administered toward the end of the school year, the principal was

concerned about viewing the results as summative data. "Either the students made it or

didn't make it, but what's going to happen if they don't? That's always my question. Do

they just go on to the next grade and you can forget about them? We don't want to go

there; we want to make sure that we are still thinking like this is important."

Consequently, the principal believed the Woodland staff needed more professional

development before they were fully able to use the results to differentiate instruction.

Like Maple Grove and Woodland, Oakdale administered the MAP three times a

year. For Oakdale, however, the test mainly served a summative purpose because the

school tested only selected students at mid-year. The tested students were those the

teachers felt needed a mid-year check on their academic growth. At the time this study

was conducted, the Oakdale principal was satisfied with September/May testing which

would show the full growth of students from the beginning of the year to the end. The

principal decided not to require the January testing until the Oakdale teachers really knew

what to do with the data and understood how useful it could be. Until they did, the

principal believed testing three times a year would be an issue with the Oakdale staff:

> I think it will become more and more useful as we—and I say we because
> that means me, too—learn more and more about it. I still don't know
> everything about it but I think I know more than a lot of the teachers do.
> But I don't know nearly enough about it yet, and nearly enough about all the
> scores, what everything means and how we can use all this. I don't know
> nearly what I know I can learn.

Like the other principals, the Oakdale principal valued the immediacy of MAP

results and believed that the MAP test provided much more information about "where

students are in their learning." Comparing it to the WKCE, the principal preferred MAP

because the results provided information about a child's reading level, where a child

stood in comparison to grade level peers, where a child ranked in terms of national

percentile information.

District Reading Assessment

Results from the district reading assessment were recorded in students' reading folders and the information was used to place students in instructional groups. Two of the three principals placed less value on this assessment compared to the MAP test.

In fact, the Maple Grove principal had not looked closely at the data from this particular assessment over the years. "It gets compiled and it gets on a shared drive but we haven't really used it here so my thinking is that if we have really been doing this for a long time and we've had this data, and I am a data person and I haven't used it—and and now we have other things—I just don't know how important it is." The principal was also concerned about the time it took teachers to administer this assessment because the time spent assessing came at the expense of instruction. As will be discussed later, the Maple Grove teachers who participated in this study had an entirely different perspective on the value of this assessment.

The Woodland principal called the district reading assessment a "tricky" instrument, claiming that a student could score poorly on the whole assessment due to fluency even though they understood what they were reading. The principal was also skeptical as to whether all the Woodland staff understood the assessment and how to use the results to inform instruction. This skepticism proved to be accurate, as will be discussed later.

In contrast to the Maple Grove and Woodland principals, the Oakdale principal believed that most of the Oakdale staff used the district level assessment results very well, both for classroom instructional decisions and placement decisions. The principal reported that teachers discussed the results between grade levels and shared information

about the student's reading level the previous year and what the student needed the following year. The Oakdale principal also spoke about Oakdale reading specialist's active involvement in the administration of the district level reading assessments. In describing the specialist's involvement, the principal said, "She likes the [district level assessments] and makes sure everyone uses them and uses them the way they should be and are using them properly, you know not inflating anyone's reading levels. They are used exactly like they are supposed to be used," the principal said.

The Oakdale principal valued the results from this assessment and felt that the results could be used in tandem with MAP results to provide a more complete picture of the child. For example, the principal said that the district assessment measured fluency while the MAP assessment did not. Furthermore, the principal noted that some students scored poorly on the MAP either because they did not take the assessment seriously, or because they lacked the stamina to do their best throughout the test. Consequently, the Oakdale principal perceived the district level reading assessment as something that provided another window into a child's achievement. Incidentally, as we will see later, these comments were similar to those made by the Oakdale reading specialist.

I did not find any evidence to suggest that results on the district level reading assessment were regular topics of discussion during the PLC meetings. Participants spoke about using WKCE data, MAP results, or data from common assessments. Furthermore, it seemed that even when teachers focused on the results of their common assessments, most of those common assessments were linked to weaknesses identified through the WKCE or the MAP.

Professional Learning Communities

As mentioned earlier, early release days had been added to the district calendar.
These additional staff development days were part of the structure the district had
established so that teachers would have formal time to examine data, analyze student
work, and make instructional decisions. As might be expected with any new venture,
PLC implementation varied from building to building. For example, Maple Grove
teachers had worked in PLC groups previously and appeared to be ahead of the other two
buildings in the level of sophistication with the process. The Principal and a Maple Grove
teacher had attended workshops and had visited schools that used the PLC approach.
They had worked together to implement the process at Maple Grove. The quote below
captures the complexity of what was needed to make PLCs operational at Maple Grove.
The Maple Grove principal explained it this way:

> First, we started with professional learning communities and we taught
> people how to collaborate. Then, we went into formative assessment and
> hooked teachers in by using Marzano's book [The Art and Science of
> Teaching: A Comprehensive Framework for Instruction]. It had all the data
> on what gives you the most percentile gains for students. When teachers saw
> that doing daily formative assessments and giving kids feedback would help
> students learn at the most rapid pace, they bought into it. So we would in-
> service them on a few different types of formative assessments and then we
> would give them homework, like, ok, for the next two months go back into
> your classrooms and try it and then when we meet at our next in-service date
> we are going to ask you to get up and share how it is going.

In order to elaborate upon how the PLC structure was working in Maple Grove,
the principal produced the Maple Grove building SMART Goals and PLC Team Plans.
These documents provided more detail about the type of data-driven work that was
happening at Maple Grove. What was most interesting about the documents was that they
appeared to demonstrate the varied ways grade level teams approached their work. Some

teams used MAP data as a starting point for designing instruction that was integrated into the regular curriculum while others adhered to the materials provided by the test company.

In the SMART Goals examples that follow, both grade level teams hoped to see increases in MAP scores as a result of their Action Steps. The Action Steps developed by a primary grade team followed an October to January timeline. It begins with a baseline assessment followed by whole group instruction. A second assessment identifies students who will need additional help with the concepts. It doesn't identify what will be done to extend the learning of those who were successful. The team wrote:

1. Give a Common Formative Assessment (CFA) on story mapping.
2. Within each classroom we will teach the four story elements: character, setting, problem, solution.
3. During intervention time we will focus on each element for one week.
4. After all elements have been taught, we will give a second CFA on story mapping.
5. After the second CFA, we will group students based on the story elements they have not yet mastered.

In contrast, the Action Steps planned by an upper elementary team were less specific. This plan was selected by the researcher to illustrate the ways in which two teams from the same building approached the task of data-driven instruction. This example illustrates the influence of MAP data on the teachers' work. The team wrote:

1. We will use the Descartes item strands to create lessons and assessments for flexible groups. (Descartes is an online tool that helps teachers identify instructional materials matched to students' MAP scores)
2. We will research sample questions from the MAP testing to guide instruction.

Copies of PLC Weekly Team Plans also provided a window into the data-driven actions undertaken by members of the Maple Grove staff. The portions cited below are taken from the teams' notes about the accomplishments of their meeting. The first log

entry was selected because it demonstrates how a primary grade team thought about their role in preparing students for the Grade 3WKCE. The second log entry provides a brief look at how an upper elementary team used common assessments. Each passage is taken directly from the team logs.

Primary Grade Log
Looked at WKCE 3rd Grade Testing booklet to see about types of questions and passages that students will be expected to read and answer. Also discussed how we can develop similar questions to help students develop the needed skills to answer these types of questions. We discussed vocabulary and inferential questions. We also talked about how language based games can help develop critical thinking and oral language. We are contemplating whether we should follow the format of a test having a "bubble" test as an added formative assessment.

Upper Elementary Log
Began organization of common formative assessments to be used during reading and reading intervention/extension. Know that we are starting with compare and contrast. Discussed how well/not well students did on baseline assessment. Know that there is teaching that needs to be done on that skill with all students.

By collecting various documents such as those described above, the Maple Grove principal could analyze the teachers' work, provide feedback, and coach them accordingly. The Woodland principal faced an entirely different situation because teachers at Woodland needed intensive training in the PLC model. The staff had not yet reached the point of working collaboratively during PLC time. "They just thought they got together and talked," the principal said. In response, the Woodland principal used Dufour's (1998, 2008)questions from his work on professional learning communities to guide the Woodland teachers' conversations. These questions included:

- What is it we want students to learn?
- How are we going to know if they learned it?
- What are we going to do if they didn't learn it?
- What are we going to do if they did learn it?

The Woodland principal expected those questions to guide the PLC discussions each week. "We are not going to talk about field trips, we not talking about other things. It's all about learning," the principal said. To illustrate this point, the Woodland principal recounted a staff meeting that focused on the process of goal-setting, the task of creating common assessments to monitor progress toward those goals, and the steps for responding to students who needed intervention or an extra challenge. This segment of the interview is included in its entirety because it captures the level of detail the principal used when coaching the staff. It was also selected because it illustrates how the principal's knowledge of elementary reading supported the work at hand. (Note: The phrase *extends and evaluates text* is a test category label used by both the WKCE and the MAP.)

> I said ok, for example, if our goal is extends and evaluates text, that's where we fell apart on the WKCEs, that's where we fall apart on MAP, so our kids don't know how to do that well. So I said, ok, here's the standard, so I went through what it is, and then the things they are supposed to be able to do – make connections, etc. The goal is students will extend the text to their own situation by evaluating the characters. Students were to complete a Venn diagram comparing themselves to a character in the story. After completing this they'll complete a 4-question assessment that will be collaboratively scored. So, the teachers would then sit down and say, here is the story we taught, here are the VENNS the kids produced, here are the four questions we made up to ask them. How are you and the character alike, different? What did you learn about the characters? And, why do you think the author included this character in the story? Ok, we've created a four-question common assessment which is no big deal. And so now I'm just getting them into how we collaboratively score.

Next, the Woodland principal coached the staff to reflect on the steps that might come after scoring an assessment. The principal reported using this approach to help teachers consider options for students who didn't learn, as well as how to challenge students who were successful. The principal recounted her discussion with the staff:

And then how do we respond when they don't learn? Choose a more familiar story to do the same activity—like little Red Riding Hood—something they've heard a million times. Model that Venn diagram with those struggling students with the teacher being compared so that they can help you determine it. Use our intervention calendars to spiral support for those students struggling with the concept of extending authors. And what are we going to do if they do learn? So, we've already determined that. So maybe they are going to do an advanced organizer, a tri-Venn, write themselves into the story as a character, compare and contrast that story with another story, read a higher level story to make world connections. So again, is this rocket science? No. So that was my thought in getting them to be thinking about what is it you want kids to learn, how do you know they learned it, etc.

The Woodland principal didn't foresee such an elaborate discussion every time teachers met. However, the principal expected that Woodland teachers would eventually reach the point where they automatically thought about Dufour's questions and then made decisions about what would make sense for a particular learning activity in light of those questions.

The assessment examples above illustrate the kinds of common classroom assessments that seemed to really matter to principals. Despite the fact that the principals valued typical on-going assessments used by individual teachers (i.e., running records, anecdotal notes, and observational logs), they were more interested in the assessments that showed whether or not students were making progress in areas identified for improvement based upon MAP or WKCE data. Ultimately, they were looking for common assessment data that would later correlate to an improvement in WKCE or MAP scores. These concrete data appeared to be more significant to the principals than some of the on-going data collected at the classroom level. For example, when the researcher asked one principal to comment on an observation log as a form of assessment, the principal remarked, "That's interesting but how do I know the students are learning?

In summary, much of the responsibility for moving staff forward fell to the building principals who were charged with implementing district-initiated change within the unique context of their respective buildings. Multiple factors contributed to the distinctive nature of each building's context, including: the principal's experience with elementary reading; the overall achievement level of the students within the building; the length of tenure of the building reading specialist, as well as the specialist's role in the building's goal-setting process; and the teachers' skill level in terms of data analysis, goal setting, and instructional decision-making.

Each principal was actively involved in the data analysis and goal setting process. The Oakdale and Maple Grove principals had reading specialists with long tenure in those buildings, a factor which enabled the specialists to fully assist the principals with data analysis and goal-setting. As we will see later, these specialists also helped staff design instruction and assessments related to the building goals. The leadership of these reading specialists was particularly useful to the Oakdale and Maple Grove principals who did not have experience teaching reading at the elementary level. In contrast, the Woodland principal had a reading specialist who was relatively new to the building and was becoming acclimated to the staff and their needs for reading support. Therefore, this principal facilitated much of the building goal-setting process. Fortunately, the principal's familiarity with goal setting from a previous experience as a principal, along with previous experience as a elementary teacher who had taught reading, gave this principal the skills needed to train the Woodland staff on the PLC model, the data analysis and goal-setting process, and the development of common assessments.

When it came to accountability, the principals were responsible for meeting their building goals, goals that were linked to student achievement as measured by the WKCE. They also faced the responsibility of ensuring academic achievement as measured by the MAP assessment. The principals reported the results of both these measures to the superintendent and the board of education. The report to the board made their accountability very public.

The three building principals spoke about assessment as an on-going classroom practice that goes hand in hand with instruction, and they were aware of the types of classroom assessments that were used in their respective buildings. However, no principal monitored these kinds of assessments, nor were they involved in assisting teachers in daily reading instruction. The principals relied upon on their reading specialists as teacher-leaders who had the expertise to assist teachers in ways that would advance student achievement in reading. The next section examines the role of the reading specialists in each of their respective buildings and discusses the ways in which their actions influenced the way data were used to inform reading instruction.

Reading Specialists

The reading specialists at Oakdale, Woodland, and Maple Grove had multiple responsibilities. On any given day a reading specialist might serve as an intervention specialist, a teaching partner, a literacy coach, a staff developer, and/or a program coordinator. As might be expected, assessment data guided the decisions they made regarding the ways in which they worked with staff and students. In the paragraphs that follow, I will examine the reading specialists' roles in light of the various kinds of

assessments that guided their work. These assessments include: the WKCE, MAP, district level reading assessments, and classroom assessments.

The WKCE

When the reading specialists spoke about the results from the WKCE, they referred to the data as a "snapshot" of student achievement. Nevertheless, these snapshots influenced classroom instruction because, as discussed earlier, building goals were linked to improved student achievement as measured by the WKCE. It appeared that a minimum amount of instructional time was allocated to formal test prep. As the Oakdale reading specialist explained, "We know that effective instruction and good curriculum is the best test prep, and then we know that kids need some support in learning test-taking skills, so we try to provide them with that shortly before the test." Sample passages and sample test questions published by the state were used for this purpose. The reading specialist estimated the teachers in her building spent a week or two on this kind of targeted preparation.

Although this type of targeted preparation was kept to a minimum, the reading specialists were instrumental in designing test prep activities that were integrated into classroom lessons. For example, primary grade students at Oakdale were beginning to learn about the kind of writing that would be expected of them on the test. "We start to teach them how to construct a response to a question. We know we can't leave it all to third grade. So, we took the rubric from the WKCE and tried to put it into kid language so that the kids could understand how, if they wrote a response to a question, how it would be scored," explained the Oakdale reading specialist.

Other test prep activities were integrated into guided reading. For example, when I observed a primary grade classroom at Maple Grove, students were expected to respond in writing to a passage and then write the page number where they found the information. "We are having them do that because that is getting them ready to find examples from the text when they write their responses for the WKCE," said the teacher. The teacher credited the Maple Grove reading specialist for designing this particular strategy to prepare students for the test.

The WKCE results also influenced decisions the reading specialists made about overall instruction within their respective buildings. For example, the Maple Grove specialist spoke about analyzing the results of the state-mandated third grade reading test, which had been replaced by the WKCE format in 2005. The specialist discovered that students "did beautifully on literature but bombed on the non-fiction passages." In response to those results, the specialist not only wrote a grant but also used a portion of the reading budget to buy non-fiction materials for the school's book room. Furthermore, the Maple Grove specialist created non-fiction lessons for guided reading and read alouds, and added non-fiction materials to the classroom libraries. This emphasis on non-fiction continued in response to needs identified through the WKCE results.

The Oakdale specialist had made similar decisions in order to improve reading achievement at Oakdale. The specialist had enlarged the collection of non-fiction materials and developed lessons to help students master the skills they needed in this area. The importance of non-fiction was evident when I observed two classrooms in which the Oakdale reading specialist was teaming with the classroom teachers on the teaching of non-fiction reading skills.

Preparing students for the WKCE appeared to be a routine task for the experienced specialists at Oakdale and Maple Grove. In fact, when I showed the Maple Grove reading specialist a page from the long range strategic plan that referred to using data to guide instructional decisions, the reading specialist shrugged, and remarked, "In terms of data-driven instruction, I feel that's what we have always been working on and will continue to work on. It isn't something new."

Speaking from her experience as a reading specialist and as a former classroom teacher in the district, the Woodland specialist's comment was similar: "When it says schools and teachers use readily available achievement data, we do that. That's how we plan for balanced literacy instruction or at least it's how we are supposed to plan for balanced literacy instruction.". After looking at other parts of the strategic plan, the Woodland specialist added: "I think it's a bigger picture that just classroom assessment. But we have had data retreats where we get together with groups of teachers to plan building goals for many years."

This history of using WKCE data may be why the reading specialists spoke in such a matter of fact way about the WKCE. Their attitude was straightforward: we identify weak areas and determine what we will do about it. What is interesting is the way their tone changed when they spoke about the MAP assessment. The change in tone may have been influenced by the fact that all three of the reading specialists had varying levels of comfort with MAP and were still trying to figure out how those results might be useful in guiding instruction. Nevertheless, their reactions to the test are important and will be discussed below.

<u>Measures of Academic Progress</u>

As mentioned earlier, the Measures of Academic Progress or MAP assessment was in its first full year of implementation within the district. Unlike the WKCE, which reported student achievement in broad terms, the MAP assessment gave teachers more detailed information about students' strengths and weaknesses in four areas: vocabulary, comprehension, analyzing text, and extending text. Teachers could easily see where their students ranked in each of these areas which enabled them to design targeted intervention or enrichment.

One can argue that MAP challenged the status quo because it was a standardized instrument that provided results in real time. This meant that teachers were expected to base instructional decisions and set goals for their current group of students based on MAP results. Additional rounds of MAP testing told them whether their interventions were successful.

Of the three reading specialists, the Maple Grove reading specialist appeared to be the most comfortable with MAP testing. That may be because Maple Grove was an early adopter of the PLC concept. The principal and staff had learned about MAP during their visits to other PLC schools and had supported the use of this test in the district. It is also important to remember that the Maple Grove principal was a proponent of MAP; the principal said that MAP testing had "single-handedly changed the way teachers set goals and go about it. They can't wait for data now thanks to MAP because it is real; it is in real time."

This perspective was confirmed by the Maple Grove reading specialist. The specialist believed Maple Grove teachers appreciated the MAP because the results helped

them assess how well their students were doing in relation to skills identified as weak on the WKCE. Furthermore, the MAP results provided feedback to the teachers on whether or not their instructional interventions were successful. "It also gives you a list of skills for each range—the skills the student is secure in, the skills to be reinforced, and the skills to be taught and introduced. So you can get a good gauge of where you want to go next with this child," explained the Maple Grove reading specialist. The Maple Grove reading specialist did not express any reservations about the MAP assessment. Since this specialist was the most reticent of all participants who were interviewed, it may be that the specialist decided not to comment. In contrast, the other reading specialists spoke about the MAP in greater detail.

The Woodland reading specialist spoke at length about the MAP. "We piloted it last year but this is the first year where teachers started studying their scores and figuring out what it meant," the reading specialist said. Woodland teachers examined the scores in order to determine growth, as well as to make decisions about appropriate instruction for the students who struggled, as well as for those who were high achieving students. "During our PLC time, especially when we have fresh scores, they are using the data to form groups, to determine which areas of math and reading their kids are struggling in and how to provide intervention for that," Woodland reading specialist said.

Overall, however, the Woodland reading specialist said that the staff was still learning about how to read the various reports that accompanied the test results. The specialist characterized the reports as "overwhelming." The specialist was still learning how to read the various reports and how they could be used to guide instructional

decisions. The Woodland reading specialist was also skeptical about the usefulness of the

MAP results and also questioned the validity of the test for younger students.

First of all, the Woodland reading specialist was cautious about relying too much

on the MAP results of second grade students who were the youngest students assessed on

this test. The specialist not only wondered if all second grade students had the computer

skills and the stamina needed for the test, but also wondered if the test truly provided

information about the comprehension levels of such young students. "When I look at

those scores for 2nd graders and I see that they are struggling with extending and

evaluating text, the first question that comes to mind is why did they get that score? Is it

because they have a problem with extending and evaluating, or is their fluency so poor

that they do not comprehend?" remarked the Woodland reading specialist.

The Woodland reading specialist felt the test did not give enough information

about those aspects of reading. Consequently, the specialist found it hard to depend on

the results from this one measure to accurately assess the needs of second graders as well

as third graders who struggled with reading. In fact, the specialist believed strongly that

MAP should not be the only formal assessment used to make decisions about students'

reading ability and whether or not they needed further intervention. "I think we have to

look at multiple types of data, we have to triangulate the data," the Woodland reading

specialist said.

The Woodland specialist approached the triangulation process by examining the

scores of students who had been targeted for intervention the previous year. The

specialist compared the students' district reading assessment scores from the previous

year with the current Fall MAP scores. "If they struggled last year and they struggled

again this Fall on the MAP, then that's a student we really want some more information

on so we can best support their learning," the specialist said. "Is it because of a

comprehension issues, or do we need to give them another district level reading

assessment to discover what problems they are still having with fluency and decoding?"

the specialist asked. (It is interesting to note that the students who struggled the previous

January but scored well on the Fall MAP still stayed on the specialist's radar. The

specialist was determined to find out if these students truly made progress or if the MAP

assessment was not valid for some reason.)

In the Woodland reading specialist's mind, the district level reading assessment

could provide information that the MAP could not. The specialist characterized the

district level reading assessment as "incredibly valuable for planning" because an

individually administered benchmark assessment like that provided information about a

student's word level strategies and fluency. "It's not that it is the end all and be all either

but it is something to give you information about what they are encountering difficulty

with and what they are doing when they are encountering difficulty," the specialist said.

When it came to younger students, the level of specificity provided by the district level

reading assessment was more valuable to the reading specialist than MAP results. The

Woodland reading specialist commented:

> I think it does not make up for sitting next to a child and listening to them,
> watching their behaviors. I don't want to sound like I'm anti-MAP. I
> mean, I think there are some things in there that are going to provide us
> with a lot of information, but my fear is that people are gonna like the
> ease of just sticking their kid on a computer and be like, "great it's gonna
> tell me everything I need to know." I don't want people to start thinking
> that this is an assessment that can override those really important
> classroom assessments they give and I think that's my biggest fear.

The Woodland reading specialist was more confident about the MAP results for 4th and 5th grade students because the specialist believed most upper elementary students were capable decoders who read with acceptable fluency. Consequently the specialist felt more comfortable in accepting the MAP scores as a relatively accurate measure of the students' comprehension abilities. Still, the specialist had concerns that the MAP results were not influencing the core curriculum but were only affecting the interventions students were receiving. "I don't know if the answer is just intervene, intervene, intervene, all day long. I think what can happen is we fall into intervention, intervention, intervention, and it is wearing on our kids," the Woodland reading specialist said.

