2009

A Descriptive Study of Teachers' Instructional Use of Student Assessmetn Data

Nancy Hoover
Virginia Commonwealth University

Virginia Commonwealth University
School of Education
Ph.D. in Education

Dissertation Defense

Title of Dissertation (Copy of dissertation to be filed with Office of Graduate Studies in Education):

A Descriptive Study of Teachers Instructional Use of Assessment Data

This is to certify that the undersigned have read and approved the Dissertation of:

Nancy R. Hoover (Student) presented on 11/23/09 (Date)

Comments:

Lisa M. Abrams
Dissertation Committee Chair

Committee Member

Vicki B. Wilson
Committee Member

Committee Member

Committee Member

Copies to: Office of Doctoral Studies, Dissertation Chair, Student

A Descriptive Study of Teachers' Instructional Use of Student Assessment Data

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at Virginia Commonwealth University.

by

Nancy R. Hoover
Bachelor's of Science, Virginia Commonwealth University, August 1996
Master of Teaching, Virginia Commonwealth University, August 1996

Director: Dr. Lisa Abrams
Foundations of Education
School of Education

Virginia Commonwealth University
Richmond, Virginia
November 2009

Acknowledgement

Significant journeys are rarely completed in isolation; for me I can say I was in very good company for the entire journey leading to the completion of this dissertation. Many people joined me at one time or another throughout the Ph.D. process and I would be remiss if I did not acknowledge some of the most significant.

I would like to thank the members of my committee for their service and support. Dr. Vicky Wilson offered guidance and encouragement, even before my dissertation process began. I am grateful for all her support and most especially a conversation we had along the way where she noted that as a profession educators have "perfected the art of teaching but still have a ways to go to perfect the science of teaching." That profound statement planted the seed that was to become this study, and it is particularly gratifying to have had her serve on my committee. Dr. Chris Corallo took on the challenge of serving on my committee and his encouragement during my Prospectus Hearing gave me the confidence to believe my study could possibly make a positive contribution to future teacher staff development. Dr. Jonathan Becker was instrumental in helping me formulate the first three chapters of my dissertation under his guidance during EDUS890. As a member of my committee, his comments and suggestions for organization and content of my final dissertation provided valuable input that I am sure made my writing better.

I must reserve a special thanks to my Dissertation Director, Dr. Lisa Abrams. Her patience seems to know no bounds, though at times I'm sure it must have waned for me. Dr. Abrams proved to be a wonderful instructor during my course work leading up to this final step, and her help and support as the chair of my committee surpassed all expectations. She was instrumental in helping me fine-tune my research topic, navigate the IRB process, and provided the exact amount of encouragement and support to give me the confidence to keep moving

forward. I'll be forever grateful for her assistance, encouragement, and willingness to take on the task of my Dissertation Director.

I had the good fortune of having several excellent instructors join me at one point or another along my journey. Dr. Robert Triscarri, Dr. James McMillan, and Dr. Lisa Abrams helped put statistics and data analysis in context and made sense out of an otherwise confusing subject. Dr. Samuel Craver, Dr. Maike Philipsen, Dr. Nita Bryant, and Dr. Cheryl Magill, to name a few, helped make history and the foundational study of education meaningful and relevant for this otherwise math/science mind.

I am eternally grateful to the members of my cohort. No one but you knows the time and work that went into getting this far. The journey would have been very difficult without you and certainly much less enjoyable. I am especially grateful for the good friends I have found in my colleagues Randi Smith and Patti Fox. Through group projects and endless discussions to help flesh out paper topics, concepts, and SPSS questions, both have become an integral part of who I am as an educator. Special thanks go to Joe Tylus and Beth Teigen. Both these friends served as cheerleaders, counselors and task-masters all at the same time. I am honored to have traveled this journey with these two friends at my side. I can safely say that without them, I am not sure I would have gotten this far.

I have countless friends and colleagues, too many to mention, that have offered me their encouragement over the last four years. My family, comprised of brothers and family, has been a stalwart of support. I owe a special thanks to my mom, Kitty Belcher, whose faith in me has never been anything but full throttle. My mom never ceases to surprise me with all the things she has accomplished and still accomplishes. She serves as my inspiration of what a woman can

do if she puts her mind to it. My pops, Tiny Belcher, gives the unconditional love that keeps one going even when it seems it cannot be done.

My sons, Matthew and Benjamin, have always been amused at their mother's continuing, and seemingly never-ending, educational pursuits. As one once asked me, "You know there is no degree *AFTER* a Ph.D, don't you, Mom?" My answer, "Well, there is always a second Ph.D." sums up their amusement. They have given me loving support that provided perspective when I was overwhelmed, and put my feet back on the ground when I had my head in the clouds. And finally thanks to my wonderful husband, Roger, who has always given me his unconditional and undivided love, support, and attention. His love and encouragement has been the one thing that has accompanied me on every step of this journey and I know for certain that without him, I could not have made it to this point. Thank you.

Table of Contents

Acknowledgement iv

Table of Tables x

Table of Figures xii

Abstract xiii

Chapter 1 Introduction 1

Statement of Problem *