Before moving on, it's important to take a moment to compare the comment of the Woodland reading specialist with those of her building principal. Some of the reading specialist's skepticism about MAP might be attributed to the specialist's status as a relatively new member of the Woodland staff, as well as the specialist's lack of experience in interpreting MAP reports and then using the results to inform instruction. The specialist also wondered about the implications MAP might have for the reading specialist's role, a role the Woodland specialist believed was to support teachers in the classroom and provide interventions to targeted students. The specialist wondered if MAP meant reading specialists would be providing more intervention than classroom support. All of these concerns existed in an environment in which the building principal valued MAP.

As mentioned earlier, the Woodland principal had a high level of comforting with interpreting and using MAP data. The principal viewed MAP as a formative measure that could guide instruction between each administration of the test and opted to test all

Woodland students in Fall, Winter and Spring. "With the MAP testing we can look at the different strands where kids are falling down and then we can make sure that we look at lessons and decide how we can best meet the needs of groups of kids through flexible grouping," the principal said. Furthermore, the Woodland principal viewed MAP as more useful than the district level reading assessment because the principal believed teachers struggled in using the district level assessment results to make instructional decisions.

I don't want to leave the impression that there was tension between the Woodland principal and the Woodland reading specialist. This was not the case. The reading specialist acknowledged that the principal had experience with MAP and saw the potential for the test to give immediate feedback to teachers and students. The point I want to make is that the Woodland principal could not rely on the reading specialist as a full partner who could assist with the MAP implementation. Instead the principal had to coach the reading specialist while trying to move the rest of the staff along. Fortunately, both the principal and the specialist valued the concept of "triangulating" data. This factor gave them a common ground for their conversations surrounding the school's reading data.

A different dynamic existed at Oakdale. Like the Woodland reading specialist, the Oakdale reading specialist was also tentative about the usefulness of the MAP results. "To be honest, all we've used this for so far is basically a number—comparing how they did to other kids", said the Oakdale reading specialist. "There's a possibility that it could help us in the upper grades, although it's still yet to be determined if it's really going to be able to be an effective tool to help inform our instruction. It may help document progress, but I'm not sure that it is going to be easily used [to help us] identify exactly

what we need to do next with students. I don't know that is going to be easily done, although maybe it will just take us some time to figure that out," the reading specialist said.

One can't help but wonder if the comments above were influenced by factors other than the nature of the assessment. Although it may well be that the reading specialist's "wait and see attitude" was driven by concern over the usefulness of MAP, political factors may have influenced the reading specialist's response to this new district-wide practice. The specialist appeared to feel left out of the conversations surrounding the MAP implementation, remarking that in the past reading specialists would have had a voice about matters such as this. The specialist was trying to have a voice in the current situation but didn't feel that the voice was heard as much as it had been in the past. "I think it has to do with, I'd say, certain administrators and perhaps they already have some preconceived notions of the way they think things should go, and that's the way they are going," the reading specialist said with a sigh.

The Oakdale specialist also shared the Woodland specialist's perspective about the test's usefulness with younger students. Referring to students in grades two and three, the Oakdale specialist said, **"**I am very skeptical because I felt like there were so many things about it that did not necessarily seem to be able to give us a really accurate assessment of what a child could do." The Oakdale reading specialist's concerns ranged from the format on the screen, which the specialist considered to be small print, to the difficulty a student might have in reading directions. For example, the Oakdale reading specialist did not consider it to be a reliable assessment when a child could not read the directions which asked the child to identify which objects start with the letter B. "They

can tell you the beginning sound but they can't read those directions yet," the specialist said. The Oakdale reading specialist also noted that the test did not provide information about students decoding skills or their fluency. This opinion of the computer-adaptive assessment was further reinforced by the fact that there could be a discrepancy between a student's score on the MAP test and the same student's performance in the classroom, as well as on the district level reading assessment.

Despite the fact that the Oakdale specialist felt the usefulness of the results was limited, the specialist made some instructional decisions from observing how students interacted with the computer-adaptive test. "When we saw that some kids had performed really low—I mean we would have picked out those kids anyway probably—but it really helped us realize oh, oh, the test [WKCE] is coming soon and they are not very good test-takers, as well as that they may be struggling readers, " the specialist said. In observing these students, the reading specialist noticed a lack of stamina as well as poor test-taking skills. In response, the Oakdale reading specialist decided the test prep for the WKCE should be done in small groups for some students. "They would get more out of it if I was sitting here and there were just four of them around me and I was explaining what they need to know about how to take a test as opposed to being one of 22 and in the back of the room and tuning out very easily and not getting the benefit of that instruction," the Oakdale reading specialist said.

Although the Oakdale principal spoke positively about MAP and believed the assessment would become more useful as teachers learned how to use the results, the principal was a proponent of the district level reading assessment, too. "You get good information from MAP, I've seen that already, but, I like the extra stuff we get out of the

district level reading assessments that we don't get out of MAP." Like the Oakdale reading specialist, the Oakdale principal noted that MAP did not provide information about a child's fluency, and that good information could be gained from the one-on-one district level assessment that was administered at the primary grades. The principal also believed the computerized nature of MAP put some younger students at a disadvantage because it was tough for them to sit and read the screen. Overall, he believed that the two measures should be used together to provide an understanding of the child's needs.

Without asking them, it's impossible to know if the Oakdale reading specialist and principal arrived at their conclusions about the MAP independently or if one person's opinion influenced the other's opinion. Although the principal seemed far more positive about the MAP compared to the reading specialist, it is interesting to note how their comments about the use of the test with younger students mirrored one another.

District Level Reading Assessments

As noted above, the Oakdale reading specialist placed a high value on the information generated by the district level reading assessment and was a big proponent of this assessment. (The district level assessment was administered in Fall to new students and to students who were identified as below the standard during the testing conducted the previous Spring.) "We get the information we need about decoding and fluency, as well as specific comprehension information and we use that pretty much to identify who needs additional intervention," the Oakdale reading specialist said.

Here, it might be helpful to recap the nature of the district level reading assessment. As prescribed by the district, primary students were assessed every January with an individually administered benchmark assessment. Fourth and fifth grade students

were assessed with a group administered version of the QRI (Leslie and Caldwell, 2001). These assessments were re-administered to students who fell below the expected standard in May to determine if progress had been made. Students who were targeted in January and May were also reassessed in October to help determine instructional focus for the classroom as well as intervention needs. New students were tested in September. Primary students read a passage orally while the teacher took a running record to determine their accuracy and analyze their errors to determine their accuracy and self-correction rate. The test was also timed for fluency and a rubric was used to assess the quality of fluency in terms of expression and phrasing. In addition, the teachers also scored students' retelling with a retelling rubric. Students were expected to meet a predetermined standard in all five areas. If not, a student was tested at progressively lower areas until he or she could meet all five standards.

The data from the district level assessment strongly influenced instruction at Oakdale, especially at the primary level, but also in Grades 4 and 5. (As a matter of fact, one of the upper elementary grade teachers who participated in this study told me that the teachers on this particular upper elementary team routinely administered the district level to all their grade students each Fall.) The Oakdale specialist believed the district level reading assessment results would continue to influence instruction for the younger students. "It's possible we could discontinue our 4th and 5th grade district level reading assessments if we can figure out how to use MAP well enough," the Oakdale reading specialist said. However, based upon the reading specialist's remarks it seemed that until Oakdale teachers gained confidence about the reliability of MAP, and gained a better understanding of how to use results to guide instruction, it appeared that classroom

instruction at Oakdale would be more heavily influenced by the results from the district level reading assessments and from on-going classroom assessments. It is worth considering that the above remarks may have reflected the reading specialist's perspective on the Oakdale staff in general. As will be discussed later, one Oakdale upper elementary teacher who participated in this study was enthusiastic about MAP and was using the results to guide some the teacher's instructional decisions.

Like the Oakdale specialist, the Woodland specialist also appeared to think that district level reading assessments and on-going classroom assessments currently had a greater influence over teachers' decisions than the MAP did. The Woodland reading specialist said that although the district did not require teachers to administer the district reading assessment in September, some Woodland teachers administered the assessment anyway. Some gave the assessment to students who had been previously targeted as struggling; others administered it to all students in a class. The reading specialist explained that during the month of September when teachers are getting to know their students, "the information that their assessment tells them is crucial to helping them determine the direction they are going to take with instruction. It helps them pinpoint exactly what their instructional focus should be," the Woodland reading specialist said.

According to the reading specialist, some Woodland teachers used the tool periodically year-round. "We have teachers that all semester long will sometimes pull a kid out and do an assessment on them just to see what's happening. And certainly, they don't do that with everyone because it's time-consuming, but you know, people will pull them out throughout the semester," the Woodland specialist said. As one might expect, experienced teachers used the district reading assessment more than required because

they had the ability to use the data to guide their instruction. The Woodland reading

specialist assisted less experienced teachers by administering the test to new students and

previously targeted students and then showed the teachers how to use the data to inform

their instructional decisions. The regular use of the district level assessments was

especially prevalent with Woodland's primary teachers. In contrast, the reading specialist

believed that Woodland's 4th and 5th grade teachers were less likely to make regular use

of the district level reading assessment. The specialist spoke about the reasons why this

might be the case:

> I don't think the 4th and 5th grade teachers use the district level assessments
> to plan their instruction as much as they use MAP. And, I think it's that
> they are not quite sure how to use the information [from the district level
> assessment]. It's kind of a cumbersome assessment because you are giving
> all these passages with all these questions. Because some of [the students]
> are taking these tests at different levels, you then have this pile of written
> work. [It is difficult] trying to sort through and think what the answers say
> about their comprehension and what the teacher needs to teach to improve
> that comprehension.

Unlike the other two reading specialists, the Maple Grove reading specialist

did not speak at length about the district level assessments, however, the two Maple

Grove teachers who participated in this study placed a lot of value on the district

level reading assessments. Their views will be discussed later in this chapter. All

three specialists spoke about the importance of on-going classroom assessments, a

topic which will be discussed in the following paragraphs.

Classroom Assessments

Like their building principals, the three reading specialists who participated in this

study had a high regard for the value of on-going classroom assessments. Some of the

assessments they spoke about were done, in the words of one specialist, "on the fly."

These assessments included observations, running records (where appropriate to the grade level) and anecdotal notes. Other assessments captured performance through checklists or rubrics. Still others called for some kind of written response.

In terms of content, some of the classroom assessments measured a skill that was being incorporated into the curriculum because that skill had been identified as a weakness on the WKCE or MAP. Other assessments focused on skills that teachers believed were important but that were not assessed by other measures.

The Woodland reading specialist's comments help one understand which strategies might be selected for on-going classroom assessments. The specialist explained that basic comprehension strategies aligned with the strands on the WKCE and the MAP. Consequently, when students were reasonably successful with a particular strategy, that strategy wouldn't be assessed as much. However, where students struggled—for example, inferential thinking was always an area of concern—that strategy would most likely be a regular focus of classroom instruction and assessment.

Despite the influence of the WKCE and MAP, the Woodland reading specialist believed the strategies that aligned with the genre of the book influenced instruction more than the data generated by the WKCE or MAP. The Woodland reading specialist elaborated upon her thinking:

> It's a bit more like, when you are reading this kind of book, what kind of thinking are you going to be required to use in order to understand it? Instead of saying, well, our test said we have to work on inference so we are just going to focus on inference. You want it to be more authentic than that. I guess the way I look at it is, if we see in the test that they are struggling on something, we kind of increase instruction in that area or make it a little more explicit, or support it more. But we don't let the assessment drive what we are going to teach because we already know what that is.

No matter the focus of an assessment, it appeared that staff development played a key role in supporting teachers with their classroom assessments. Some staff development was a scheduled event; other staff development came about naturally during collaboration between the reading specialist and teachers in the building.

Collaboration

Of the three reading specialists who participated in this study, the Oakdale reading specialist was the most enthusiastic about the Oakdale teachers'use of classroom assessments. The reading specialist commented on how teachers at Oakdale approached assessment:

> Our teachers can't live without it. It's part of their daily routine, in one way or another, to gather information, to decide how this child is doing; is this child coming along; what do I need to do next for each of them. I feel really great that our teachers really believe in differentiated instruction and meeting the needs of individual students and you can't do that without on-going assessment.

Here it is helpful to think about the Oakdale specialist's comments in light of this particular individual's professional status and years of experience. This reading specialist was highly regarded by the building principal and staff. The specialist's length of tenure in the building had given the specialist time to develop collegial relationships. These relationships paved the way for professional collaboration around best literacy practices and also enabled the specialist to coach the Oakdale staff, many of whom were experienced teachers.

Although the Oakdale reading specialist did not have the positional authority of the principal, the specialist was the instructional leader in her building when it came to reading. "I have a particular vision about what I think is effective and I have learned

certain things over the years that I think are good and that we should try them," the Oakdale reading specialist said.

That's not to say the Oakdale reading specialist forced ideas upon the staff. The specialist recognized the importance of collaborative relationships and spoke about approaching teachers with an idea and then asking them their thoughts about whether the concept would work. "We figure it out together. It's not me saying this is what you need to do at all," the Oakdale reading specialist said. However, what this collaborative process demonstrated is that the Oakdale staff was willingly worked together to design assessments that measured aspects of reading they considered important. (Incidentally, the Oakdale principal valued this kind of collaboration around creating assessments. "I think it is so much more important [compared to knowing a student's score] that the teacher knows where the students are and if they are learning what the teacher is teaching. Teachers talking to each other within their grade levels and outside their grade levels so they can learn from each other on what's working and what's not and where they can go—that's important," the Oakdale principal said.

We can understand how this collaborative process worked by taking a look at the classroom assessments that accompanied the reading partnership approach used at Oakdale. The partnership approach was being used by the Oakdale teachers who participated in the study. According to the Oakdale reading specialist, teachers at the school had initially implemented literature circle groups consisting of four to eight students. To their surprise, this group configuration did not meet their expectations. Teachers believed that students were not developing the desired discussion skills because the group size limited the opportunity of individual students to participate. Furthermore,

teachers believed that this lack of participation was detrimental to students'

comprehension and that students' comprehension suffered because there were no

structures in place to ensure that they adequately processed the text as they prepared for

the discussion.

"We found that there were some kids over the course of the year who were very

good at staying under the radar because they would just sit back and not participate much.

Then, all of the sudden, at the end of the year we realized that their comprehension was

not very strong," the Oakdale reading specialist said. As a remedy, the reading specialist

and the teachers implemented reading partnerships, which were small discussion groups

consisting of two to three students. They believed the smaller group size would give all in

the group the opportunity to participate in sustained discussion, the kind of discussion

that would foster greater comprehension.

In addition, the Oakdale reading specialist and the Oakdale teachers

collaboratively devised assessments to help them determine whether or not students were

internalizing the comprehension skills that would enable the students to fully participate

in their reading partnerships. "We decided what we needed to assess and what it would

look like," the Oakdale reading specialist said, adding that the teachers revised the

assessments every year based on their experiences with the assessments.

One such assessment provided a structure for a teacher's observations of desired

reading partnership behaviors. (See Appendix J.) Teachers wrote in a different color of

ink each time they used the sheet. This tactic enabled them to see how students'

responses changed from observation to observation. Teachers also noted teaching points

on the form. These notes guided their instruction the next time they met with the

partnership group.

Another example of collaboratively developed assessments is a Reading Log,

which contained two sets of rubrics. One set assessed the quality of the post-it notes the

student wrote in preparation for discussion (see Appendix K). The other set assessed the

student's level of participation in the discussion (see Appendix M). Both sets of rubrics

called for student self-assessment and teacher assessment. Different versions existed for

various grade levels and genres. According to the reading specialist, these rubrics and

checklists were not unique to Oakdale. Assessments like these were shared across the

district among teachers who were striving to create better ways to assess their students'

on-going reading achievement.

I saw the Oakdale reading specialist use one of these teacher-created assessments

when I observed the classroom of one of the Oakdale teachers who participated in this

study. The reading specialist was collaborating with the teacher on non-fiction reading

strategies and wanted to know how well students were able to apply the strategies that

had been discussed. Appendix L contains a sample of this paper and pencil assessment.

The process of "scoring" the assessment was really a matter of walking around the room

while the students were completing the task. The results helped the reading specialist and

the classroom teacher determine the next instructional steps for the class as a whole, as

well as the next steps for those who struggled.

The simplicity of this paper and pencil assessment, along with the ease of

evaluating the results can not be underestimated. Although the teacher's observation log

and the student's reading logs, and teacher/student rubrics were remarkable in terms of

the breadth of information they captured, comments made by the classroom teachers led me to believe that it was sometimes hard for a teacher to keep up with the volume of paperwork these assessments generated in comparison to the paper and pencil assessment discussed above. "If I didn't have time to work on this at home, it would be hard for me to keep up with it, said one teacher. "Sometimes it's kind of hectic to record everything," explained another teacher. "You get interrupted or get a phone call, or the class period is over," the teacher said.

The Maple Grove reading specialist also spoke about some assessments created in response to the needs at Maple Grove. The reading specialist designed these assessments to measure reading competencies that were valued by both the reading specialist and the teachers but were not measured by existing assessments. "I never found assessments that assessed things like cause-effect relationships or comparing and contrasting and I would like assessments to get down to more of those things with children," said the Maple Grove specialist. In response to this need, the reading specialist designed simple assessments with pre and post measures to assess students' ability to compare and contrast. The specialist also designed a simple discussion assessment to monitor the number of times a child contributed to a discussion along with a checklist to monitor the quality of the discussion.

One salient feature the assessments cited above is that the data they provided tended to drive the next instructional steps. It appeared that teachers did not use these assessments to formally track the overall performance of the class or to formally track the changes in an individual child's performance overtime. From what I observed or learned through the interview process, the reading specialists and classroom teachers used these

assessments to get a sense of the state of the class. That informed their decisions about

the topics they would reinforce through whole group mini-lessons or within small group

discussion.

On-going assessments also captured other useful information. For example,

students had to record the number of minutes/pages read over the course of a week or

month. This is worth mentioning because it demonstrates that teachers chose to capture

data other than the kind that documented strategy-use. I observed teachers ask students

about the log when they were conducting individual reading conferences. The logs, which

recorded in-school reading and at-home reading, gave the teacher's a sense of the

student's reading volume and pace. A reading stamina rubric helped students self-assess

their independent reading. I also noticed a poster-sized rubric posted in several classes

which was designed to help students assess if a book was "just right" for them.

Keeping Track

The assessments described above were accompanied by some kind of paperwork

or form for capturing the data. That's not to say that all classroom assessment data were

recorded. When asked about what they had noticed about teachers' assessment practices,

the reading specialists said teachers didn't write down everything they observed. The

Maple Grove specialist commented:

> Sure, some of it is in their heads, of course it is. You couldn't possibly
> record everything. But when I speak to teachers about individual students it
> always amazes me how well they know their students. And sure some of
> that is from what they actually record, but a lot of it is up here [points to
> her head]. I don't necessarily see it as one or the other because it is so
> individual. You know, teachers are kid-watchers and some of us keep it
> very well all up here and I can't tell you if there is any real difference
> between how well they know their kids compared to those who write things
> down. I don't write everything down.

The Woodland reading specialist also spoke about teachers' assessment practices. "Everybody kind of does their own thing because that is such a personal thing. You know, kind of like everything that works for me doesn't necessarily work for someone else," the Woodland reading specialist said. The specialist explained that while one teacher might have a spiral notebook for every student, another teacher might find that to be too cumbersome and opt for one binder with tabs in the binder and sheets behind the tabs for recording notes. Still another teacher might like to use post-its and carry a clipboard or use a post-it and put them in a notebook. "As long as they are taking some observation notes and using them to plan their instruction, I think the method with which they do that has been more left up to them because it is more of a management thing," said the Woodland reading specialist.

It's important to remember that the classroom teachers had a great deal of autonomy in implementing the assessments described above. They determined what to assess and how often. They also decided the method for recording the results, and how they would use the data to guide instructional decisions. These variations in practice meant that it was unlikely that teachers would ever aggregate the results and use them as a starting point for grade level conversations about the varying needs of students within a grade level. Those conversations were best facilitated by common assessments.

The Woodland reading specialist said that focused conversations about the results of common classroom assessments were beginning to occur at Woodland school. "Within the PLCs we are starting to do that a little bit better." The specialist explained that a group of teachers had developed a common assessment on making inferences about characters which they administered to the entire grade level. When the teachers came

back together as a team they examined how the students were doing and discussed possible next steps such as repeating the lesson, taking it to the next step, and/or providing more support for some students. The reading specialist reflected on the process: "So, I think in terms of developing some of those common assessments, they are talking more about what a whole class is doing. But, I think that is a relatively new concept," the reading specialist said. Incidentally, this remark illustrates a difference in practice among buildings. As will be discussed later, teachers in other buildings spoke about some of the collaborative scoring they had done with student work.

Staff Development

The reading specialists at Oakdale and Maple Grove had been assigned to their respective buildings for years. Providing staff development was routine for them. They delivered formal presentations and also provided staff development through co-teaching and collaboration. That may be why they did not speak extensively about the topic of staff development as it related to developing classroom assessments.

In contrast, the Woodland reading specialist spoke about the topic at length. The discussion flowed from a discussion of on-going classroom assessment and the level of knowledge and skill teachers needed before they could easily use assessment in their balanced literacy classrooms. "I think, for those of us who have been here for awhile, it's a natural part of what we do because we have to in order to be able to teach the way we teach here," the specialist said. In reflecting upon personal experiences as a classroom teacher, the Woodland reading specialist empathized with the day to day challenges inexperienced teachers face when trying to incorporate assessments into their daily practice. The specialist said the difficulty of incorporating on-going assessments also

caused additional challenges for these teachers because without appropriate assessments, the less experienced teachers struggled to determine the focus of a guided reading lesson.

Consequently, the Woodland reading specialist viewed staff development as critical in helping teachers meet those challenges and bemoaned the fact the less experienced teachers in the building had not received the extensive staff development in the balanced literacy model that teachers new to the district has received previously. The Woodland reading specialist believed that solid staff development prepared teachers to conduct the kinds of assessment that enabled them to be successful teachers of reading within their respective classrooms.

> I think sometimes teachers get bogged down with the day to day grind of school. And to think, oh yeah, I have to pull that kid out and assess him—I mean sometimes it gets difficult to manage it all and fit it all in. And certainly management in a balanced literacy classroom is tricky which is why staff development is so important. I mean, I find staff development to be a key. I firmly believe that programs do not teach our kids, books do not teach our kids, teachers teach our kids and they have to be expert teachers of reading to do it.

In comparing less experienced teachers to those with more experience, the Woodland reading specialist said the experience level of a teacher played a key role in the teacher's ability to use assessment data effectively. The reading specialist explained that an experienced teacher would know how to conduct an assessment, how to analyze the data, how to determine what the data say about a particular student, and, last but not least, know how to use the data to inform instruction. The Woodland reading specialist emphasized the importance of staff development as a key factor in shaping an experienced teacher's ability to use assessments well. "It's experience and staff development because you could have a very, very experienced teacher who has never

been taught how to use that data to inform instruction and then it isn't very helpful," the

reading specialist said.

The Woodland reading specialist also attributed other factors to a teacher's skill in

using assessments within the teacher's own classroom. The specialist believed a teacher's

interest in reading, or even the extent to which a teacher embraced the concept of

balanced literacy, were also factors in shaping the degree to which a teacher used or was

able to use on-going classroom assessments.

> I think it depends on how strongly they buy into the philosophy of balanced
> literacy. If you have somebody that really believes in balanced literacy and really
> understands it, then their use of and understanding of assessments is strong. I
> think some people who don't necessarily buy into balanced literacy as much and
> would prefer to have a textbook and a basal that they could flip though aren't
> going to buy into the whole assessment piece as much either.

Tacit Expectations

The three reading specialists considered on-going classroom assessments to be

tacit expectations that accompanied the balanced literacy model. The Oakdale reading

specialist spoke to the issue of accountability:

> [Teachers] are accountable for that definitely because we as teachers are
> accountable for showing parents and students how well they are doing on a
> daily and weekly basis and we need to provide them with that feedback, as
> well as have that information to make instructional decisions. There's no
> written district policy that says they have to have them. It's just understood,
> as teachers, that we need to have a balanced assessment system and those
> informal assessments are a key part of that.

The on-going assessment were also essential because they influenced students'

grades. However, teachers did not assign grades to the assessments and then calculate an

average. The Oakdale reading specialist explained the process of assigning report card

grades using a rubric scoring system, a system she helped develop. (See Appendix O.)

The specialist explained reason for the rubric scoring system:

The process was developed to help parents understand that in reading and writing you just don't average. You just can't take a percentage or average grades and get a grade. It's not that way like it is in math; it's a whole different ball game. And so what we would do is we would ideally take all of our anecdotal records, as well as what we would have gathered from our district level assessments, and then use that to help us create a picture of where the students are and we would highlight where we see them falling and then we could say most of them fall within "B" so the overall grade in reading is a "B".

The rubric served an important purpose at conference time. Classroom teachers often gave parents a highlighted copy of the rubric along with the report card so that the parents could why the teacher assigned a particular grade. This grading process meant that teachers saved a lot of student work over the grading period. Some participants said they sent work home and had parents sign it and return it. That way the teacher had the student's actual work at hand when making grading decisions and when discussing a student's performance with parents. Some participants also created a portfolio for each child for this purpose.

Allocation of Resources

The Woodland reading specialist considered assessment results as something that heavily influenced programmatic and curriculum decisions. For example, the reading specialist said data might influence something relatively straight forward like the kinds of books purchased for a school's tradebook collection. On the other hand, the reading specialist said that data also influenced the schedules of the reading specialist and the instructional aides. "We really look carefully at our data on individual children and, when we meet with teachers, we say, ok, these children are really likely to struggle in reading this year. What are we going to do as a team to meet their needs?" Those conversations would determine how the reading specialist supported the classroom teacher in improving

the teacher's classroom instruction. Data might also determine if an aide would be used for classroom support, or if the reading specialist would work with individual students to increase their learning. "All of that is driven by what we notice our kids doing on their district level assessment and classroom assessments," the reading specialist said.

Data also informed decisions about students who excelled in reading. At the time of our interview, the Woodland reading specialist was working with a small group of advanced primary students. As time went on, the specialist anticipated that a literacy aide might be assigned work with that group of students so that the specialist's time could be devoted to serving other children.

In summary, the reading specialists, like the principals and administrators, believed that the purpose of assessment was to guide instruction. However, not until we reach the reading specialist level of the organization that we see evidence of assessment actually guiding the delivery of instruction. This occurred when the reading specialists collaborated with teachers on lesson design, delivered classroom instruction, or served specific students. Although the reading specialists did not identify accountability as a purpose of assessment, this purpose influenced their actions, too. This accountability purpose of assessment is evidenced by the interventions, lessons and assessments the reading specialists created in response WKCE and the MAP data.

Classroom Teachers

Ultimately, the task of using data to inform instruction falls into the hands of the classroom teachers because they are the individuals who ideally use data on a daily basis to inform classroom instruction. Their task is complicated by the fact that classroom teachers possess varying levels of experience, training, and interest in reading instruction

and assessment. As we will see below the intersection of these factors influenced the ways in which teachers use assessment data to inform instruction.

In order to highlight the interaction of these factors, I will begin this section with a look at one teacher, an upper elementary teacher from Oakdale. I selected this teacher as a starting point for the discussion because of the way the teacher used data from multiple sources to influence instruction. I will later compare this individual's practices with those of other participants.

Three factors—experience, training, and interest in reading instruction— appeared to influence the ways in which the Oakdale upper elementary teacher used assessment to guide instruction. The teacher had over 20 years of experience and enjoyed working with data. MAP data and on-going classroom assessment data guided the teacher's planning. The teacher valued the data gleaned from the district level reading assessment, as well as the WKCE results and considered both measures snapshots of what students were able to do. This upper elementary teacher also had an interest in the latest research on reading instruction. The teacher was a devotee of Lucy Calkins (2000) and, along with other teachers from the Valley View School District, had attended some of Calkins's institutes. In fact, this particular teacher was one who collaborated with others on developing some of the rubrics that were used across the district. It would be accurate to say that this teacher was a strong proponent of the balanced literacy model and admitted devoting a lot of time planning and collaborating on the best ways to implement the model within the classroom. The teacher collaborated with the reading specialist as well as with grade level team members.

Beginning the School Year

   At the start of the school year, this upper elementary grade teacher from Oakdale used MAP data as an early indication of students' strengths and needs. "We absolutely look at that MAP testing and we study it and we base our initial groupings at the beginning of the year based on that testing because it is done in the first week of school," the teacher said. The teacher emphasized that MAP data influenced the grouping decisions that were made so that students could get started in their reading groups. The teacher anticipated the need to adjust group membership over the year. "Sometimes kids stay together because it is working and they are having great discussions but it all depends on what they need. And the kids know that and they have to be willing to move around," the teacher said.

   Out of all the classroom teachers who were interviewed, this teacher was the only classroom who spoke confidently about understanding the various reports generated by MAP. While the other participants used the data as a general indicator of achievement, this teacher used the results to guide targeted instructional decisions. During the interview, the teacher spoke about an individual student's results, the results for a group of students, and the instructional decisions made in response to the data. The teacher focused on the results of one student, 'Jack', to make a point.

> Ok, well here's a kid, Jack, he's a good one. He's really high in understanding the text but he was really low in analyzing the text. So, what we did was ask, what does analyzing text mean? This is an area we really zoomed in on as [grade level] teachers because we noticed that there were a number of kids who were low in this area and that it has to do with inferring, taking information to condense it, to summarize it, and so on. And, so really, all of our reading instruction zooms in on that.

The teacher reported that the MAP data influenced instructional decisions made by the teacher and the teacher's grade level team in three areas. First of all, the team of teachers decided it was important to press students to make more inferences when they were engaged in a book discussion. The team also decided to encourage students to elaborate when they spoke or wrote about a book, an idea, a theme, or the characters. When it came to nonfiction, the teachers agreed to prompt students to go beyond the facts of the book and to make connections. They also used several stems to scaffold the students' responses to text. Some of the stems included: "This makes me realize…"; "Another example is…," and "This is important because…"

According to the Oakdale upper elementary teacher, the value of these MAP-based instructional decisions was reinforced by the WKCE data. The Oakdale reading specialist had conducted an extensive analysis of WKCE results and had found that student also needed improvement on WKCE items which required them to analyze text. In response to the data, the reading specialist at Oakdale had provided staff development to assist teachers in designing instruction that would enhance students' abilities to analyze text. Although this particular Oakdale upper elementary teacher recognized that it was important for Oakdale students to perform well on the MAP and on the WKCE, and that it was the teacher's responsibility to ensure students had the skills they needed to do well on these assessments, this teacher did not let allow the MAP or WKCE to dictate the daily instructional program.

Classroom Assessments

The Oakdale upper elementary grade teacher used various templates to capture the reading competencies to be measured. One of the most noteworthy attributes of the

templates used is that they assisted the teacher in capturing real-time data that came in the

form of coded anecdotal notes or rubric scores. (These tools were discussed earlier and

can be found in Appendices J, K, and M.) Although these data were far different from the

scale scores or percentiles provided by the WKCE and the MAP, its format made

complete sense in light of how this Oakdale upper elementary teacher defined the

purpose of assessment. Like every other participant in this study, this particular teacher

said the purpose of assessment was to inform instruction. By examining this individual's

comments related to the purpose of assessment one can understand why a commercially

produced measure like MAP or WKCE would never capture all the information about

students that this classroom teacher valued. The Oakdale teacher explained her view of

assessment:

> The purpose of assessment in reading is to understand how my students are
> functioning as readers, what their comprehension level is, what their decoding is
> like, if they understand how books work, that they understand different genres
> and how you read differently based on the genre, how to find a "just right" book,
> if they are reading "just right books", how to turn kids on to books, are they
> engaged as a reader, do they have a list of books that are on deck and they are
> waiting to read or are they jumping through many books and not really finishing,
> not really engaged. So I want to know, basically, what kind of a reader are they.
> And then I would use that information to plan instruction. So I guess the purpose
> ultimately is to plan instruction.

In addition to using specific templates, the Oakdale upper elementary teacher

systematically collected data on students twice a week while the students were engaged

in their discussion groups. This data collection process took about five minutes per group.

I observed the teacher sit and listen to the discussion and simultaneously code the

students' responses. On this particular day the teacher was assessing whether or not

students could support their statements with evidence from the text. The assessment

ended with the teacher giving students a "powerful praise." The compliment was followed by a teaching point related to what had been observed. The notes on the template were color coded so that it was easy for the teacher to track the patterns of students' discussion over the course of an entire book. This kind of data collection was sometimes challenging, even for someone with this teacher's level of experience and confidence. "Sometimes it's kind of hectic. You get interrupted or get a phone call, or the class period is over," the teacher explained. Nevertheless, the teacher appeared to be committed to this method of on-going assessment.

The Oakdale upper elementary teacher also required students to complete a reading log that was designed to match the genre students were studying. This log contained three sets of rubrics. One set described the quality of the post-it notes students prepared for discussion, another set described the quality of the writing a student prepared as an extended response to one of his or her post-its, and the third set described the quality of the student's participation in discussion. Students marked the rubric first and then the teacher highlighted the same rubric as a way to give feedback to students on their work. (Both of these assessments were discussed earlier. Samples are found in Appendices K and M.)

The Oakdale teacher also kept a reading conference binder. The binder contained notes the teacher recorded during weekly one-on-one conferences with students. The conferences were focused on a student's *reading life*, a term the teacher used to refer to a child's independent reading habits and interests. During this conference, the teacher reviewed the child's reading log and asked questions about the independent books the child was reading.

The reading log also served as a self-assessment for the child in terms whether or not the child met the classroom reading goal of one book every week and a half. Students were expected to track their reading everyday both in school and at home in terms of starting and ending page and the number of minutes they read. At the end of each month students totaled the number of school minutes, the number of home minutes and the number of books read. In addition to the reading log, students were occasionally asked to reflect on themselves as readers. They did this reflective work in a spiral notebook which was tabbed to accommodate four topics: My Reading Life, Reading Practice, Independent Reading and Read Alouds.

The purpose for all of this record-keeping was made clear to the students at the beginning of the year. The upper elementary grade teacher reiterated the message given to students in September:

> We study you as readers. That's our job to study you. So, with all of these things, you might think, oh, I gotta do this, but really it helps us know you and it helps you know yourself as a reader.

The concept of studying students as readers extended beyond this particular teacher's classroom. The teacher spoke about collaborating with grade level colleagues to determine if students were making inferences and to learn more about the other kinds of thinking students were doing. This grade level team also wanted to determine if students were growing as readers. To answer their questions, the team worked during their PLC time to analyze students' notebook responses. This analysis led them to consider the steps they needed to take in order to help their students grow as readers.

The data the upper elementary grade teacher captured through on-going classroom assessments also served another purpose: grading and reporting. As discussed above,

students' report card grades were assigned by evaluating a child's reading progress against a grading rubric that had been establish by the district. The students in this teacher's classroom did not take reading quizzes or tests, therefore, the on-going assessment data the teacher collected became the basis for assigning grades, as well as the evidence for justifying a grade, if the need should arise.

In contrast to this upper elementary grade Oakdale teacher, a primary grade Oakdale teacher who participated in this study did not use MAP data to make beginning of the year decisions. Instead, this teacher used data from each student's reading folders to guide early group placement decisions. (The folders, which followed each child from grade level to grade level, contained a variety of information about the child's reading experiences, including the results from the student's January district-level reading assessment, the books students read throughout the year in guided reading and the student's May guided reading level.) In addition, the Oakdale primary grade teacher sought input from the previous year's teachers to place students. This teacher also observed students: "The first two weeks to a month, we are just reading books. You know, we sort of guide them to their own level based on last year and then we listen to them," the teacher said. Running records helped the teacher understand the needs of students. Those not "up to par" were given the district level benchmark assessment so that the teacher could gain a more complete understanding of the children's needs.

Another Oakdale teacher who participated in this study also worked with upper elementary students. This teacher approached the task of placing students with a different mindset, as well as a different procedure. Like the other Oakdale upper elementary teacher, this teacher also valued MAP data, and used it to gain a basic understanding of

students' relative strengths and weaknesses. However, teachers on this individual's grade level team administered the district-level assessment in September, despite the fact that it was not a district requirement. The teacher commented on the usefulness of this data:

> We give the [district assessment] at the beginning of the year and it's pretty counterproductive as far as I am concerned. Granted there is some fall back in the summer time, but, to have to use class time … Well shoot, it takes us probably a week by the time we get them all done. To take that time at the beginning of the year to reassess when we kind of already know where they are at anyway [based on the reading folder information], just seems to be counterproductive to me. I think it would be more beneficial to just get into those partnerships [2-3 students reading and discussing the same book] and hit it running and let things kind of work out.

The frustration expressed by this particular teacher has to be understood in light of the way the teacher approached classroom assessment. "I would say, the vast majority of the assessment that we do is the anecdotal stuff in the classroom. We are big on reading and writing here and so those two areas are a daily check on progress," the teacher explained. A visit to the teacher's classroom showed that informal assessment was part of the daily routine. (Incidentally, this teacher's impatience with the amount of time it took to administer the district level reading assessment was echoed by two of the three principals. They, too, were concerned about opportunity cost of this assessment in terms of lost instructional time.)

Both the Oakdale primary grade teacher and the Oakdale upper elementary teacher discussed above collaborated with the reading specialist and used some of the same reading logs and teacher observation logs used by the upper elementary Oakdale teacher who was discussed at the beginning of this section. However, the two teachers did not appear to make extensive use of MAP data. It may be that these particular teachers simply needed more time to understand how to use the results of the MAP assessment.

Or, perhaps they were influenced by some of the uncertainty their reading specialist expressed about the test. Whatever the reason, the varied approaches of three teachers within one building highlight the challenges a district faces when it sets its sights on having all teachers using data from one measure to guide some of their instructional decisions.

An entirely different assessment dynamic existed in the classrooms of the teacher participants from Woodland school. Several factors may account for the difference. For one, the Woodland teachers who participated in this study were teachers with less than three years of teaching experience. Both were serving as upper elementary grade teachers. One had previously served as primary grade teacher. In addition, both spoke of being overwhelmed by the challenge of planning for and managing a balanced literacy classroom. They expressed a lack of confidence in knowing exactly how to conduct such a classroom. Due to personnel changes at the district level, these teachers had not received the extensive new teacher training in reading that had been provided to the experienced teachers at Oakdale and Maple Grove. At the time of this study, the two Woodland teachers had just begun to participate in a staff development workshop conducted in the district. As a result, they were learning about and starting to use the assessment templates that were being used by Oakdale teachers and other teachers in the district. However, at the beginning of the year the Woodland teachers based instructional decisions on information from the students' reading records, the district level reading assessment, and MAP assessment. Each used the data in different ways.

One of the Woodland teachers began the decision-making process by examining students' reading folders from the previous year. The teacher considered that information

along with the beginning of the year MAP assessment results and the teacher's personal

observations. The Woodland upper elementary teacher explained the process:

> I tried to meet with every kid in the first couple of weeks and tried to make
> some observations. I had them read a passage and then had them do a
> retelling and then asked them a couple of questions on what they are doing.
> So I kind of did my best guess and then I picked out a couple books from
> the guided reading group and I tried to also meet with each kid to read a
> book. I tried to pick where they were from last year so I gave them a guided
> reading book and asked them to read so I could tell from there. Then I tried
> to give them a harder one and see where their frustration was. I kind of did
> my own mini [district] level assessment and that kind of helped me.

The other Woodland upper elementary teacher used information from the

students' reading folders, MAP test data, and observational data to place students into

instructional groups. MAP testing alerted the teacher to students who were potentially

struggling readers. However, the teacher wasn't willing to take the MAP data at face

value because students took the test within the first two weeks of school: "I just thought

well, you know, they are getting used to me, they are getting used to being back in

school, so it [the MAP data] might be skewed a little bit," the teacher said.

This skepticism prompted this Woodland teacher to administer the district level

benchmark assessments to learn more about the low-scoring students. As part of the

assessment process, the teacher took notes and tried to understand where students might

be having difficulty. In addition the teacher also consulted with the previous year's

teachers to gain more insight into the needs of struggling readers. When it came placing

those students determined to be average and above average students based upon their

MAP results, the teacher consulted the students' reading folders as a second source of

information. "I just took the level from their last level and started them from there," the

teacher said.

Based upon my interviews with the Woodland upper elementary teachers, as well as observing their classrooms, it appeared that these teachers struggled in conducting the kinds of on-going assessments that was second nature for Oakdale teachers. What set the teachers from the two schools apart is that the Oakdale teachers seemed to be better able to focus students on the discussion at hand. As a result, they appeared to be better able to collect more accurate data on a student's performance, whether it was the student's ability to make predictions, or to provide evidence, or to justify an opinion. If students drifted—for example, got too far afield in talking about an unrelated personal experience—the Oakdale teachers were able to bring them back by either explicitly redirecting them or asking a probing question that naturally got the students back on track. The Woodland teachers were less skillful with their questioning techniques. In one instance, a Woodland upper elementary teacher appeared to be hesitant to interrupt a student, even though the student was far off track in the discussion. In addition, the teacher's interactions with a small group were interrupted by the misbehavior of students who were to be working independently. In another case, one of the Woodland upper elementary teachers appeared to accept incomplete answers from students, even though it seemed that a probing question might have enabled the students to express themselves more completely. It is entirely possible that the younger teachers were nervous with an observer in the room, which affected their ability to think on their feet when interacting with students. But, whether it was due to nervousness or lack of skill, the on-going classroom assessments the Woodland teachers were using were not fully capturing the reading behaviors that they believed were valuable. Lack of teaching experience can't be discounted as a factor that impacted the teachers' assessment skills, particularly when one

learns about the way in which one of the Woodland teachers assessed students the year
before.

Staff Development

One of the Woodland upper elementary teachers spoke about the value of staff
development and its influence on how the teacher taught reading and used assessment to
guide instruction. The previous year, this young teacher's assessments consisted mainly
of short answer comprehension quizzes. The teacher had used these assessments with
guided reading groups because the teacher lacked knowledge about other assessments
that might be used. Furthermore, the teacher remembered comprehension quizzes from
the teacher's personal experiences as a student. Professional development in the form of a
district-wide workshop changed this teacher's view about reading instruction and
assessment. The experience led the teacher to focus on the quality of students' responses
as they discussed the books they were reading. The teacher explained the new assessment
approach by stating: "readers don't take quizzes on things they read in real life". As a
result of this new view on assessment, the teacher's plan for the current year was to
involve students more in talking about their books because, as the teacher put it, "that's
what adults do when they read a book—they discuss it." This shift from using quizzes to
a focus on assessing through discussion led to the use of the assessment tools discussed
above.

Primary Teachers

The assessment practices of the Maple Grove teachers who participated in this
study were very different from those of the other teachers, and for good reason. Given the

developmental nature of reading, these teachers had to concern themselves with skills that were far more basic than those expected by upper elementary teachers.

It is important to know that both Maple Grove teachers were experienced primary grade teachers, with seven and eleven years of experience, respectively. Furthermore, each had received extensive staff development on the balanced literacy model when they joined the district and, without prompting, spoke positively about the experience and how it had helped them know how to implement the balanced literacy model within their own classrooms. In addition, both made an effort to keep abreast in best practices through graduate level coursework and other professional development experiences. Assessment was a routine part of these teachers' practices, and, like the other teachers in this study, they used assessment results early in the year to make decisions about initial group placement and to gain insight into the strengths and needs of individual students. Also, like the other teacher participants, they said that the purpose of assessment was to guide instruction.

For example, one of the Maple Grove primary grade teachers started the year by using the district level reading assessment benchmark kits to conduct a basic running record. "It was really about determining what they needed and to be able to put them into groups at the beginning of the year," the teacher said, adding that even students with the same reading level might have different needs within that level. The running record results gave the teacher a general idea about where students were with their decoding ability. Where appropriate, the teacher also used the materials from the benchmark kits that were used for the district level reading assessments to assess comprehension. At the beginning of the year this primary grade teacher had students with widely varying

reading abilities. One student couldn't attend to the print on the page; others were reading above grade level and, as the teacher explained, "cruised through the material" put in front of them. Once the teacher established reading groups, on-going assessment became part of the regular routine. The kind of assessment the teacher selected depended on the reading situation and the child.

Sometimes the teacher took anecdotal notes while listening to a child read and then selected a strategy for the child based upon what had been observed. Other times, the teacher did a formal running record. And sometimes the teacher claimed to *just know*. "It's just a matter of being able to tell that they are ready to move on. It's just internal, like, ok, there's no struggle here we just need to move on," the Maple Grove primary teacher said.

The teacher said it took a couple of years to develop that sense, along with a sense of how to assess students and select strategies suited to individual needs. (A copy of the log sheet the teacher kept for each student is found in Appendix P.) The teacher spoke about the way in which assessment practices changed as a result of experience:

> In the beginning it was pretty much all running records and it was all the same strategy and it was almost—I don't want to say prescribed—but it wasn't individualized. Now, I am at the point where I can teach all these kids in the same group and I have a strategy in mind and I teach them all the same strategies. But when I am pulling them individually, then I might teach them something else when it's just the two of us, even though the rest of the group is doing another thing.

After moving away from doing extensive running records, the teacher began to focus more on students' comprehension abilities. For example, the teacher spoke about asking students to read to find the answer to question and then flagging the answer with a post-it. The teacher was able to assess students' ability to find the answer by noting

where each student placed the post-it. The teacher noted that when assessment shifted

from tracking decoding (which the teacher had been trained to do) to paying attention to

comprehension, the volume of books students read each year grew from 10-12 books a

year to 20-30 books a year. Although some of that change may have been due to the

teacher's growth as a professional, the thought that a change in the type of assessment

could impact a change in achievement as measured by the number of books read is worth

further investigation.

As might be expected, formal assessments were part of this teacher's repertoire,

too. Like other teachers in the district, this Maple Grove teacher administered the district

level reading assessment to all students. Despite the teacher's commitment to on-going

assessments, sometimes the district level assessment results proved to be a surprise:

> You know, assessment is on-going throughout the year but then in January,
> I'm even surprised about the results. Even though I do the anecdotal notes or
> running records, I am still sometimes really surprised about the results, like
> when you pull them individually and keep giving them blind books that they
> have never seen. I am amazed that they can do even more than what I had
> them doing.

Sometimes the results from the district level reading assessment provided data to

suggest that a student wasn't making adequate progress and would need a double dose of

reading instruction delivered by both the classroom teacher and the reading specialists.

Other times, the results suggested that a student had made wonderful progress in a

double-dose situation and no longer needed intervention.

Although much of this Maple Grove teacher's assessments focused on individual

students or small groups, some of assessment focused on the work of the entire group.

For example, the teacher recounted an assessment where students had to write details

about a character in a self-selected book. This mini-assessment aligned with classroom

work on characters and also dovetailed with a comprehension improvement goal established by the teacher's grade level team. "We've been doing these types of things to build comprehension, with the end goal being that they will include more details in their retellings and will be able to answer more of the questions," the teacher said.

The teacher further explained that the comprehension goal was developed as part of the building and district focus on improving reading. The teams' own experiences with students also influenced the goal. In addition, standardized test data played a part. Referring to the teacher's grade level team, the teacher said, "We don't do MAP and WKCE but people in the building had all this data that they put together and comprehension was what showed up."

The teacher's team established a baseline for comprehension and retelling at the beginning of the year. The January district level reading assessment provided the grade level with another data point, and they planned to administer the district level reading assessment again in May for their third data point.

Assessment was part of the regular routine for the other Maple Grove primary grade teacher who participated in the study. Assessment data guided the way she started the year because the teacher assessed all students in the class, not just those who didn't meet the previous grade's benchmark on the district level reading assessment. That's because this primary grade teacher believed that there was the possibility that students grew in their reading ability in the summer.

Students' fluency scores influenced the instructional decisions the teacher made early on. "I always look at fluency so if their scores are low we try to do a lot of repeated reading. I do a lot with poems and use them with repeated readings to foster fluency

because they are short and sweet," the Maple Grove teacher said. The students' scores on retelling and comprehension also influenced the teacher's decision-making process, but to a lesser degree. That's because the teacher incorporated opportunities for retelling and comprehension in all the guided reading groups.

When it came to guided reading time, the teacher wove several assessment tasks into instruction. I observed one vocabulary assessment in which the teacher had students mark a list of words to show how well they knew the meaning of a word. This assessment was developed in response to the fact that students were able to read words fluently but didn't know the meanings. Students marked a "K" if they knew the word and could explain the meaning and marked an "L" next to words they wanted to learn. When they were done with this very brief assessment, the teacher used this information to make on-the-spot decisions as to which words to emphasize so that students had the word knowledge they needed to comprehend the story.

The importance of comprehension also led to the implementation of another assessment in which students marked the parts of a story with IDGIs, which were post-its that stood for *I don't get it*. The teacher said IDGIs served as a tool to reinforce the idea that students need to understand what they are reading. The IDGIs provided a quick assessment, as well. "If someone has a lot of IDGIs, I think, *they are really in tune with their reading, they know they are not getting it.* But I kind of worry more about the child who doesn't have any IDGIs because I think they might not be catching the part they don't get," said the Maple Grove primary grade teacher.

When it came to recording informal assessments like the ones mentioned above, this teacher's method was fairly informal. The teacher's comments are repeated in detail

below in order to capture how a teacher's personal style meshed with the teacher's

instructional decision-making.

> I just take post-its and scribble notes to myself and leave stuff on the
> [teacher's] chair. So, when the kids are getting up, I take quick notes on what I
> need to do, where we are leaving off, who struggled on this, etc., and I put the
> note down. Should I have a notebook, yeah, but I tried it and ended up with
> stickies. I don't know why. I think it's just me to veer in and pick up where I
> left off. Like with all the groups, I wanted to focus on making connections and
> they weren't getting it. So, I just left a note for myself on this chair so that I
> made sure that, with everything we were hitting, I kept bringing them back to
> that.

Even the running record format didn't serve this teacher well. The teacher

attributed this to personal style which meant the teacher preferred to listen, take notes,

and then tell the child about what had been recorded. The teacher explained this

personal approach:

> I am so unorganized in that way, but in my mind, I meet with them so often
> that it works. So, I have them do cold reads and use note paper and write the
> words they missed. Then when everyone is done, I come back and say, *All
> right, let's go to this page and let's go here.* I wish I had a beautiful system,
> but I tend to take notes.

Although informal assessments guided some instructional decisions, MAP data

informed others, especially when it came to intensive interventions. The teacher said that

the Winter MAP results identified six students in the classroom who were below level.

These students received a double-dose of instruction from the teacher, four times a week

for a half hour.

Despite the fact that this teacher had missed some of the initial training on MAP

the teacher spoke positively about the potential the instrument might have. The teacher

believed MAP to be a "good overall snapshot and hoped to learn more about how to

interpret the assessment. "I want to know if the test can tell me if they struggle with

characters or setting. It's a goal for me to learn how to read and understand the reports," the teacher said.

In summary, like every other participant group, teachers said the purpose of assessment was to guide instruction. And, that is how they used it. Based upon my observations, the data they gathered was in the form of anecdotal notes, observations, or rubric scoring systems. These data captured the reading behaviors of individual students or small groups of students. Although the data gave teachers a sense of how students were progressing, they did not use the information to formally track a student's growth in such a way that they were able to assess the progress of all students on a particular competency. I observed a few occurrences where teachers assessed the entire class with what might be considered an informal classroom assessment. For example, I watched an upper elementary grade teacher purposefully listened to students while they were practicing non-fiction reading strategies. In another instance, I observed a primary grade teacher, in partnership with the reading specialist, using a writing activity to assess students' understanding of nonfiction reading strategies. In both cases, I noticed that these informal group assessments informed the whole group instruction that followed the assessment.

In other instances, whole class analysis was linked to goals established in light of the WKCE or MAP. As mentioned earlier, one Maple Grove primary grade teacher, collected whole group data on students' ability to give details about a character. In another instance, a Woodland upper elementary grade teacher had just begun to examine class data as a result of a common assessment the teacher and a teaching partner had developed in conjunction with their work with MAP data. Prior to analyzing the common

assessment data with a colleague, most of the data the teacher collected analyzed related to the work of individual's or small groups of students.

It is important to appreciate the volume of paperwork that accompanies the task of collecting data about student achievement. It appeared to be challenging for the teachers to keep up with the workload. Although all intentionally collected data on a regular basis, it seemed that the data got the most thorough analysis when it came time to assigning report card grades. "I use the information from the notebook I keep on each student, along with their work, to justify my grades," said one upper elementary grade teacher.

For the most part, the data that teachers captured and analyzed provided a "teaching point," which was essentially a written or mental note to self on what to do next. The concept of selecting a teaching point did not seem to present a problem for experienced teachers who had honed their observation and decision-making skills. However, on-going informal assessments appeared to be more problematic for less experienced teachers. These teachers were less skillful in facilitating a discussion, both in terms of keeping the discussion focused and in asking probing questions or follow-up questions. As a result, a session might end with the teacher having very little information to inform the next lesson's teaching point. A prime example of this occurred during an observation of a guided reading group in which a teacher was trying to reinforce the day's mini-lesson on making an inference. Although the teacher asked students for examples of inferences from their reading, students kept making personal connections to the text instead of making inferences. The teacher did not redirect them to reporting their inferences nor did the teacher ask probing questions that might have prompted students to build an answer from the connections they had already made. When the lesson ended, the

teacher had no information about the students' ability or lack of ability to make an inference within the context of that guided reading session.

When done well, the on-going assessments appeared to have a synergistic effect on classroom instruction. An assessment identified a teaching point, which informed another lesson, which led to another teaching point, and so on. The Reading Observation Log in Appendix J demonstrates how the process might work in practice.

Perspectives in Review

The sections above discussed the practices and perceptions of each participant group—district administrators, building principals, reading specialists and classroom teachers. Although each group had unique roles and responsibilities, they shared the common belief that the purpose of assessment is to inform instruction and that on-going classroom assessments were the most valuable in improving student learning. Yet, despite the shared belief, the groups diverged when it came to action. District administrators espoused the value of on-going classroom assessments, but they devoted their energies toward developing a district-wide plan which promoted the use of WKCE and MAP data as the driving force behind setting student achievement goals. Building principals espoused the value of on-going assessments, too. However, when it came to discussing the particulars of assessment the principals spoke of instruction related to WKCE or MAP data and of student achievement as measured by these instruments.

It is only when we get to the reading teachers and classroom teachers that we see more of an alignment between beliefs and actions. Participants in these groups stated that the purpose of assessment was to guide instruction and this was evident through their comments, through classroom observations, and assessment samples. These participants

understood the importance of the WKCE and worked to prepare students for the skills needed to do well on the assessment, but I found no evidence to suggest that test content influenced the daily program to the exclusion of other skills. Likewise, it did not appear that MAP data influenced the daily program to a great degree. However, there was evidence in one school's PLC logs that the data were beginning to have a systematic effect. Working collaboratively, teachers at this school were developing common assessments and common instruction linked to MAP data.

In the next chapter, I will discuss the actions and perspectives of the participants from the district perspective.

Data Driven Decision-Making:
A Case Study of How a School District Uses Data to Inform Reading Instruction

Chapter 5 – Discussion

We are a district of schools, we want to be a school district.
-Valley View District Administrator

At the time of this study, the Valley View School District was in the process of

transforming itself from a 'district of schools' with strong site autonomy to a 'school

district' with uniform expectations for goal-setting, testing, and professional

collaboration focused on data. This shift in organizational structure appeared to be

influenced by two factors: district administrators were concerned about the lack of

consistency in instructional practices across schools, and they were concerned about the

disparity in student achievement among schools. The administrators attributed these

differences to site-based management practices that had been put in place by previous

administrators. In essence, district administrators were faced with the challenge of

changing a culture that they had inherited.

On a local level, Valley View administrators were discovering school culture

issues reminiscent of those identified in a series of case studies conducted by the Center

for the Study of Reading in the early to mid 1990s. The studies were conducted in four

districts in Illinois—Alpha, Beta, Delta and Gamma. In a cross site analysis of these

studies, Stephens et al. (1995) found that "the meaning of particular concepts—

assessment, curriculum, accountability—varied so significantly across districts that to 'do

school' in one district was not the same as to 'do school' in another" (p. 9). To be fair, the

cultural differences in the Illinois studies were considerable; researchers discovered that

districts ranged from those that valued strong teacher autonomy to those that embraced a

strong top-down decision-making process. It would be an overstatement to say that the schools in Valley View occupied opposite ends of the continuum, however, as will be discussed later, there were some palpable differences—enough to say that to 'do school' in one part of Valley View was not the same as 'doing school' across town. As one principal remarked, "It's pretty autonomous here; each building is pretty much its own castle." Valley View administrators had a remedy for that autonomy, a remedy administered from the top-down.

It is important to recall the context in which that top down structure emerged. In review, Valley View's district level administrators were relatively new to the district, and as sometimes happens with a change in leadership, the board and community were looking to the Superintendent and the district administrators to usher in change. Prior to the Superintendent's arrival there had been some contentious issues in the community related to enrollment and school closings. There was a small but vocal group of parents strongly opposed to the district's balanced literacy curriculum because they believed the model hampered the achievement of some students. In addition, there was a concern among parents that the literacy curriculum was not uniform across the district which led to disparities in the way students were instructed. In short, various factions were fracturing the school community.

In response to the tensions within the school community, the Superintendent envisioned fostering harmony among stakeholders and wanted to redirect everyone's energies around a common purpose. Therefore, the Superintendent orchestrated the development of a long-range strategic plan in order to redirect those energies. "I thought it was time to shift the focus to what we are supposed to be about and to bring some

people together. We brought together some of the people that were the loudest and the angriest in the mix to put a long range plan together to help us move our focus in a different direction," the Superintendent said. The "different direction" the Superintendent had in mind was a data-driven focus on student learning, a focus guided by the long range strategic plan.

Before proceeding further into this discussion, it is important to acknowledge that this study is framed in context of NCLB. Undoubtedly this factor influenced the questions I chose to ask, the focus of my observations and document collecting and the lens through which I analyzed and report my findings. A different study might have selected other portions of the long range strategic plan for analysis. This study examines an aspect of the strategic plan that is closely linked to the intent of NCLB: the achievement of academic standards.

The strategic plan addressed six action plan areas: curriculum, post high school preparation, community service, student conduct, communication, and employment practices. One specific part of the plan, *Wisconsin Model Academic Standards/Analysis of Curriculum*, made numerous references to reading, assessment and data analysis. Overall, the plan established a global focus on student achievement not a specific focus on achievement in reading; however, it acknowledged the "critical importance of reading and writing," dubbed "reading and writing K-12 as foundation skills for success in all areas," and identified reading as an area of focus in the district.

One can appreciate the global focus on student achievement by examining some of the objectives and strategies of the plan. For example, a measurable objective states that "all students will demonstrate proficiency in the Wisconsin Model Academic

Standards." This objective is complemented by a strategy which states that the district

will "perform a comprehensive, in-depth, data-driven analysis of student performance

and current academic curricula, and will formulate an improved scope and sequence for

each academic core area that will maximize achievement for all students." Broad

statements like these are supported by a series of action steps which focus the

organization on data, instruction and student achievement. The steps also provide

evidence of the top-down decision making model at work in the district and are

summarized as follows:

1. Gather data continuously and systematically from multiple sources.
2. Train district personnel to use data to make instructional decisions.
3. Analyze data to determine strengths and needs.
4. Let data guide instructional decisions, including intervention and enrichment. Assess and monitor short-term progress in foundational skills.
5. Compare student data annually and track long-term achievement gains in core areas across schools and among subgroups.

At the time of this study, the Superintendent and two other district administrators,

the Director and the Assistant Director, were all focused on implementing initiatives that

supported the goals of the plan. Some initiatives were not affecting participants directly at

the time of the study; however, these initiatives are worth mentioning because they were

part of the environmental shift from a site-based culture to a culture guided by district-

wide expectations. For example, curriculum revisions were in process. In the future, these

revisions would inform the content of common assessments. A reading textbook adoption

was also underway with the intent that all elementary students would be using a new core

reading series the following year.

Other initiatives were affecting the participants in this study directly. These

included the implementation of the Measures of Academic Progress (MAP) assessment,

the expectation that WKCE and MAP data would guide the principals' building goals, and the implementation Professional Learning Communities (PLC), a staff development structure that fostered professional collaboration focused on data-driven instruction.

It is useful to consider the top-down structure the Valley View district administrators were establishing in light of the series of studies conducted in the Alpha, Beta, Gamma and Delta districts in Illinois (Shelton et al.,1993; Stephens et al.,1993a; Weinzierl et al., 1993; and Stephens et al.,1993b). The purpose of these studies was to understand the relationship between assessment and instruction in light of the decision-making process in each district. The authors who conducted these studies defined *instructional decision-making* in a broad sense so that the term included topics like textbook purchasing policies, curriculum development and site management, as well as the monitoring of student progress as measured by both standardized and classroom assessments. This broad-based definition of decision-making matches the type of decision-making structure in Valley View because curriculum development, the adoption of a district wide reading series, the use of the MAP assessment, uniform goal-setting linked to MAP and WKCE data, and the PLC initiative were all driven from the top down. In a cross-site analysis of these studies, Stephens et al.(1995) contended that "the salient relationship was not the one between assessment and instruction, but rather the relationship of these to the decision-making model in the district" (p.19). In other words, the authors found that standardized tests appeared to influence instruction in districts with top-down structures; conversely, they had less influence over instruction in districts with collaborative structures or in those that valued teacher autonomy.

Similarly, standardized tests were having a growing influence on instruction in Valley View. The top-down structure embedded in the Valley View long-range strategic plan ensured that WKCE and MAP data influenced building goals, which in turn influenced the work of the new PLC teams. The task of these teams was to use data to inform instruction—instruction that would lead to success on the objective stated in the strategic plan: "All students will demonstrate proficiency in the Wisconsin Model Academic Standards." The WKCE measured proficiency in the Wisconsin Model Academic Standards and the MAP assessment was aligned to the WKCE.

Such a tight alignment might lead one to think that Valley View was under pressure to dramatically improve its test scores. This was not the case. For example, 90% of district students scored at proficient or advanced levels on the reading portion of the WKCE. As the Superintendent explained it, "We are a high achieving suburban district, we are doing ok." Yet, the Superintendent was pragmatic about the reality of the test. "You know that you need to have a focus on it, you just can't ignore it, to pretend it's not important because it is." Although the Superintendent recognized the need to monitor the progress of subgroups, at the time of this study subgroup performance was not alarming. Two subgroups surpassed the 74% benchmark in reading required by the state during the 2007-08 school year, with one subgroup scoring at 79% proficient and advanced and the other group scoring at 77%. Only one subgroup fell below the mark, with 58% of those students scoring at proficient or advanced levels. Regrettably, I failed to probe more deeply about the superintendent's reactions to the score of that particular subgroup or to the performance of subgroups in general.

Despite the fact that Valley View was, as the Superintendent put it, "a high achieving suburban school district," it is interesting to compare its organizational structure to that described in studies which investigated improved performance on mandated tests. For example, in studies conducted in Ohio, North Carolina, and Texas (Kercheval & Newbill, 2002; Cobb, 1997; Charles A. Dana Center, 2001), we can see a top down organization structure, too. Although the authors of these studies do not discuss the concept of autonomous and top down decision-making structures, these studies show evidence of organizational structures that support the use of data in schools which are experiencing external pressure to improve. In such systems, leaders ensure that data are organized, analyzed, and disseminated so that teachers can devote their time to making instructional changes in response to the data. The leaders also provide professional time for data-driven work and they dedicate resources toward the purchase of practice materials or curriculum materials aligned with the test content. Likewise, in Taylor et al (2000) we see that organizational structures that supported the use of data was one characteristic of schools that showed greater than expected improvement in reading achievement

As an aside to this discussion of top-down organizational structures, it is interesting to think about the concept of systematic support for the use of standardized test data even when organizational structure or the external pressures to improve are not part of the discussion. Brozo and Hargis (2003) used data from standardized reading assessments test in a low-stakes fashion as tools to improve teaching and learning. The authors, in the role of researchers/consultants assembled and shared reading test data with high school teachers. The teachers' response to the data led them to implement changes

within their high school program. Although one can't support an argument from one study, it is worthwhile noting that the support for using data came from outside the classroom. As in the situations with top-down administrative structures, data were organized and provided to the teachers; teachers made the instructional decisions.

The Inherited Culture

The Valley View strategic plan established systemic expectations in three areas: the use of WKCE data and MAP data to establish building goals, the implementation of the Measures of Academic Progress (MAP) assessment, and the implementation Professional Learning Communities (PLC), a staff development structure that fostered professional collaboration focused on data-driven instruction. These initiatives were being implemented within a culture already accustomed using data to inform instruction. It is important to understand these embedded practices because they provide a context for understanding the interactions between the new initiatives and the established ways of using data to inform instruction. An examination of these practices also provides *partial* answers to the following research questions.

- What reading competencies do districts measure and why?
- Who selects what competencies should be measured, how they should be measured and the context in which these assessments occur?
- How do conceptions of data and conceptions of assessment shape policies and procedures?
- How do conceptions of data and conceptions of assessment inform programmatic or curricular decisions?
- How do conceptions of data and conceptions of assessment inform instruction at the classroom level?

Based upon the comments of several participants, it appeared that the task of using WKCE data to inform instruction had been in place for at least five years and that data from a previously mandated state reading test had informed instruction before that.

The data-analysis task was structured so that during the summer a team of teachers in each elementary building analyzed the building data and then made recommendations to the staff on areas that needed improvement. Those recommendations led to building level goals related to improvement on the WKCE and changes in instruction.

For example, the Oakdale reading specialist spoke of expanding the number and scope of the non-fiction selections in the school bookroom in response to the fact that a WKCE item analysis showed that students were weak in interpreting non-fiction. The reading specialists also designed classroom lessons and assessments to support students' growth in non-fiction reading strategies. The Fantastic Four Assessment in Appendix L is an assessment designed to monitor students' understanding of non-fiction reading strategies. This assessment was observed in use in a primary grade classroom where the reading specialist was collaborating with the classroom teacher on a non-fiction reading strategy lesson.

The practice of using WKCE data to inform instruction was also illustrated by an observation of a primary grade classroom at Maple Grove. Students were expected to respond in writing to a passage and then write the page number where they found the information. As an aside to me after the students had been given directions, the Maple Grove teacher said, "We are having them do that because that is getting them ready to find examples from the text when they write their responses for the WKCE." The teacher credited the Maple Grove reading specialist for designing this particular strategy to prepare students for the test.

Overall, the evidence gathered through interviews and observations suggested that the use of WKCE data to inform instruction was fairly low-key. Teachers adjusted

instruction within the balanced literacy curriculum to address topics in which students appeared to be weak based upon WKCE data. The only exception to the practice was that participants reported that formal test prep in Valley View took place a week or two before the test. Teachers used the sample items released by the state to give students an opportunity to get used to the test format. What teachers reported doing appeared to match the philosophy of a phrase spoken by several: "The best test prep is good instruction."

<u>District Level Assessments</u>

District administrators also inherited a culture in which classroom teachers administered a district level reading assessment once a year, in January. Students who did not meet the January benchmark were retested in May. In terms of reading competencies, the district level assessment measured some competences that cannot be measured on a standardized test like the WKCE. Competencies listed in the district curriculum documents included: fluency, operations on print, self-monitoring, cross-checking, self-correcting, and retelling. The district level assessments also measured students' literal and inferential comprehension, competencies also assessed on the WKCE.

Results were recorded in each student's reading record folder, a folder that followed the child from grade to grade. Teachers used the records at the beginning of the year as one piece of information in determining reading groups. (Most teachers spoke of using other informal measures as well, such as listening to students read or asking them to retell a story.) Some reported using the assessment occasionally outside of the mandated testing period to gauge the progress of individual students during the year. This information was then used, along with observational or informal assessment data, to

change the student's group placement and/or determine appropriate classroom or reading specialist interventions. I did not uncover any evidence to suggest that the data from these assessments informed decisions outside of the individual classroom, except in cases where the results showed that a student required reading intervention from the reading specialist.

Another cultural aspect is that teachers expressed varying levels of comfort with this assessment. Experienced teachers spoke highly of it. For example, one primary grade teacher spoke of "loving" the assessment. The teacher noted that it took a number of years to become good at administering the assessment and using the results. The teacher had received staff development on the assessment as part of the training that was provided to teachers who were new to the district. Although it was not required, this teacher assessed all students at the beginning of the year to see who might have grown or who might have regressed in comparison to the previous scores on their reading record. The teacher used the results as a basis for assigning "just right" books and for determining students' immediate instructional needs, both individually and as a group.

In contrast, the two less experienced teachers were less confident about using the results to inform their instruction decisions. This lack of confidence was likely due to the fact that changes in district resources meant that the young teachers had not received the same extensive staff development that was provided to new teachers in the past. In addition, teachers at this grade level administered a group test, which lacked the one-on-one interactions that might have given them clues into the needs of individual readers. As a result, the assessment had little influence over instruction. "It's a struggle for me to

understand what we are supposed to take from this; what we were supposed to do with teaching time because of it," said one of the less experienced teachers.

The principals' reactions to the assessment varied. For example, the Woodland principal called the assessment a "tricky" instrument and was skeptical as to whether all Woodland teachers understood the assessment and how to use the results to inform instruction. The Maple Grove principal was concerned about the time it took to administer the assessment and that the time spent assessing came at the expense of instruction. In contrast, the Oakdale principal valued the assessment and believed that most of the Oakdale staff used the assessments for instructional decisions and placement decisions. The principal also reported that Oakdale teachers talked about the results between grade levels and exchanged information about students' needs.

As a group the reading specialists and some of the experienced primary grade teachers expressed great confidence in the value of the district level reading assessment This assessment provided immediately useful information about a child's reading performance that guided the kind of reading interventions a child might need. This assessment was integral part these participants' professional practice.

Classroom Assessments

Through interviews and observations, I learned about the culture of classroom assessment in Valley View. The data teachers gathered were in the form of running records, anecdotal notes, reading group logs, or rubric scoring systems. These tools enabled them to capture the reading behaviors of individual students or small groups of students and the results informed each teacher's decisions about the next instructional steps. These tools captured some of the same competencies that were measured by the

WKCE and/or the district level assessment—skills ranging from literal and inferential comprehension to fluency to self-correcting. However, a remarkable feature of some of these classroom assessments is that they assessed competencies not assessed by the other instruments. For example, all participants had students keep a log of their reading. This information helped the teachers track the student's pace of reading and reading interests. In another instance, several teachers used observation logs to capture information about the students' use of visualization during their reading partnership discussions. Another competency that was assessed with a rubric was the ability to generate new ideas based upon evidence presented in the story (Appendix N). Still another competency was embedded in a small group discussion rubric (Appendix M) that established the importance of students' abilities to piggyback upon one another's ideas during discussion. It is important to note that some of these assessments were created collaboratively by some of the reading specialists and teachers who happened to participate in this study. These tools were shared across the district, as evidenced by the fact that teachers in buildings separated by some distance used the same assessments. Why were these competencies valued? The answer comes from a Woodland upper elementary grade teacher in the context of her comments about her transition from a new teacher who used reading quizzes to a teacher who was gaining an understanding of the ways in which she might implement the balanced literacy model. She explained that she wanted to involve her students more in talking about books because "that's what adults do when they read a book—they discuss it."

In summary, district administrators inherited a culture in which teachers already used assessment data to inform instruction. Data based upon the reading competencies

measured by WKCE influenced building goals which in turn influenced some

instructional decisions. Data from the reading competencies measured by the district level

reading assessment influenced group placement and intervention decisions. Classroom

assessments measured some of the same competencies captured by the WKCE and/or the

district level reading assessment; however, the classroom assessments also captured

additional competencies that were valued by teachers but not assessed in other ways. This

type of data-driven decision making based upon competencies not assessed in other ways

may set the practices of the Valley View participating teachers apart from the practices of

teachers described in the literature related to data-driven decision-making. This

suggestion is made cautiously because literature reviewed did not describe the

competencies that might have been the focus of portfolios, conferences or student self-

assessments. For example, Gamma students in Illinois received sticky notes that told

"Things I Need to Work On" or "Things I Do Well" (Shelton, et al. 1993, p. 25) but there

is no mention of the competencies embedded in those "things."

The New Initiatives

The fact that some educators in Valley View were already familiar with the

concept of using data to inform instruction provided a foundation for the new initiatives

because teachers were accustomed to the concept of using data to establish goals. The

long range strategic plan expanded upon the way data were used within the district by

establishing systemic expectations in three areas: the implementation of the Measures of

Academic Progress (MAP) assessment; the expectation that WKCE and MAP data would

guide the principals' building goals; and the implementation Professional Learning

Communities (PLC), a staff development structure that fostered professional

collaboration focused on data-driven instruction. The paragraphs that follow will discuss

the ways in which these new district initiatives complement existing practices, as well as

the ways in which they focused professional practice on new ways of using assessment to

inform decisions. This section will also provide some answers to the study's research

questions:

- What reading competencies do districts measure and why?
- Who selects what competencies should be measured, how they should be measured and the context in which these assessments occur?
- How do conceptions of data and conceptions of assessment shape policies and procedures?
- How do conceptions of data and conceptions of assessment inform programmatic or curricular decisions?
- How do conceptions of data and conceptions of assessment inform instruction at the classroom level?

The long range strategic plan heightened the expectations that district

administrators had for building level goals. Past practice had established the principal's

role in establishing building goals and reporting on the achievement of the goals. These

goals were wide ranging. For example, an operational goal might have focused on the

purchase of playground equipment while an achievement goal might have aimed for

improved WKCE math scores. The long-range strategic plan created a climate in which

operational goals were passé and achievement goals de rigueur. A statement made by the

Director captures the top-down nature of this expectation for principals' goals:

> We ask them to have the over-riding goal be linked to WKCE. The
> MAP testing is the progress-monitoring goal, and then they have an
> action plan. It's like we need to say, this is what we are tight on. You
> can be loose about how you get there. And if somebody has some good
> success, then of course we will share the wealth.

One can see how these expectations were translated into action by examining the

some sample goals. For example, one building goal aimed to "increase the number of

students reaching proficient and advanced levels in reading by 2 percentile points on the 2009-10 WKCE." Another goal from the same building referenced the MAP assessment scores. The goal aimed for an increase in points in comparison to the national norms: "Grade 3 is 6 points below the national norm. Our goal is to achieve the national norm by Spring."

The goal form included space for recording Strategies and Action Steps, Responsibility, Timeline, and Evidence of Effectiveness. Some of the strategies and actions steps are worth closer examination because they illustrate the link between Valley View's student achievement goals and professional development. For example, one action step called for teachers to "Collaboratively analyze the results of formative assessments to identify strengths and weaknesses in student learning and in instructional practices." Another addressed the need to "constructively examine present instructional practices and their effect on student learning." These action steps remind one of the kind of work that occurred in effective high poverty schools described by Taylor et al. (2005). Teachers in these schools engaged in faculty-wide discussions about school change, had the opportunity to discuss reading instruction and curriculum, and set goals based upon data. Although Taylor et al. do not discuss the concept of Professional Learning Communities, the work they wrote about appeared to be similar to PLC work.

The work toward achieving building goals was supported by the Professional Learning Community (PLC) model that being formally implemented in the district during the year in which this study was conducted. The district calendar ensured that teachers would have time each month to do the work of a professional learning community: create common assessments based on standards, analyze data, and make instructional decisions based

upon the data. This PLC occurred during three district-wide early release days each month. As discussed in Chapter 3, the scope of this work was communicated to the community through a brochure that listed the essential questions and focus activities that were to guide the PLC work.

We can appreciate the way in which a building goal—one influenced by WKCE and MAP data—can in turn influence professional practice by examining the SMART Goals and PLC Weekly Team Plan from Maple Grove school. A detailed explanation of the documents is in Chapter 3. In brief, SMART goals were established in response to areas identified as weak based upon the results from the WKCE and/or the MAP assessment. The goal was to improve performance through targeted classroom instruction. To that end, teachers created lessons and common formative assessments linked to WKCE/MAP competencies. The lessons were used as part of a whole group or small group intervention with the hope that students would show gains on MAP and/or the WKCE as a result of the intervention. Teachers use a team planning log to record their plans. These plans were submitted to the building principal. Excerpts from some of the plans demonstrate the close connection between test items and instruction. For example, one upper elementary grade team wrote:

- We will use the Descartes item strands to create lessons and assessments for flexible groups. (Descartes is an online tool that helps teachers identify instructional materials matched to students' MAP scores)
- We will research sample questions from the MAP testing to guide instruction.

An excerpt from primary grade team planning log contained this entry:

Looked at WKCE 3rd Grade Testing booklet to see about types of questions and passages that students will be expected to read and answer. Also discussed how we can develop similar questions to help students develop the

needed skills to answer these types of questions. We discussed vocabulary and inferential questions. We also talked about how language based games can help develop critical thinking and oral language. We are contemplating whether we should follow the format of a test having a "bubble" test as an added formative assessment.

It is critical to note that the reading competencies measured by the MAP mirrored those measured by the WKCE. This tight connection helps us understand the relationship between district level decisions and assessment and instruction.

When district administrators selected MAP as a required district assessment, the result of this decision was that, by default, the reading competencies measured in the district matched the competencies measured by the WKCE. In establishing these competencies as the premier competencies to be measured, district administrators fostered conditions that might lead to success on a key objective of the strategic plan: "All students will demonstrate proficiency in the Wisconsin Model Academic Standards." (The WKCE measured proficiency in terms of the standards; hence proficiency on the WKCE indicated proficiency on the standards.)

As we end this discussion it is important to a highlight a key characteristic that set Valley View apart from districts and schools cited in the literature. Valley View had implemented MAP, which served as a formal process for tracking student achievement, throughout the year. Although schools cited in the literature tracked progress toward identified goals (Kerchval and Newbill, 2002, Charles A. Dana Center, 2001), no mention is made of any school or district using an electronic tool like MAP which automatically compiled and disaggregated data in real time. Furthermore, MAP provided data three times per year on a group of students both as a class and as individual students. Teachers were able to see which students were making acceptable grade level progress as

determined by the test and which students struggled with specific concepts. This kind of

feedback shaped the way some teachers planned within their PLC teams. We saw

evidence of that earlier in a log entry cited above. The entry said:

- We will use the Descartes item strands to create lessons and assessments for flexible groups. (Descartes is an online tool that accompanies the MAP assessment. The tool helps teachers identify instructional materials matched to students' MAP scores)
- We will research sample questions from the MAP testing to guide instruction.

At this point it is impossible to speculate about the degree to which the MAP

competencies would influence future instruction across Valley View because at the time

of the study, the participants at Oakdale and Woodland were just beginning to learn about

the ways in which the MAP assessment could be used within their respective buildings.

However, comments made by the Maple Grove principal demonstrate the test's impact on

that particular school's culture. "MAP testing, I think, has single-handedly changed the

way teachers set goals and go about it. They can't wait for the data now, thanks to MAP,

because it is real; it is in real time." Recalling teachers' reactions to the September

testing, the principal said:

> Teachers were walking around with their little notepads seeing what the kids' scores were because they now realize that when they go in there they can figure out by strand [students' strengths and weaknesses]. What was great for the teachers is that they could go, ok, my student is at 223 so he or she has most of this down over here. Now I want to go and see what I can do to extend the children.

The principal also spoke about teachers' reactions to the January testing:

> Our building retested in reading in January and they could not wait to see which kids made gains and which kids didn't. With our MAP testing the strands we worked on we improved on, which was great for teachers to see. Our MAP scores were phenomenal.

This comment is worth noting because of a point made earlier: MAP tested the same reading competencies measured by the WKCE. The main difference between the two is the frequency of administration and time of year when the test is administered. The once a year WKCE is a strictly summative measure. On the other hand, the MAP can be used as a summative assessment at the end of the year and as formative assessment that guides instruction throughout the year. This leads one to wonder about the degree to which MAP will influence instruction in the future. One can speculate that MAP/WKCE competencies might eventually shape instruction to such a degree that competencies measured by some of teacher-designed classroom assessments would become passé. If that happens the content measured by high-stakes tests would become the content that is taught.

Limitations and Deficiencies of the Study

There are several factors that contribute to the limitations and deficiencies inherent in this study. First of all the study was conducted at a time when the Valley View School District was in a state of change. Change is not a linear process, a factor acknowledged by the Superintendent who assessed the process across the district by saying, "It's awesome in some places already; others are struggling. Administrators are different in how they operate through the system so it's not a linear path for sure." In terms of conducting research, this non-linear aspect of change meant that participants who volunteered to participate in the study were figuring things out; they were growing and changing in their understandings of assessment and the ways in which they used it at the same time the research was being conducted. If the research design had allowed for multiple interviews over time, I would have been able to capture how each participant's

perspective had evolved. The inclusion of a member check would also have allowed me to further probe participants' perspectives and might have led to changes or refinements in the way I analyzed and reported the results.

My decision to situate this study within the context of NCLB is also a factor that must be considered in a discussion of limitations and deficiencies. Although I sought information from multiple sources—district documents, interviews, and documents and artifacts shared by the participants—it is possible that I filtered information through my NCLB lens and therefore failed to capture all the key attributes of the Valley View culture and the ways in which data was used to inform instruction.

Despite the limitations and deficiencies in this study, I believe it can serve as a catalyst for generating a discussion which might move the field of assessment forward. With this thought in mind, I created a model designed to expand the ways in which one might think about assessment in public education. It seems that the challenge for researchers and practitioners is to devise an assessment system that serves traditional accountability purposes and holds teachers accountable for developing students who are life-long readers and deep thinkers. The model, which was influenced by my review of literature in Chapter 2, as well as by my findings in Valley View, provides the backdrop for a thought experiment conducted in Chapter 6.

Assessment in Public Education: A Model

The Assessment Categories in Public Education model retains the distinction between summative and formative assessment as described by Stiggens et al. (2004) and Popham (2008). I retained the distinction because I wanted to reinforce the idea that the term *summative* refers to measures that take place at a fixed point in time for the purpose

of accountability, while *formative* assessment is an on-going part of instruction that provides feedback to the teacher and the students in relation to the targeted goals.

The unique feature of my model is that the summative and formative categories occupy two different strands, the Pragmatic Strand and the Transformative Strand. I created these strands to illustrate an argument: When we conceptualize assessments as either summative or formative assessment we run the risk of not thinking deeply enough about the range of competencies the assessments measure and the kind of data we collect as result of the assessment. This is particularly worrisome in a high-stakes era where the realities of accountability may serve to promote a test-driven curriculum.

The need for two strands was inspired by Serafini's (2000) concept of *Assessment as Procedure*. As he defines it, this type of assessment stems from an external mandate; it is a measure pressed upon students and teachers. Consequently, "teachers are burdened with another set of procedures given to them by their administration in order to provide scores for an external authority (p. 210)." The Pragmatic and Transformative strands in my model serve to spark awareness of this possibility. Table 5.1 provides details about the strands. Further explanation follows the table.

Table 5.1 Reading Assessment Categories in Public Education

	Summative Assessments	Formative Assessments
Strand 1: Pragmatic Purpose	**Purpose:** Track and report student *achievement* annually. **Measures**: Literacy competencies linked to predefined standards. **Instrument:** Externally developed standardized instruments. **Timeline:** Fixed **Reporting:** Publicly reported annually. **Data Informs:** Improvement plans with an eye toward increased student performance on the assessment. **Data Tracking::** Year-to-year by individual, class or grade level, cohort group.	**Purpose:** Assess student *growth* over time. **Measures**: Literacy competencies linked to predefined standards. **Instrument:** Externally developed standardized instruments; or teacher selected or created. **Timeline**: Fixed for progress monitoring; flexible for on-going assessments. **Reporting**: On-going. Report growth internally; externally, if required. **Data Informs:** Teaching. Student self-assessment and goal-setting expected. **Data Tracking**: On-going by individual, class or grade level, cohort group.
Strand 2: Transformative Purpose	**Purpose:** Track and report student *achievement* annually. **Measures:** A wide range of literacy competencies and non-cognitive traits linked to predefined standards. Includes competencies in the Pragmatic/ Summative quadrant above. **Timeline:** Fixed. **Reporting:** Publicly reported annually **Data Informs:** Improvement plans with an eye toward increased cognitive and non-cognitive competencies. **Data Tracking::** Year-to-year by individual, class or grade level, cohort group.	**Purpose**: Assess student *growth* over time. **Measures**: A wide range of literacy competencies linked to predefined standards and non-cognitive traits. **Instrument:** Externally developed standardized instruments; or teacher selected or created. **Timeline**: Fixed for progress monitoring; flexible for on-going assessments. **Reporting**: Monitor growth; report internally, where appropriate. **Data Informs**: Teaching. Student self-assessment and goal-setting expected. **Data Tracking**: On-going by individual, class or grade level, cohort group.

The Pragmatic Strand might just as easily be labeled the High-Stakes Strand.

However, I selected the word *pragmatic* to label the strand because I believe the word

captures reality: schools have no choice but to respond to annual accountability demands.

Assessments in this category provide one snapshot of student achievement that is

reported publically. Results drive improvement plans which are focused on improving

student achievement as measured by the assessments in this quadrant.

Moving clockwise to the Pragmatic/Formative slot, notice that content of the

assessment remains the same; it is dictated by the same competencies measured in the

Pragmatic/Summative quadrant. The primary distinction is the use of the word *growth*. The implication here is that the assessments are on-going and are used to inform teaching and learning, including student self-assessment and goal-setting. All efforts are focused on skills measured with assessments used in the Pragmatic/Summative quadrant.

The strand labeled *Transformative Purpose* serves to highlight the need to transform public education into a system that provides opportunities for all students to learn at high levels. For that reason, assessments in both quadrants of this strand measure students' growth on a wide range of competencies, including non-cognitive traits. Competencies in the Transformative/Formative quadrant are measured in order to ascertain progress. The assessments are on-going and are used to inform teaching and learning. Student self-assessment and goal-setting on cognitive and non-cognitive traits are a critical aspect of this quadrant. The content of the assessments in the Transformative/Summative quadrant mirror those in the Transformative/Formative slot. The difference is that achievement of cognitive and non-cognitive competencies is publically reported. The data inform improvement plans linked to the range of competencies and traits that were measured.

The Transformative Strand requires more of districts, of teachers, and of students because these assessments capture achievement on a wide range of competencies. The Transformative Strand recognizes the reality of high-stakes tests but expands the scope of what is assessed to encompass more than can be measured with standardized instruments. Several authors inspired the Transformative strand. The importance of measuring a wide range of competencies, including non-cognitive traits was inspired by Afflerbach (2007b) who contends that "accomplished readers share important characteristics that are related

to but different from their cognitive skills, strategies, and achievement" (p. 156). He

charges that our assessment agenda is incomplete when we ignore how students develop

in terms of motivation, perseverance, and self-esteem because these factors can either be

either powerful facilitators or obstacles to a child's reading development. Without this

information, he argues we "limit the inferences we can make about students' reading

development, teacher effectiveness, and the value of the reading curriculum" (p. 156).

Allington (2006) also influenced the scope of the Transformative Strand through his

comments about the reader who succeeds with higher level literacy tasks such as the

ability to "search and sort through information, synthesize and analyze information, and

summarize and evaluate the information." Allington also prompts us to consider that it

might be worthwhile to know if our student-reader is one who is not only able to read

difficult texts but one who chooses to read a variety of texts in his or her spare time.

Additionally, Johnston (1999) provided inspiration with his comments about the reader

who is able to engage in "collaborative sense making" and can use others' literate

thinking as a tool to extend both individual and collective thought"(Johnston, p. 19).

Data Driven Decision-Making:
A Case Study of How a School District Uses Data to Inform Reading Instruction

Chapter 6: A Thought Experiment

At the end of Chapter 5, I proposed a model, Assessment Purposes in Public Education, as a way to think about varying conceptions of data and assessment within a high stakes climate. In this chapter, I will discuss the types of assessments that I learned about in Valley View in light of the proposed model. I will then use my analysis of these assessment categories, along with my model of Assessment Purposes in Public Education, as the basis for a thought experiment[2]. In this experiment, I will hypothesize one way in which Valley View might evolve as a data-driven district despite the public accountability mandates imposed by either NCLB or future high stakes legislation.

Assessment Categories in Valley View

I want to emphasize that I am using this the model of Assessment Purposes in Public Education to categorize what I observed or discussed with participants within a very limited timeframe. Participants may have not mentioned assessments that they use within other units or at other times of the year. Or, perhaps within such a limited timeframe, I failed to probe deeply enough about the district's assessment practices and therefore failed to uncover key information. The context in which I conducted my research might have also influenced observations or discussions. Not only was this study situated within the context of NCLB, it was also conducted during a time when specific top-down changes were being implemented. Principals faced increased accountability for setting and monitoring academic goals. The MAP assessment was in its first full year of implementation. Teachers were becoming accustomed to the collaborative work called

---

[2]A thought experiment is a cognitive process first used by philosophers in Presocratic Greece for developing their ideas. (Rescher, 1991).

for by the PLC model and were working together to plan instruction and create

assessments in light of MAP data. The district was in the process of adopting a new core

reading series, a step that would likely be accompanied with assessments matched to that

series. For all these reasons, the assessments listed in the chart below are a snapshot; they

may not represent the assessments used in Valley View when the study was conducted.

Table 6.1 Assessment Categories in Valley View

	Summative	**Formative**
Pragmatic	WKCE MAP	MAP Common Assessments linked to MAP or WKCE competencies
Transformative		MAP Common Assessments District Level Reading Assessment Classroom Assessments reflecting additional competencies, i.e. Book talk Discussion Rubric Student Reading Log Literature Response Rubric

I placed the names of the assessments discussed or observed over the course of

the study into the quadrants to demonstrate the prominent characteristics of these

assessments. The WKCE naturally falls into the Pragmatic/Summative slot because it is a

state-mandated accountability measure. I also placed the district-mandated MAP

assessment here because the MAP served as an end-of-year accountability measure

within Valley View. Both WKCE and MAP results were publically reported. Data from

each measure informed building achievement goals which principals presented to the

school board in Fall and reported on in early Summer. The district used both the WKCE

and the MAP data as a way to track the achievement of groups of students over time—as

grade level groups and cohort groups per building and across the district.

The MAP assessment also appears in the Pragmatic/Formative quadrant because

Fall MAP data informed teachers' instructional decision with the assumption that

students would score better on the Winter MAP as a result of those decisions. Common

assessments share the slot with MAP because common assessments were used as both

baseline measures and interim measures to determine if students were making progress as

a result of the intervention determined by MAP.

The Transformative/Formative quadrant accommodates standards-based measures

of reading achievement, as well as authentic measures that assess, in the words of one

Valley View teacher, "what real readers do.". This quadrant provides a spot for the

assessments designed collaboratively by reading specialists and teachers. It captures the

kind of assessment Lucy Calkins (2001) writes about:

> Assessment is the thinking teacher's mind work. Assessment is the stance
> that allows us to learn from our students and thus to teach them. Assessment
> is the compass from which we find our bearings and chart our course, the
> map on which we do this. Assessment is also the thinking student's work. It
> is the student's growing awareness of what it means to do good work and his
> or her own sense of progress, goals and next steps. Clear goals and honest
> frequent assessment, including self-assessment allow students to become
> managers and authors of their own learning lives. (p. 137).

I placed the district level reading assessments into the Transformative/Formative

quadrant as well. The assessments measured skills that, if mastered, would likely lead to

a strong performance both on the MAP and the WKCE. However, these benchmark-style

assessments relied on teacher judgment/expertise both for administering and scoring. In

terms of tracking and reporting, the district level assessment benchmarked each student's

reading achievement from year to year, or, in the case of struggling students, from one

testing period to the next. Results from this assessment were not a factor in building goal

setting or public reporting. They informed students' placement in guided reading groups,

and guided interventions by the classroom teacher or reading specialist. If they existed, on-going assessments that measured other reader competencies and non-cognitive traits would also be placed in this quadrant.

The Transformative/Summative quadrant is blank. Recall that this quadrant calls for a range of cognitive and non-cognitive assessments with the results tracked by grade level or cohort group. These results inform improvement plans with an eye toward increasing cognitive and non-cognitive competencies. Valley View goal sheets focus only on achievement as measured by the WKCE and MAP. In order to meet the criteria for the Transformative/Summative quadrant, other competencies would have to be identified and tracked as part of the building goals. They were not.

A Thought Experiment

The state of Valley View at the time of this study provides the backdrop for this thought experiment. Recall that the Valley View School District was in the process of transforming itself from a "district of schools" with strong site autonomy to a "school district" with uniform expectations for goal-setting, testing, and professional collaboration focused on data. The district's long range strategic plan guided the change process, a process which focused on student learning to the degree that an objective in the strategic plan called for "All students [to] demonstrate proficiency in the Wisconsin Model Academic Standards." District initiatives that emerged as a result of the long-range strategic plan included the implementation of the Measures of Academic Progress (MAP) assessment, specific expectations of principals in terms of using WKCE data and MAP data to establish building goals, and the implementation of Professional Learning Communities (PLC), a staff development structure that fostered professional

collaboration focused on data-driven instruction that was informed by the WKCE and

MAP. It is important to remember that the WKCE measured proficiency in the Wisconsin

Model Academic Standards and the MAP assessment was aligned to the WKCE.

Consequently, it is not surprising that Valley View's Transformative/Summative

quadrant is empty.

The purpose of this thought experiment is to imagine what might happen if Valley

View expanded the scope of the assessments used within the district and created a menu

of assessments to fill the Transformative/Summative quadrant. In doing, it could become

a district that graduated students who both demonstrated proficiency on the Wisconsin

Model Academic Standards, as well as students who demonstrated the following

characteristics: motivation, perseverance, and self-esteem (Afflerbach, 2007b); the ability

to read difficult texts and that they chose to read a variety of texts in their spare time; the

ability to search and sort through information, synthesize and analyze information, and

summarize and evaluate information (Allington, 2006); and the ability to engage in

collaborative sense making as well as the ability to use others' literate thinking as a tool

to extend both individual and collective thought (Johnston, 1999).

Let's proceed with the thought experiment by imagining what might happen if

Valley View moved in the Summative/Transformative direction with existing parameters

in place: The district maintains its top-down decision-making structure, retains the MAP

assessment, the expectations for building goals, and practice of using the PLC model for

data-driven decision making. Hypothetically, here is one scenario that could play out.

In this scenario, imagine that Valley View district administrators revisited what

they believed about purpose of assessment. In doing so, they acknowledged that high

stakes testing was a reality even in a relatively high performing district like theirs, and that it was important for Valley View students to continue to perform well. At the same time they recognized that they spoke about the value of on-going classroom assessments and believed that the primary purpose of assessment was to guide instruction. In this respect, the district administrators were pleased to hear from the building principals that teachers, working in their PLC teams were creating formative classroom assessments linked to the content of the WKCE and MAP, and that these formative assessments informed daily instruction. Administrators also learned from the principals that teachers were using other kinds of classroom assessments, assessments the teachers claimed were designed to capture some of the things "real readers do." When Valley View administrators investigated the nature of these assessments, they began to wonder if there was any grade level or cohort data to show that students in Valley View had attained some of these competencies. There wasn't.

This fact propelled Valley View administrators into action. They considered the ramifications of the organizational change they had envisioned, a change captured by a remark made by the Superintendent: "Teaching to the test, the way it is set up structurally, is the way we are trying to move our organization." In response, they revised their long-range strategic plan so that it called for all students to achieve proficiency in competencies measured by the WKCE and MAP and to demonstrate proficiency in other reading competencies similar to those demonstrated by capable adult readers.

As a result of this adjustment to the plan, Valley View moved from a Pragmatic district to a Transformative district. Valley View expanded its view of excellence by implementing a full menu of assessments—a menu of assessments that measured more

than the key skills and strategies valued in a high stakes environment. Valley View's revised assessment program generated ample grade level and cohort data about students motivation, persistence and self-esteem as readers; their habits of reading widely on their own time; their capacity to accomplish higher level literacy tasks; their ability to engage in collaborative sense making and their ability to use others' literate thinking as a tool to extend both individual and collective thought (Johnston, 1999, p.19). To complement the new assessment program Valley View adjusted its goal setting process so that it would capture growth and achievement in all the assessed areas. In an era of high stakes accountability, Valley View allowed itself to become transformed.

Recommendations for Further Study

Will that scenario play out in Valley View? It's hard to predict. Electronic assessments are becoming more prevalent across the country. Within the last year, the Wisconsin Department of Public Instruction has decided to abandon the WKCE in favor of what will likely be an electronic assessment. Some say the assessment will be administered more than once a year and that it will provide information about student growth, similar to what is provided by an assessment like the MAP. If that comes to bear, what will happen within a district like Valley View? Will the district get caught in an accountability loop that traps it in the Pragmatic strand? Or will the district stretch itself and strive for transformative assessment practices that nurture students who are capable of using reading as a tool to extend both individual and collective thought? It is hoped these questions will be the catalyst for a larger discussion of how schools might use data-driven instruction with the end result being that we become more thoughtful of how we use data to guide instruction that truly improves reading achievement.

Bibliography

Afflerbach, P., Kapinus B, & Winograd, P. (1994). Developing alternative assessments: Six problems worth solving. *The Reading Teacher*, 47(5), 420-423.

Afflerbach, P. (2007a, February). *High Stakes testing and literary learning: Where do we go from here?* Paper presented at the Wisconsin State Reading Association, Milwaukee,WI.

Afflerbach, P. (2007b). *Understanding and using reading assessment K-12*. Newark, DE: International Reading Association.

Allington, R. L. (2006). *What really matters for struggling readers: Designing research-based programs*. Boston: Pearson Education, Inc.

Allington, R. L. (2002). What I've learned about effective reading instruction from a decade of studying exemplary classroom teachers. *Phi Delta Kappan,* 83(10), 740-747

Alverman, D.E. (2003). *Seeing themselves as capable and engaged readers: Adolescents and re/mediated instruction*. Naperville, IL: Learning Point Associates.

Amrein, A. L. & Berliner, D.C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18), Retrieved April 21, 2006 from http: epaa.asu.edu/eppa/v10n18/

Atherton, J.S. (2005). *Learning and teaching; what works and what doesn't.* Retrieved November, 2007 from http://www.learning and teaching.infor/ teaching/ what_works.htm.

Bailey, A.L. & Heritage, M. (2008). *Formative assessment for literacy grades K-6: Building reading and academic language skills across the curriculum*: Thousand Oaks, CA: Corwin Press.

Bintz, William (2000). Using freewriting to assess reading comprehension. *Reading Horizons*, 40 (3), 205-223.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education,* 5(1), pp.7-74.

Black, P. & Wiliam, D. (1998). Inside the blackbox: Raising standards through classroom assessment,. *Phi Delta Kappan,* 80(2), 139-148.

Black, P. Harrison, C., Lee, C., Marshall, B, & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 9-21.

Bogdan, R.C. & Biklin, S.K. (1998). *Qualitative research for education: An introduction to theories and methods* (3rd ed.). Boston: Allyn and Bacon.

Booth, D. and Rowsell, J. (2002). *The literacy principal: Leading, supporting, and assessing reading and writing initiatives*. Markham, Ontario: Pembroke Publishers Limited

Boston, C. (2003). Effect size and meta-analysis. *Eric Digest*. College Park, MD: Eric Clearing House on Assessment and Evaluation. (ED470590)

Brozo, W. G. & Hargis, C.(2003). Using low-stakes reading assessment. *Educational Leadership*, 61(3), 60-64.

Caldwell, J.S. (2002). *Reading assessment: A primer for teachers and tutors*. New York: The Guilford Press.

Calkins, L.M. (2001). *The art of teaching reading*. New York: Longman

Campbell, M.B.(2001). Inquiry into reading assessment: Teachers' perceptions of effectiveness. Reading Horizons, 41(1), 1-20.

Charles A. Dana Center (2001). *Opening Doors: Promising lessons from five Texas high schools.* Austin, TX: University of Texas at Austin, Charles Dana Center.

Cobb, C. (1997). Improving schools study: Another way to look at raising student achievement (A preliminary study). Raleigh, NC: Evaluation Section Division of Accountability Services North Carolina Department of Public Instruction Retrieved Winter, 2006 from www.eric.org/ericdb/ED421482.htm

Creswell, J. W. (2003). *Research design: qualitative, quantitative, and mixed method approaches.* Thousand Oaks, CA: Sage Publications.

Eight-Step Instructional Process (2007). Retrieved July 17, 2007 from http://brookview. warren.k12. in.us/instructional/8step.asp.

Farr, R. (1992). Putting it all together: Solving the reading assessment puzzle. In M. Baker (Ed.). (2004*). Preparing reading professionals: A collection from the International Reading Association*: Newark, DE: International Reading Association.

Fiene, J., & McMahon, S.(2007). Assessing comprehension: A classroom-based process. *The Reading Teacher*, 60(5), 406-417.

Foertsch, M. (1998). *A study of reading practices, instruction, and achievement in district 31 schools.* Oak Brook, IL: North Central Regional Educational Laboratory.

Fox, D.(2001). Three kinds of data for decisions about reading. In L.O.Simons, (Ed.), *The newsletter of the comprehensive center, Region VI,* 6(1), 11-13. Madison, WI: Wisconsin Center for Education Research

Garica, G.E. (2000), Bilingual children's reading. In M.L.Kamil, P.B. Mosenthal, P.D. Pearson, & R.Barr, R. (Eds.), *Handbook of reading research* Vol. 3 (pp. 813-834). Mahwah, NJ: Lawrence Erlbaum Associates.

Glazer, S. M. and Brown, C. S. (1993). *Portfolios and beyond: Collaborative assessments in reading and writing.* Norwood, MA: Christopher Gordon Publishers, Inc.

Glesne , C. (1999). *Becoming qualitative researchers: An introduction.* New York: Longman.

Gove, M. K. (2006, May). *Classroom teachers' perceptions of the DIBELS assessments.* Poster Session presented at the meeting of the International Reading Association, 51st Annual Convention, Chicago, IL.

Guthrie, J.T.. Schafer, W.D., & Huang, C.W. (2001). Benefits of opportunity to read and balanced instruction on the NAEP. [Electronic version].*The Journal of Educational Research*, 94 (3), 145-162

Guthrie, J.T. & Wigfield, A.(2000), Engagement and motivation in reading. In M.L.Kamil, P.B. Mosenthal, P.D. Pearson, & R.Barr, R. (Eds.), *Handbook of reading research* Vol. 3 (pp. 403-422). Mahwah, NJ: Lawrence Erlbaum Associates.

Hill, B.C. (2001). *Developmental continuums: A framework for literacy instruction and assessment K-8.* Norwood, MA: Christopher Gordon Publishers.

International Reading Association & National Council of Teachers of English (2007). *Standards for the assessment of reading and writing.* Retrieved November, 2009 from http://www.reading.org/General/CurrentResearch/Standards/ AssessmentStandards.aspx.

International Reading Association (1999). *High-stakes assessment in reading: A position statement of the international reading association*[Brochure]. Newark, DE: International Reading Association.

Invernizzi, M.A., Landrum, T.A., Howell, J.L., & Warley, H.P. (2005). Toward a peaceful coexistence of test developers, policy makers, and teachers in an era of accountability. *The Reading Teacher*, 58(7), 610-618.

Johnson, J. H.(1997). *Data-driven school improvement*. (Eric Digest, Number 109). Retrieved May 9, 2006 from www.eric.org/ericdb/ED401595.htm

Johnston, P. (1999). Unpacking "literate" achievement. (CELA Research Report Number 12007*).* Albany, NY: Center on English Learning and Achievement. ED427336

Johnston, P. & Costello, P. (2005). Principles for literacy assessment. *Reading Research Quarterly,* 40(2), 256-267.

Kame'enui, E.J.(2002). *An analysis of reading assessment instruments for K-3.* Eugene, OR: Institute for the Development of Educational Achievement, College of Education, University of Oregon.

Kercheval, A. & Newbill, S. (2002). A case study of key effective practices in Ohio's improved schools. The Indiana Center for Evaluation. Retrieved March, 2006 from http:/www.indiana.edu/~iuice

Kerr, K.A., Marsh, J.A., Ikemoto, G.S., Darilek, J. & Barney H. (2006, June). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts [Electronically published]. *American Journal of Education,* 112, 496-520.

Leslie, L. & Caldwell, J. (2001). *Qualitative Reading Inventory III*. New York: Longman.

Leslie, L. & Jett-Simpson, M. (Eds.). (1997). *Authentic literacy assessment: An ecological approach*. New York: Addison Wesley Longman.

Lincoln Y.S.. & Guba, E.G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications

Lyons, C.A. & Pinnell, G. S.(2003). Reading and writing assessment. *National Staff Development Council*, 24(2), 23-31.

Marzano, R.J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R.J., Pickering, D.J., & Pollock, J.E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

McGill-Franzen, A. (2000), The role of text in classroom reading. In M.L.Kamil, P.B. Mosenthal, P.D. Pearson, & R.Barr, R. (Eds.), *Handbook of reading research* Vol. 3 (pp. 889-908). Mahwah, NJ: Lawrence Erlbaum Associates.

McNamara, M. (2003). Reading assessment options for teachers of junior secondary students. In M. Baker (Ed.). (2004*). Preparing reading professionals: A collection from the International Reading Association*: Newark, DE: International Reading Association.

McKenna, M.C. & Stahl, S.A. (2003). *Assessment for reading instruction*. New York: The Guilford Press.

McKenna, M.C. & Walpole, S. (2005). How well does assessment inform our reading instruction? *The Reading Teacher*, 59(1), 84-86.

Meisels, S.J., Atkins-Burnett, S., Xue, Y., Nicholson, J., Bickel, D.D. and Son, S.H. (2003). Creating a system of accountability: The impact of instructional assessment on elementary children's achievement of test scores, *Educational Policy Analysis Archives*, 11(9). Retrieved from http:eppa.asu.edu/eppa/v11n9

Merriam, S. B. (1998). *Qualitative Research and Case Study Applications in Education.* San Francisco: Jossey-Bass.

Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Miller, L.P. (1992). Alternatives to norm-referenced testing for assessing student achievement. *ERS Spectrum Summer, 1992*, 10(3), 3-8.

Mitchell K.J.(1997). What happens when school reform and accountability testing meet? *Theory into Practice*, 36(4), 262-265. Retrieved November 28, 2006 from http://jstor.org

Mokhtari, K. & Reichard, C.A. (2002). Assessing students' metacognitive awareness of reading strategies. [Electronic version]. *Journal of Educational Psychology*, 94(2),

Mokhtari, K., Rosemary, C. A., & Edwards, P. A.(2007). Making instructional decisions based on data: What, how and why. *The Reading Teacher*, 61(4), 354-359.

National Research Council (2001). *Knowing what students know: The science and design of educational assessment.* Committee on the Foundations of Assessment. J.Pelligrono, N. Chudowsky, & R.Glaser (Eds.). Board on Testing and Assessment, Center for Education. Washington, DC: National Academy Press.

Nelley, E. & Smith, A. (2000). *Rigby PM benchmark kit: A reading assessment resource for grades K-5*. Barrington, IL: Rigby

Oldfather, P. (1994). When students do not feel motivated for literacy learning: How a responsive classroom culture helps. (Reading Research Report No. 8). Retrieved Winter 2006 from www.eric.org/ericdb/ ED365996.htm

Paris, S. G. & Hoffman, J.V. (2004). Reading assessments in kindergarten through third grade: Findings from the center for the improvement of early reading

achievement. [Electronic version]. *Elementary School Journal*, 105 (2), 199-217.

Pearson, P. D. (2005). In S.G.Paris & S. A. & Stahl (Eds.). (2005). *Children's reading comprehension and assessment*. CIERA Center for the Improvement of Early Reading Literacy. Mahwah, New Jersey: Lawrence Erlbaum.

Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Pressley, M. (2006). *Reading instruction that works: the case for balanced teaching* (3rd ed.). New York: The Guilford Press

Pressley, M., Wharton-McDonald, R., Allington, R. Block, C.C., & Morrow, L. (1998). *The nature of effective first-grade literacy instruction*. (CELA Research Report Number 11007). Albany, NY: The National Center on English Learning & Achievement. Retrieved April, 2006 from www.eric.org/ericdb/ED427338.htm

Pressley, M., Yokoi, L., Rankin, J, Wharton-McDonald, R., & Mistretta-Hampton, J. (1997). *A survey of the instructional practices of Grade 5 teachers nominated as effective in promoting literacy*. (Reading Research Report No. 85). National Reading Research Center. Retrieved April, 2006 from www.eric.org/ ericdb/ ED405558.htm

Reeves, Douglas B. (2000) *Accountability in Action: A Blueprint for Learning Organizations.* Denver Colorado: Advanced Learning Press

Rescher, N. (1991) Thought experimentation in pre-Socratic philosophy. In T. Horowitz & G.J. Massey (eds.). *Thought experiments in science and philosophy*. Savage, MD: Rowman and Littlefield. Retrieved February 2010 from http://www.pitt.edu/ ~pittcntr/About/center_publications/ thought_experiments/ thought_experiments_ toc.htm

Richardson, V., Anders, P., Tidwell, D. & Lloyd, C. (1991). The relationship between teachers' beliefs and practices in reading comprehension instruction. *American Educational Research Journal*, 28(3), 559-586.

Rumery, Karen (2000). *Using data to inform instruction-stories from five states: "Maine" the way life should be.* Paper presented at the annual meeting of the National Council of Teachers of English, Milwaukee, WI. Retrieved April 2006 from www.eric.org/ericdb/ED452540.htm

Shellard, Elizabeth and Protheroe, Nancy (2001). *What principals need to know about teaching reading.* Alexandria, VA: National Association of Elementary School Principals and Educational Research Service. Retrieved Winter, 2006 from www.eric.org/ericdb/ED468343.htm

Shellard, Elizabeth G. (2003). Using assessment to support reading instruction. *Principal*, 83(2), 40-34 (EJ676118).

Snow, C.E. (2002). *Reading for understanding: toward a research and development program in reading comprehension*. Santa Monica, CA: RAND

Stephens, Diane, Pearson, P. D., Stallman, A., Shelton J., Commeyras, M., Roe, M.,Gilrane, C.P.((1993). Assessment and decision-making in Alpha. (Technical Report No. 589). Urbana, IL: Center for the Study of Reading. Retrieved Summer, 2007 from www.eric.org/ericdb/ED362836.htm

Stephens, D., Pearson, P. D., Gilrane, C., Stallman, A., Roe, M., Shelton J., Stallman, A., Commeyras, M. (1995). Assessment and decision-making in schools: A cross-site analysis. (Technical Report No. 614). Urbana, IL: Center for the Study of Reading. Retrieved Summer, 2007 from www.eric.org/ericdb/ED381757.htm

Stephens, D., Rodriguez, A., Commeyras, M., Stallman, A., Shelton, J.,Pearson, P.D.… Gilrane, C. P. (1993). Assessment and decision-making in Delta. (Technical Report No. 591). Urbana, IL: Center for the Study of Reading. Retrieved Summer, 2007 from www.eric.org/ericdb/ED362838.htm

Shelton, J., Stephens, D., Stallman, A., Commeyras, M., Pearson, P.D., Roe, M.,Gilrane, C.P.(1993). Assessment and decision-making in Gamma. (Technical Report No. 592). Urbana, IL: Center for the Study of Reading. Retrieved Spring, 2006 from www.eric.org/ericdb/ED362839.htm

Serafini, F. (2000/01). Three paradigms of assessment: measurement, procedure, and inquiry. In M. Baker (Ed.). (2004). *Preparing reading professionals: A collection from the International Reading Association*: Newark, DE: International Reading Association.

Simons, L.O. (Ed.) (2001). *The newsletter of the comprehensive center, Region VI*. Madison, WI: Wisconsin Center for Education Research

Reading Success Network (2002). *Taking a reading: a teacher's guide to reading assessment*. Downey, CA: Southern California Comprehensive Assistance Center Region XII

Spradley, J.P. (1979). *The ethnographic interview*. New York: Holt, Rinehart, Winston.

Stiggins, R. J, Arter, J. A., Chappuis, J. & Chappuis, S. (2004). *Classroom assessment for student learning: doing it right-using it well*. Portland, Oregon: Assessment Training Institute.

Taylor, B., Pearson, P. D., Clark, K., & Walpole, S. (1999). Effective schools and accomplished teachers. *The Reading Teacher*, October, 1999, 53(2), 156-164.

Taylor, B., Pearson, P. D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: lesson about primary grade reading instruction in low-income schools. In B.M. Taylor, & P.D. Pearson(Eds.). (2002). *Teaching Reading: Effective Schools, Accomplished Teachers.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Taylor, B. M., Pressley, M. & Pearson, P. D. (2000). *Effective teachers and schools: trends across recent studies.* Ann Arbor, MI: Center for the Improvement of Early Reading Achievement. Retrieved Winter 2006 from www.eric.org/ericdb/ ED450353.htm

Taylor, B. M., Pressley M. & Pearson, P. D. (2000). *Effective teachers and schools: trends across recent studies.* Ann Arbor, MI: Center for the Improvement of Early Reading Achievement. Retrieved Winter 2006 from www.eric.org/ericdb/ ED450353.htm

Taylor, B. & Pearson, P. D. (2005). Using study groups and reading assessment data to improve reading instruction within a school. In S.G. Paris & S.A. Stahl (2005). *Children's reading comprehension and assessment.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Taylor, B. M., Pearson, P. D., Peterson, D. S. and Rodriguez, M. C. (2005). The CIERA School Change Framework: An evidence-based approach to professional development and school reading improvement. *Reading Research Quarterly,* 40 (1), 40-69

Thorn, C.A. (2002). *Data use in the school and classroom: The challenge of implementing data-based decision making inside schools.* (WCER Working Paper No. 2002-2). Madison, WI: Wisconsin Center for Education Research.

Valencia, S.W. & Wixson, KW. (2000), Policy-oriented research on literacy standards and assessment. In M.L.Kamil, P.B. Mosenthal, P.D. Pearson, & R.Barr, R. (Eds.), *Handbook of reading research* Vol. 3 (pp. 909-935). Mahwah, NJ: Lawrence Erlbaum Associates.

Vaughn, S. & Briggs, K.L. (2003). *Reading in the classroom*: *Systems for observation of teaching and learning.* Baltimore, MD: Paul H. Brooks Publishing Company.

Wade, S.E. & Moje, E.B.(2000), The role of text in classroom reading. In M.L.Kamil, P.B. Mosenthal, P.D. Pearson, & R.Barr, R. (Eds.), *Handbook of reading research* Vol. 3 (pp. 609-627). Mahwah, NJ: Lawrence Erlbaum Associates.

Webb, N.L.(2002). *Assessment literacy in a standards-based education setting.* (WCER Working Paper No. 2002-4). Madison, WI: Wisconsin Center for Education Research.

Weinzierl, J., Stephens, D., Staffleman, A., Pearson, P. D., Shelton, J., Rodriguez, A., Gilrane, C. P.(1993).Assessment and decision making in Beta. (Technical Report No. 590). Urbana, IL: Center for the Study of Reading. Retrieved Spring, 2006 from www.eric.org/ericdb/ED362837.htm

Wilkeson, L.C. & Silliman, E.R. (2000). Classroom language and literacy learning. In M.L.Kamil, P.B. Mosenthal, P.D. Pearson, & R.Barr, R. (Eds.), *Handbook of reading research* Vol. 3 (pp. 337-360). Mahwah, NJ: Lawrence Erlbaum Associates.

Winograd, P., Flores-Duenas, L., & Arrington, H. (2003). Best practices in literacy assessment. In L.M. Morrow, L.B. Gambrell & M. Pressley (Eds.). *Best practices in literacy instruction* (2nd ed.). New York: Guilford Press.

Wisconsin Department of Public Instruction (n.d.) WINNS Successful School Guide – Data Analysis. Retrieved Winter 2008 from http://data.dpi.state.wi.us/ data/ selschool.asp

Young, V.M. (2006, June). Teachers' use of data: Loose coupling, agenda setting, and team norms [Electronically published]. *American Journal of Education*, 112, 521-547.

Yin, Robert K. (2003). *Case study research design and methods (*3rd ed.*)*. Thousand Oaks California: Sage Publications.

Appendix A
Interview Questions

(Note: Questions are similar. The three question sets were created in recognition of participants' various roles within the district.)

District Administrators
- What do you see as the purpose of assessment?
- I found this document on your website. (Investigator shows copy of document.) What can you tell me about the district vision or goal that influenced its creation?
- I am particularly interested in reading assessment. What kinds of reading assessment data do you gather in your district? Who decides which reading competencies should be assessed and how they will be assessed?
- To what extent does the data you collect about reading inform the decisions you make regarding: Policies and procedures? Programmatic decisions and curricular decisions? Staff development plans?
- To what extent do your building principals use reading assessment data? Your teachers?
- Can you describe for me the ways in which reading assessment data does or does not influence their actions?
- How do you think your principals and teachers view reading assessment data in comparison to the way you view it as a district level administrator?
- Is there anything else you would like me to know so that I can best understand how reading assessment data is used in this district?

Building Principals and Reading Specialists
- What do you see as the purpose of assessment?
- I found this document on your website. (Investigator shows copy of document.) What can you tell me about the district vision or goal that influenced its creation?
- I am particularly interested in reading assessment. What kinds of reading assessment data do you gather in your district? In your building? Who decides which reading competencies should be assessed and how they will be assessed?
- To what extent does the data you collect about reading inform the decisions you make regarding: Policies and procedures? Programmatic decisions and curricular decisions? Staff development plans?
- To what extent do you as a building principal/reading specialist use reading assessment data? To what extent do your teachers use reading assessment data?
- Can you describe for me the ways in which reading assessment data does or does not influence your actions? Your teacher's actions
- How do you think your teachers view reading assessment data in comparison to the way you view it as a building principal/reading specialist?
- Is there anything else you would like me to know so that I can best understand how you use reading assessment data in this district?

Appendix A
Interview Questions (continued)

Teacher Questions

- What do you see as the purpose of assessment?
- I found this document on your website. (Investigator shows copy of document.) What can you tell me about the district vision or goal that influenced its creation?
- I am particularly interested in reading assessment. In terms of reading assessment in your building, who decides which reading competencies should be assessed and how they will be assessed?
- Do you assess reading competencies that may not be assessed by others? What made you decide to measure these competencies?
- To what extent, do you as a classroom teacher use the data that you collect?
- Can you describe for me the ways in which reading assessment data does or does not influence your actions?
- To what extent does the data you collect about reading inform the decisions you make about curriculum and materials?
- How do you think other teachers view reading assessment data in comparison to the way you view it?
- Is there anything else you would like me to know so that I can best understand how you use reading assessment data in this district?

Appendix B
Agreement of Consent

MARQUETTE UNIVERSITY
AGREEMENT OF CONSENT FOR RESEARCH PARTICIPANTS
Data Driven Decision-Making:
A Case Study of How a School District Uses Data to Inform Reading Instruction
Principal Investigator: Terese A. Brecklin
(Educational Policy and Leadership)

<u>District Administrator, Building Principal, Reading Specialist Consent Form</u>

You have been invited to participate in this research study. Before you agree to participate, it is important that you read and understand the following information. Participation is completely voluntary. Please ask questions about anything you do not understand before deciding whether or not to participate.

PURPOSE: I understand that the purpose of this research study is to understand the ways in which data are used to guide reading instruction in a district, schools within the district and classrooms within those schools. For the purposes of this study, the term, *data*, is defined as information—either quantitative or qualitative information—that is intentionally captured, recorded, and analyzed at either the individual student or group level, and then used to inform instruction. I understand that I will be one of approximately 15-18 participants in this research study.

PROCEDURES: I understand that I will be audio taped during the interview portion of the study to ensure accuracy. The tapes will later be transcribed. Transcriptions and the tapes will be destroyed after 10 years beyond the completion of the study. For confidentiality purposes, my name will not be recorded. I understand that I will be asked to provide documents related to the topic of data-driven instruction and assessment. These might include: copies of assessments that are used in my district or school, copies of policies or procedures that describe how assessments are used, examples of how data are analyzed and reported. These documents will be destroyed after 10 years beyond the completion of the study.

DURATION: I understand that my participation will consist of one 60-90 minutes interview with the possibility of an additional 30-60 minute follow-up interview.

RISKS: I understand that the risks associated with participation in this study are generally no more than I might encounter in everyday life. I understand that every effort will be made to maintain the confidentiality of the participants through the use of pseudonyms. This is important because I may express an opinion that is different from my supervisors, colleagues or other professional peers. A breach in confidentiality could influence a supervisor's evaluation or might prompt peers or co-workers with different viewpoints to ostracize me. I might also be ostracized simply by agreeing to participate in the study. Although I will not be identified by name, and even though the investigator will make every effort to omit specific descriptive information that might give clues to my identity, the possibility exists that quotes used in a summary report could be linked to me if it is read by others who know me well. Additionally, if my role in the district is unique, others employed in the district who participate in the study and read the summary report will be able to link the quote to me.

Initials: _____

Page 1 of 2
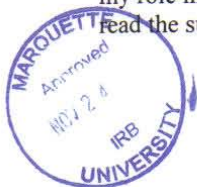
Date: _____

Appendix B
Agreement of Consent (Continued)

Protocol Number: _____

BENEFITS: I understand that the benefits associated with participation in this study include the opportunity to discuss and reflect upon my own professional practice which may lead me to refine and improve my practice. Because this study has the potential to add to the body of knowledge about the ways in which data are used to guide reading instruction, I understand that I will potentially be contributing to that body of knowledge through my participation in this study.

CONFIDENTIALITY: I understand that all information I reveal in this study will be kept confidential to the extent possible, given the design of the study. All my data will be assigned an arbitrary code number rather than using my name or other information that could identify me as an individual. When the results of the study are published, I will not be identified by name. However, the possibility exists that quotes used in a summary report could be attributable to me if the comments or language in the quotes are known to be unique to me or my role in the district. I understand that the data will be destroyed by shredding paper documents, destroying audio tapes, and deleting electronic files 10 years after the completion of the study. (Research records may be inspected by the Marquette University Institutional Review Board or its designees, and (as allowable by law) state and federal agencies.) I will do my part to ensure confidentiality by not discussing my participation in the study with others in my building or district.

VOLUNTARY NATURE OF PARTICIPATION: I understand that participating in this study is completely voluntary and that I may withdraw from the study and stop participating at any time by submitting a written statement to the researcher that I wish to withdraw from the study. If I withdraw from the study, any data I provided will be destroyed.

CONTACT INFORMATION: If I have any questions about this research project, I can contact Terese Brecklin at   or via email at terese.brecklin@marquette.edu. If I have questions or concerns about my rights as a research participant, I can contact Marquette University's Office of Research Compliance at (414) 288-7570.

I HAVE HAD THE OPPORTUNITY TO READ THIS CONSENT FORM, ASK QUESTIONS ABOUT THE RESEARCH PROJECT AND AM PREPARED TO PARTICIPATE IN THIS PROJECT.

_____   _____
Participant's Signature   Date

_____
Participant's Name

_____   _____
Researcher's Signature   Date

Appendix C
Performance Data

District Level Data – All Students
Students Enrolled for a Full Academic Year
Test Date: November 2006

	Minimum	Basic	Proficient	Advanced	Proficient/Advanced Combined
Grade 3	1%	10%	31%	58%	89%
Grade 4	2%	8%	36%	53%	89%
Grade 5	1%	5%	31%	63%	94%

Percent of Students Scoring at Proficient + Advanced Levels in Reading	Percent of Students Scoring at Proficient + Advanced Levels within a Specific Subpopulation (Scores signal the district in danger of missing AYP of 87% by 2011-12)
90%	Economically Disadvantaged (77% proficient/advanced) African American (79% proficient/advanced) Students with Disabilities (58% proficient/advanced)

Appendix D
The Coding Process

Through the process of open coding I created numerous categories based upon the exact word of the participants. This was a multi-step process because varying roles of the participants. My first step was to code the transcript of each participant. I used the comment feature in Microsoft Word to accomplish this. A sample is shown in Figure 1 below.

Figure 1: Initial Coding
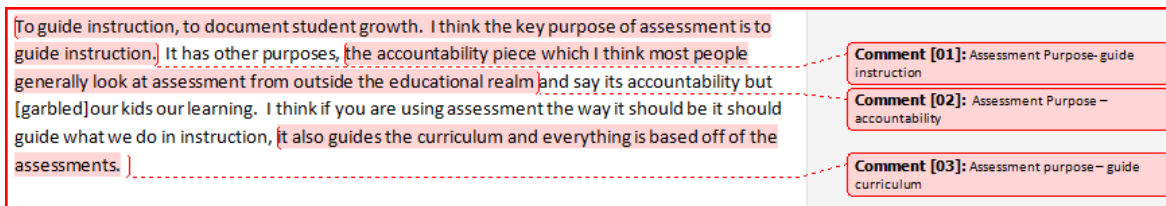
The use of the comment feature shown above proved to be fairly cumbersome. Therefore, I used marginal notes in the second round. This proved to be fairly efficient because I could label the transcript by hand faster than I could with the comment feature. The second go-round was more revealing because I was able to pick apart sections that initially had been coded with one general label. A sample is shown in Figure 2 below.

Figure 2: Coding – Round 2
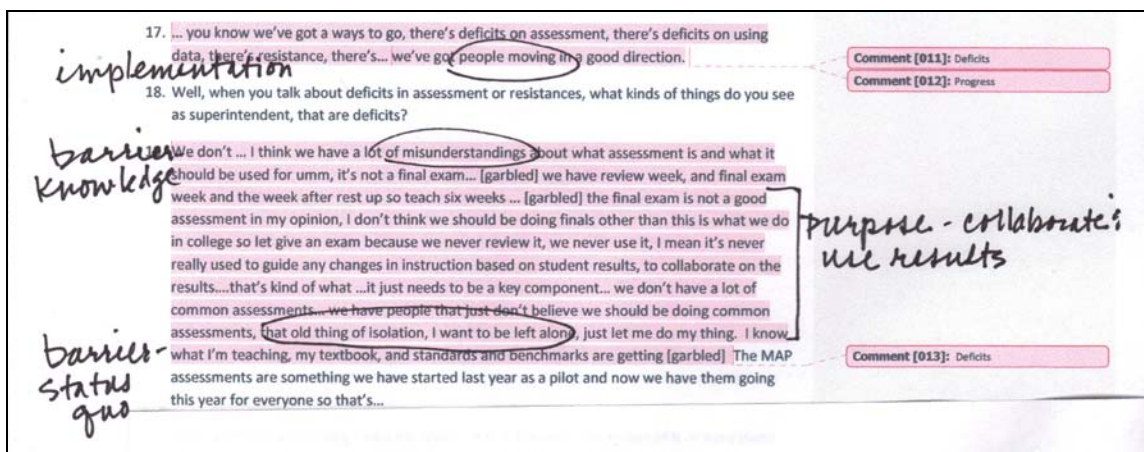
Next, I used the codes I created as a means of comparing the comments of participants who had similar roles. This was accomplished by cutting and pasting like categories into a new document. The process helped me understand the similarities and differences within each group as well as the similarities and differences among the varying roles. Figure 3 demonstrates this process. In this instance principals are commenting about the purpose of assessment.

Appendix D (Continued)
The Coding Process

Figure 3: Comparing Texts

| Purpose of Assessment – Guide Instruction |
|---|
| 1. **Maple Grove Principal**: I view it as teachers using assessment to follow students, help follow their students and bring them to the place they need to be. When you ask me how I view assessment, it's the most important thing we do along with feedback umm, it's the most important thing that we do. |
| 2. **Woodland Principal:** To drive instruction**.** Absolutely and that's the only thing I have ever known and I am thinking of my experiences as a teacher… that's what I first learned… I have never been on that page with teachers who do assessment for the sake of we're giving a test. How are you using this to make this kid smarter or better or learn it to use it in another context? That's how I see it. I can't see it another way. |
| 3. **Oakdale Principal:** To see if the students have learned what had been taught by the teachers, um, obviously, whatever at the time is being taught to make sure that the students have gained that knowledge. And then that will then guide the instruction of the teachers, of can we move on from where we are now or do we need to continue where we're at from that that point I think are the keys things because you know the assessment will tell you I just taught X, have they learned it, um or do I need to keep doing it and tell you for every child in the class where they are at and where they are on whatever you are teaching at the time I think are the key components because that's what… I think ties to all instruction is to know where the students are and you know the assessment part of it –and there are all different types of assessment—the daily assessment, the final assessment and all that –but that all has to be done, there's the on-going assessment so you know when you say assessment I don't take it as… I guess way back in the past assessment was final tests but that's not what it is any more, it's daily assessment you give kids by asking questions, by little tiny couple question assessments to see if they are catching on as you go through units of study or whatever you are doing at the time. So all that sort of guides instruction I think.   [Later in interview] as I said in the beginning I think, is just so much more important so the teacher knows where the students are, if they are learning what they are teaching… um teachers talking to each other amongst in their grade levels and outside their grade levels so they can learn from each other on what's working and what's not and where they can go. |

Appendix D (Continued)
The Coding Process

During the entire process, I kept running grocery-style lists of the codes and the categories that were emerging. Each word on the list became a word in a textbox in Inspiration®. Inspiration® is a software program that enables one to create graphic organizers. Once all the words were in inspiration, I combined arranged and rearranged the textboxes until I felt I had captured the key ideas that had emerged from my analysis. I created a concept map for each participant group and one that captured the district level of analysis. See Figure 4 for an example. The concepts maps follow in Appendices E-H.

Figure 4: Capturing Ideas in Inspiration

231
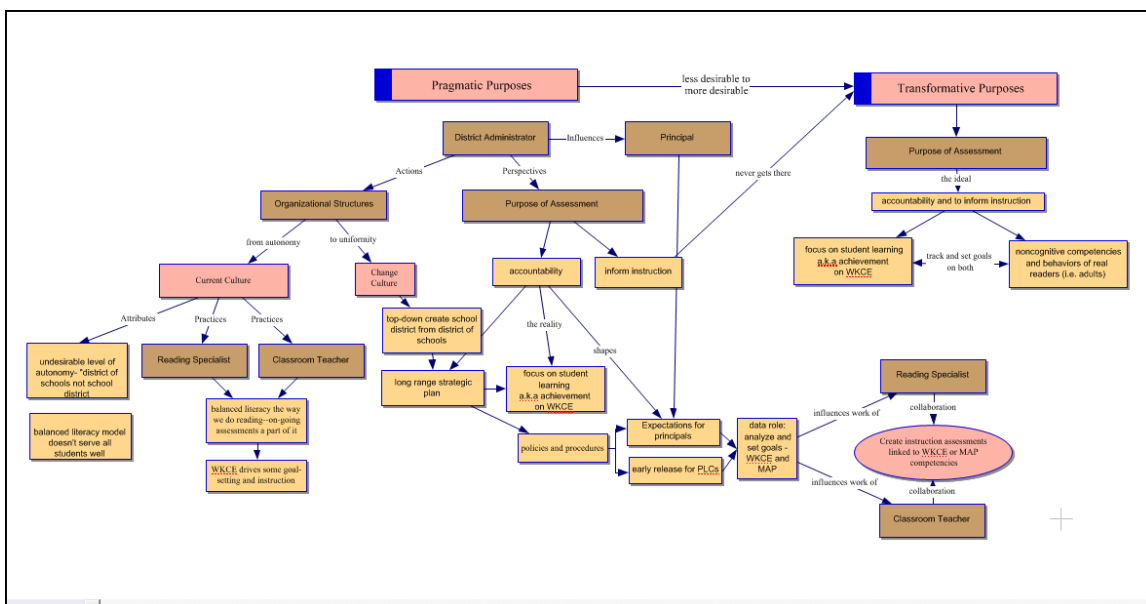
Appendix E
Data Analysis Concept Map - District Administrators
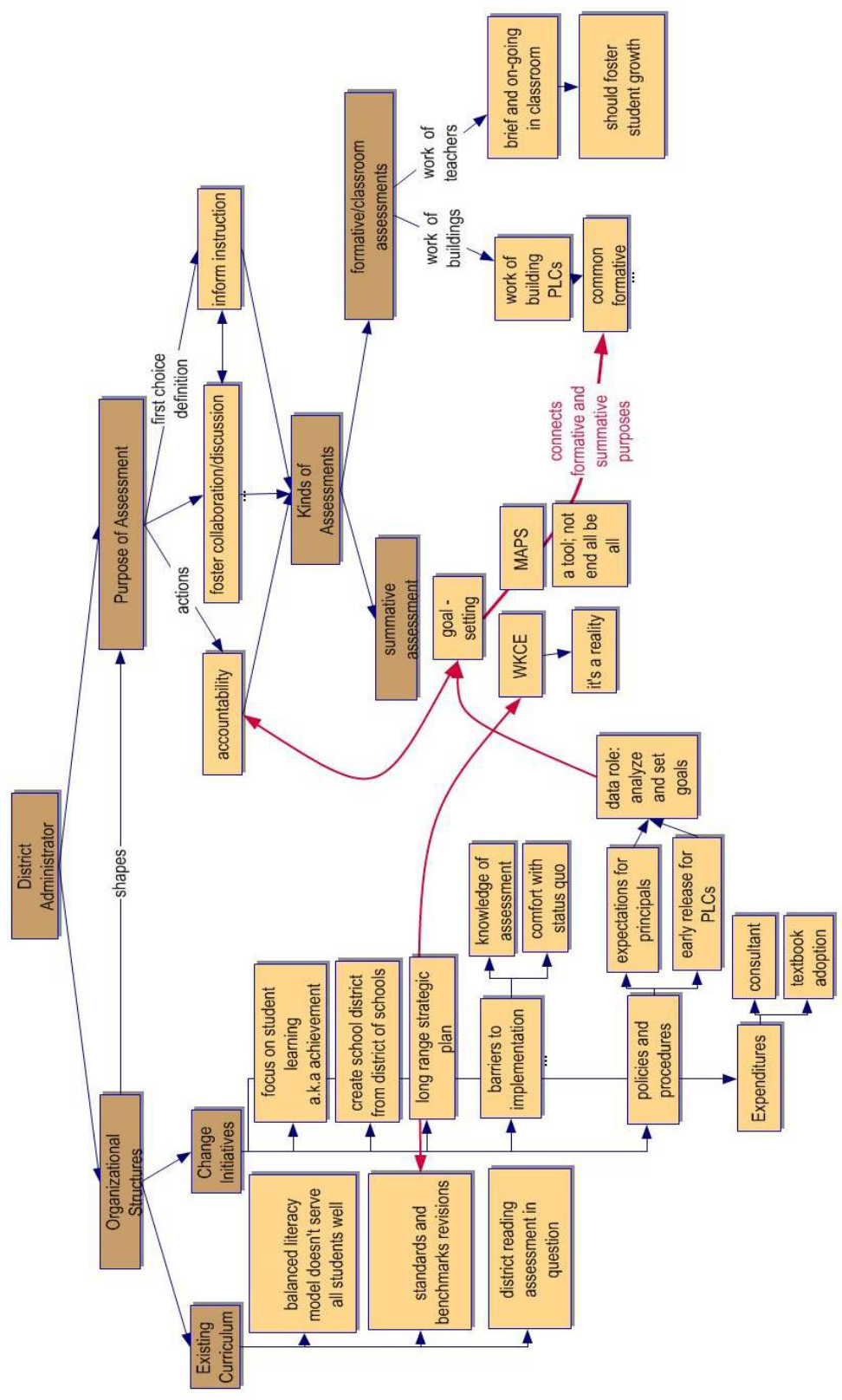
Appendix F
Data Analysis Concept Map – Principals

Appendix G
Data Analysis Concept Map – Reading Specialists
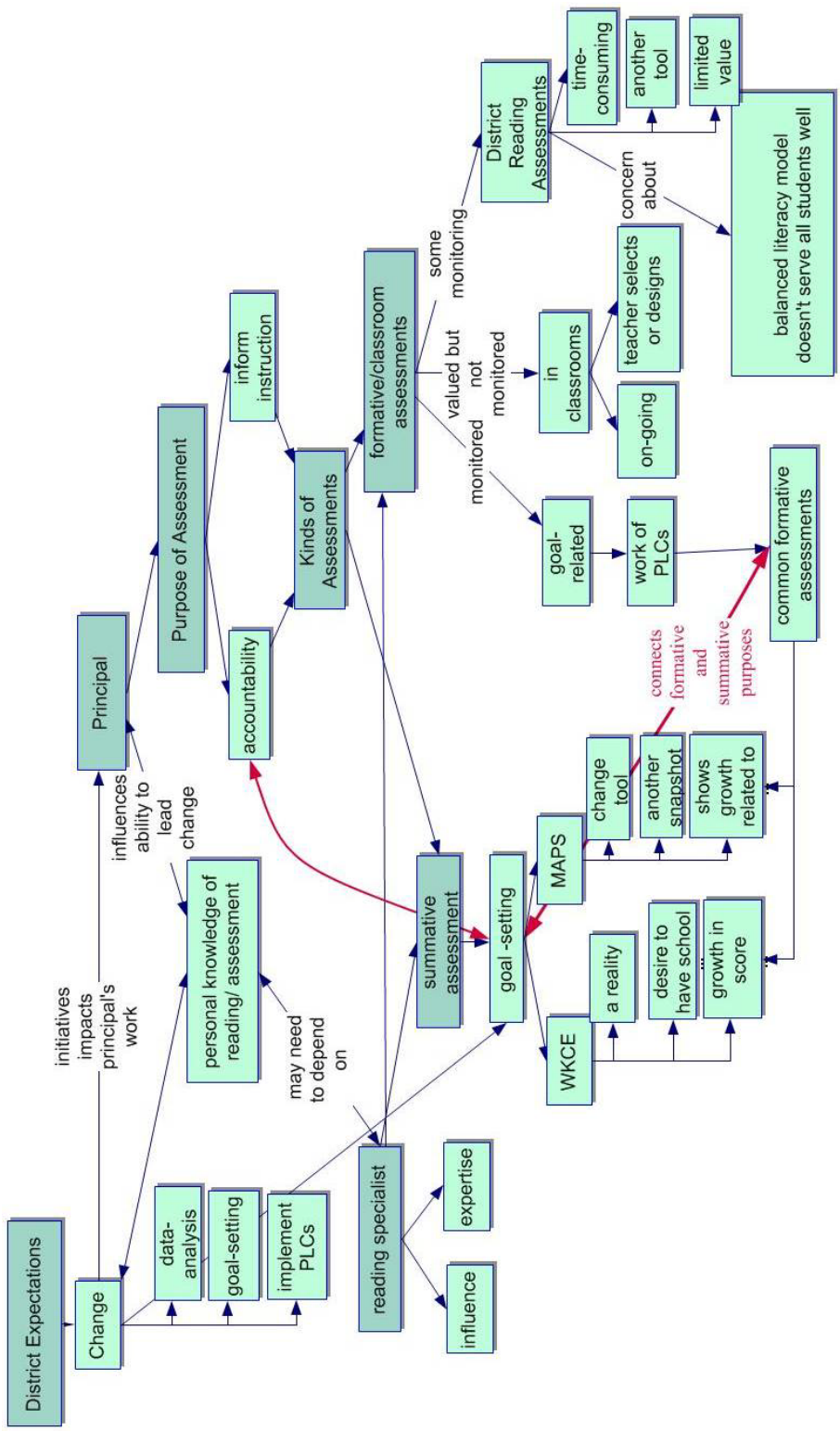
Appendix H
Data Analysis Concept Map – Classroom Teachers
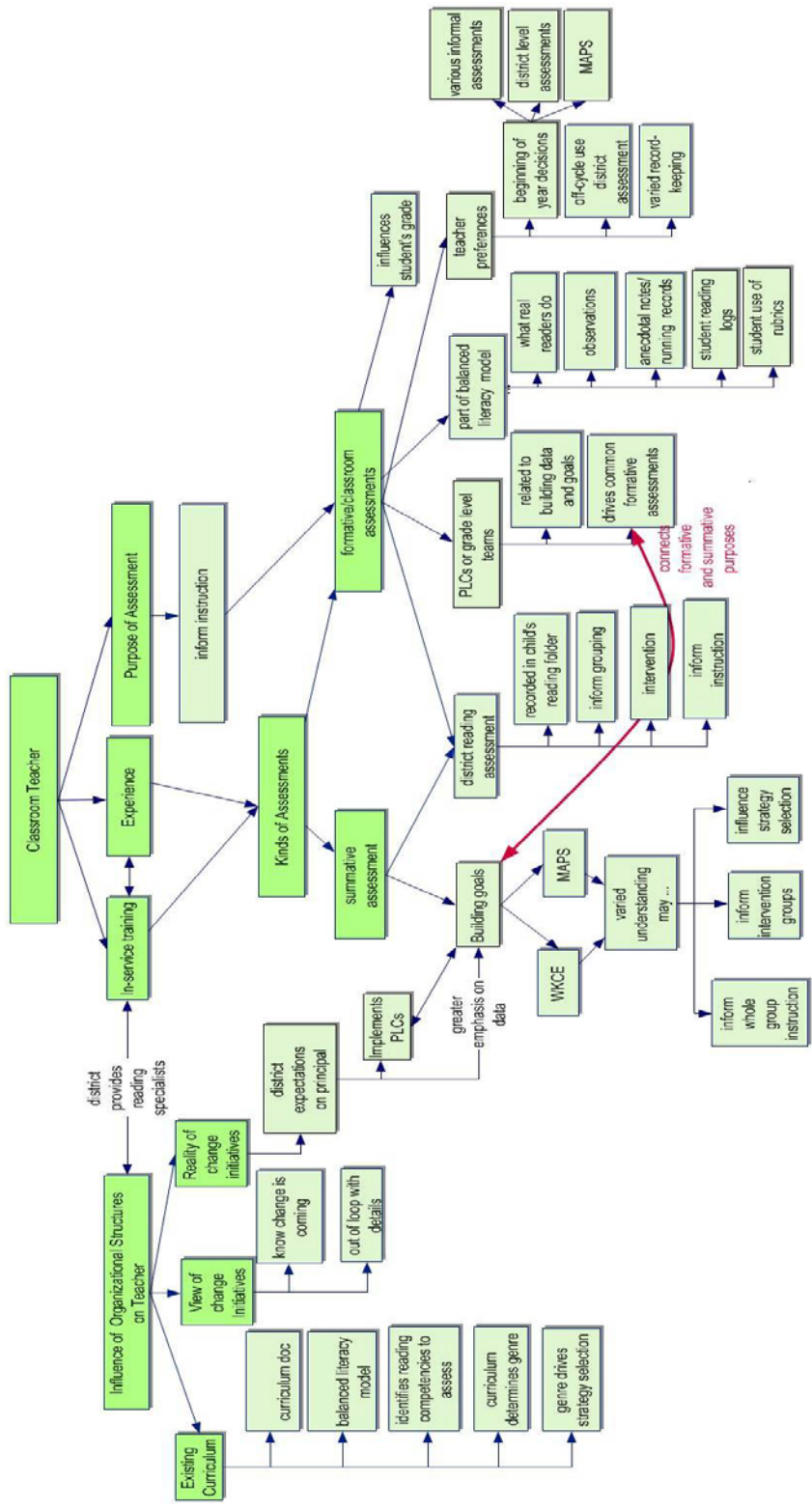
Appendix I

Summary Table – Selected Categories
Sample Responses of Participant Groups

Response Categories	Participant's Responses			
	District Administrators	**Building Principals**	**Reading Specialists**	**Classroom Teachers**
Purpose of assessment	"If you are using assessment the way it should be used, it should be to guide what we do in instruction," "Accountability"	"In the past, assessment was final tests but it's not that way anymore. It's daily assessments to see if the students are catching on." "Accountability"	"I need to be able to document progress over time and determine the effectiveness of our program and our instruction and I also need to be able figure out where to go next with these students, what have they learned, what do they still need to learn."	"The purpose of assessment is to understand how my students are functioning as readers… and then I would use that information to plan for instruction. So, I guess the purpose is to ultimately plan for instruction."
District Reading Assessment	"It is going through a review process and we will make some determinations about what is the best use of it."	"You know when we get together and our Superintendent asks us for our data, it's never [the district level reading assessments] we haven't had those conversations at our principal meetings." "I like the extra stuff that we get off [the district level reading assessment] that we don't get out of MAP. "It's a tricky instrument".	"We get the information we need about decoding and fluency, as well as specific comprehension information and we use that pretty much to identify who needs additional intervention." "[Some teachers] are not quite sure how to use the information from the district level assessment because you are giving all these passages with all these questions."	" She had the fluency part down and the decoding, and the word ID for the known words, but then the comprehension, no." "I knew she was making good progress, and then I did the [district level reading assessment], and it was like, wow," she just kept going."

Appendix I (Continued)

Response Categories	Participant's Perspectives			
	District Administrators	**Building Principals**	**Reading Specialists**	**Classroom Teachers**
WKCE	"You can't just ignore it and say it's not important because it is. As a district we have to score well." "Teaching to the test, the way it is set up structurally, is the way we are trying to move our organization."	"I will be honest with you, I am a little bit sad about our test scores." "We have an educational debt to these students. We have had them since 4K. We don't have an excuse. "It's not very high to be proficient [on the WKCE]. I rather have them all advanced…"	"We start to teach them how to construct a response to a question. We know we can't leave it all to third grade…" "We know effective instruction and good curriculum is the best test prep." "If we see they are struggling on something, we kind of increase instruction in that area or make it a little more explicit, or support it more. But we don't let the assessment drive what we are going to teach. We already know what that is."	"We really hope that just our day to day teaching is quality, thoughtful, that we don't really need to put in anything special in place. Now a couple of days before the WKCEs ….we print off those [practice tests] and have the kids do the practice readings. We talk about good test-taking skills; about getting sleep; about not stressing out; about doing your best. We just let the kids take them and they have been pretty successful."

Appendix I (Continued)

Response Categories	Participant's Perspectives			
	District Administrators	**Building Principals**	**Reading Specialists**	**Classroom Teachers**
Measures of Academic Progress	"Philosophically, it is hard for elementary people to adjust because they are not used to getting that kind of data; they are not used to using standardized data like MAP."	"I think it has single-handedly changed the way teachers set goals and go about it. They can't wait for data now, thanks to MAP because it is real; it is in real time. "In Fall we see it as our formative assessment and we work, work, work, work. And after we test in December or January, then we say, ok, this is where we saw our deficits, we really gotta kick it up and see what happens in May" "I think it will be more useful as we get used to it…"	"It gives you a list of skills for each range—the skills the child is secure in, the skills to be reinforced, and the skills to be taught and introduced. So you can get a good gauge of where you want to go next with the child." "I think it does not make up for sitting next to a child and listening to them and watching their behaviors…" "It's possible we could discontinue our 4th and 5th grade district assessments if we can figure out how to use MAP well enough. "	It's a good overall snapshot…I want to know if the test can tell me if they struggle with characters or setting. It's a goal for me to learn how to understand the reports. "We are just starting to figure out how to more effectively use this in the classroom. It does break it down very nicely into all these strands … and then we use part of the [PLC] early release time to look at all this stuff and see how we can apply it to the classroom."

Appendix I (Continued)

Response Categories	Participant's Perspectives			
	District Administrators	**Building Principals**	**Reading Specialists**	**Classroom Teachers**
On-going Classroom Assessments	"We have got to make sure that if we are going to remediate, we are going to remediate. You don't take a child and put them in a separate group for a year and just leave them there. And the only way you are going to know if they are making those gains is through assessing."	"What you see in the classroom everyday with your assessments is to me the most important and authentic, second is MAP testing, third is the WKCE." "How are you using this [results]to make a kid smarter or better?"	"Teachers can't live without it. It is part of their daily routine, in one way or another, to gather information, to decide how this child is doing." "For those of us who have been here for awhile, , it's a natural part of what we do because we have to in order to be able to teach the way we do around here."	"Sometimes it's kind of hectic to record everything. You get interrupted or get a phone call, or the class period is over." "We study you as readers. That is our job to study you. So, with all of these things, you might think, oh, I gotta do this, but really it helps us know you and it helps you know yourself as a reader."
Common Assessments	"It's what they are learning not what I am teaching." "We said we did [assess to common standards] but we didn't have the assessments in place to do that"	"Students were to complete a Venn diagram comparing themselves to a character in the story. After completing this they will complete a four -question assessment that will be collaboratively scored."	"Within the PLCs we are starting to do that [focused conversations around common assessments] a little bit better." "So in terms of developing some of those common assessments, they are talking more about what the whole class is doing. But I think that is a relatively new concept."	"Began organization of common formative assessments to be used during reading and reading intervention/ extension…discussed how well/not well students did on baseline. Know that there is teaching to be done…"
Organizational Structure	It is time to go from a "district of schools to a school district." "It's like we need to say, this is what we are tight on. You can be loose about how you get there."	"It's pretty autonomous here; each building is pretty much its own castle."	"I'd say certain administrators… perhaps they already have some preconceived notion of the way they think things should go, and that's the way they are going."	

Appendix I (Continued)

Response Categories	Participant's Perspectives			
	District Administrators	Building Principals	Reading Specialists	Classroom Teachers
Collaboration Around Data	We never review it [final exam data], we never use it. It's never really used to guide any changes in instruction based on student results… to collaborate on the results…needs to be a key component…	"If they [teachers] come up with their own plan because they saw the data and then came up with the strategies, then it becomes their idea about what they are going to teach in their class." "I have to know what [the data] is first, otherwise, I don't know if they are on the right track or not."	"We figure it out together. It's not me saying this is what you need to do." "We found some kids over the course of the year who were very good at staying under the radar… we decided we needed what we needed to assess and what it would look like."	"This is an area [on the MAP] that we really zoomed in on as [grade level] teachers because we noticed that there were a number of kids who were low in this area and that it has to do with inferring so really all of our reading zooms in on that."
Professional Learning Communities	"What do we want our students to know? How will we know each student has acquired targeted standards and benchmarks? How will we respond when students do not learn? How will we extend and enrich learning for students who have mastered the benchmarks?"	"Teachers talking to each other… so they can learn from each other on what's working and what's not and where they can go—that's important."	"In terms of data-driven instruction, I feel that's what we have always been working on and will continue to work on. It isn't something new." "It's experience and staff development because you could have a very, very experienced teacher who has never been taught how to use data to inform instruction and then it isn't very helpful."	"…We have been doing a lot of stuff with the WKCE and goals–setting for the PLC groups so that's where we have been using a lot of our data. We hope that we can get more time to develop common assessments and stuff throughout the building and grade levels." This coming [PLC] we are going to be grading our [district level] writing assessments so this [PLC] will be more focused on it.

Appendix I (Continued)

Response Categories	Participant's Perspectives			
	District Administrators	**Building Principals**	**Reading Specialists**	**Classroom Teachers**
Professional Development	"I think we have a lot of misunderstanding about what assessment is and what it should be used for."	"So we would in-service them on various types of formative assessments and then we would give them homework…when we meet at our next in-service date we are going to ask you to get up and share how it is going."	"I find staff development a key…programs do not teach kids, books do not teach our kids, teachers teach our kids and they have to be expert teachers of reading to do it."	"When I first got here, every Monday afternoon I left and met with the district reading specialist and she coached us all year long."
Long range strategic plan	The district will "perform a comprehensive, in-depth, data-driven analysis of student performance and current academic curricula, and will formulate an improved scope and sequence for each academic core area that will maximize achievement for all students."	"We really have to start looking at how our kids are doing, not necessarily how we are doing as teachers, but how well our kids are learning. That's the shift we have to get to."	"When it says schools and teachers use readily available achievement data, we do that… but I thinks it's a bigger picture than just classroom assessment."	"I've heard of it." "I haven't read it specifically but I am aware of the purpose of data and we are doing a lot with this one [points to part of document] in that we are doing the MAPS which has helped with finding our where the students are so we have been using the MAPS testing quite a bit so it's not a surprise to me."

241

Appendix J

Reading Discussion Group Observation Log

Appendix K

Post-it Rubric

Assignment: _____   Due Date:_____

Directions: come to each discussion with 4-6 thoughts written on Post-it notes. After the discussion, place your Post-its (no more than 6) here:

Post-It Note/Response Evaluation

Student-Highlighter Teacher-Marker		
Extraordinary	**Good**	**Dreadful**
Post-it notes and responses show deep thinking (inferences) about the characters and important events in the story.	Post-it notes and responses show some thinking about the characters and important events in the story.	Post-it notes and responses show little or no thinking about the characters and little or no understanding of the story.

Appendix L

Non-fiction Reading Strategies Assessment

PREDICTING: Use headings, titles, and graphics to predict what the text will be about; also use prior knowledge.

CLARIFYING: If you don't understand something (IDGI), you need to ask questions; things that may need clarifying include word identification, word meaning, or ideas.

QUESTIONING: Identify information that would make good teacher questions.

SUMMARIZING: If you were to explain what you read about today to a friend, what would be important to include?

Appendix M

Book Discussion Group Rubric
(Criteria generated by students.)

Book Talks	• Expand Your Thinking	
	• Deepen Your Understanding	
Best Ever	So-So	Oops!
Everyone actively participated.	Most people were active.	One or two people dominated the group while the others sat quietly.
Lots of piggy-backing on each other's ideas.	Some piggy-backing.	No piggy-backing.
Questions helped expand our ideas and understanding.	Few questions so ideas were kind of skinny.	Not really any good questions.
Thoughtful evidence from the reading was given to support ideas.	Some evidence was given.	Little to no evidence was given.
Talk stayed close to the book.	Talk occasionally drifted away from the book.	We lost the book!
Real listening.	Some real listening.	Fake listening.
All disagreements were handled politely with care.	Some polite and some critical disagreements.	Harsh disagreements.
One voice at a time.	Some interrupting.	Confusion of voices.

Next time I will try to help the book talk by...

Appendix N

Write More Prompt and Rubric

Assignment:_____   Date due:_____

Write More

Choose a post-it note that shows deep thinking about the characters or important
events in this part of the story and copy it here:

```
┌─────────────────────────┐
│                         │
│                         │
│                         │
│                         │
│                         │
│                         │
│                         │
└─────────────────────────┘
```

Now, write more about this idea to explore and expand your thinking.   Be sure to
include details and examples from the passage to support your thinking.

Write More Evaluation			
Student - highlighted Teacher - marker			
Extraordinary!	**Very Good**	**Satisfactory**	**Unsatisfactory**
• Writing shows extended thinking about the selected post-it note.	• Writing shows extended thinking about the selected post-it note.	• Writing shows some extended thinking about the selected post-it note.	• Writing shows little additional thinking about the selected post-it.
• Ideas are fully supported with evidence from the story.	• Ideas are fully supported with evidence from the story.	• Ideas are supported with some evidence from the story.	• Ideas are not supported with evidence from the story.
• Ideas focus on an important part of the story.	• Ideas focus on an important part of the story.		
• New ideas are presented and are also supported with evidence from the story.			

Appendix O

Grade 4 and 5 Reading Achievement

Grades 4 and 5 Reading Achievement Rubric

	A	B	C	N
Comprehension	• Student's comprehension is detailed and complete. • Student applies knowledge of text organization to comprehend. • Student is consistently and independently able to apply strategies to comprehend a wide variety of material. • Student reads at grade level or above.	• Student's comprehension is complete but does not show great detail. • Student generally applies knowledge of text organization to comprehend. • Student generally applies strategies well to comprehend a wide variety of material. • Student reads at grade level.	• Student maintains basic understanding of text but has some misconceptions. • Student inconsistently applies knowledge of text organization to comprehend. • Student consistently needs assistance to apply strategies to comprehend; may apply strategies at a lower level. • Student needs some support to apply strategies to comprehend. • Student reads one level below grade level.	• Student has many misconceptions about what they read; does not understand. • Student does not apply knowledge of text organization to comprehend. • Student consistently needs assistance to apply strategies to comprehend; may apply strategies at a lower level. • Student reads one or more levels below grade level. • No evidence in understanding has been shown. Not enough to accurately assess student's comprehension.
Decoding Skills	• Student is extremely independent, flexible, and fluent in his/her strategy use to decode unfamiliar words. • Student has excellent fluency when reading; adjusts reading rate appropriately.	• Student is fairly independent, flexible, and fluent in his/her strategy use to decode unfamiliar words. • Student maintains fluency when reading; may not always adjust reading rate.	• Student needs some prompting to use different strategies to decode unfamiliar words. • Student has fairly good fluency when reading; reads everything as the same rate.	• Student needs much assistance and prompting to use different strategies to decode unfamiliar words. • Student has little fluency when reading; word by word reading with some stretches of fluency. • May be able to apply strategies at a lower level.
Independent Reading	• Student always uses independent reading time effectively. • Student reads a wide variety of genres. • Student shows great confidence as a reader. • Student always makes appropriate book selection choices by self.	• Student uses reading time effectively most of the time. • Student reads different genres at times. • Student shows confidence as a reader. • Student usually makes appropriate book selections by self.	• Student sometimes uses independent reading time effectively. • Student reads few different genres. • Student shows self-confidence as a reader a times. • Student may make appropriate choices; may need assistance.	• Student shows avoidance behaviors during independent reading time. • Student tends to read in only one genre. • Student has little confidence in self as a reader. • Student does not make appropriate book selection choices; needs much assistance.

Appendix P

Grade 1 Teacher's Log

Guided Reading Group Anecdotal Record Sheet

Students: ▓▓▓▓▓▓▓

Date: 1/12/09

Title: Brave Father Mouse

Level: 6

Instructional Focus:

1st + end + SL

Comments:

up- |©| (F) ✓   Wh- | (TP) | With | (TP) - (T)
upstairs   Where |

p.8 walks   Here | SC   Br-|
    wakes   He | Bread

Date:

Title:

Level:

Instructional Focus:

Comments:

Guided Reading: Making It Work   Scholastic Professional Books

255

Appendix Q

Grade 1 Running Record

Running Record

Student _____ Date _____ Teacher _____ Lesson _____

Title _The Clock That Couldn't Tell Time_____ Level __14__ RW _125_ WPM _____

# of Errors _4_ # of SC _0_ SC:E ___:___ Accuracy [(RW-E)/RW] _97_ % 95-100% (Ind.) _X_ 90-94% (Instr.) ____ <90% (Hard) ____

Powerful Praise:	did it on your own!
Teaching Point:	look through the whole word

Notes / Observations:

	-T-	E	SC	E MSV	SC MSV
✓ ✓ ✓ ✓ ✓					
✓ ✓ ✓ ✓ ✓					
✓ ✓ ✓ ✓ e|TP|en-|TP|T ✓ ✓	I				
enough					
✓ ✓ ✓ ✓					
✓ ✓ ✓ ✓ ____					
✓ ✓ ✓ ✓ ✓					
✓ ✓ ✓ ✓					
✓ ✓ ✓ ✓ ✓					
✓ ✓ ✓					
✓ ✓ ✓ ✓ ✓ ✓ (PP) rereading					
✓ ✓ ✓ ✓ e|TP|ch-|-im|TP|chim|TP		I			
chime					
✓ ✓ ✓					
✓ ✓ e|ch|T	I				
echoed					
✓ ✓ ✓					
✓ ✓ ✓ ✓					
✓ ✓ ✓					
✓ ✓ ✓ ✓ ✓					
✓ ✓ ✓ ✓ ✓					
✓ ✓ perfect ✓ ✓ (TP)					
sh-|sh-|sh-ut- ✓ (PP)		I			
shutters					
✓ ✓ ✓ ✓					
✓ ✓ ✓ ✓					
✓ ✓ ✓ ✓					
✓ ✓ ✓					
✓ ✓					
✓ ✓ ✓ ✓					
✓ ✓ ✓ ✓ ✓ ✓					
✓ ✓ ✓					

2-2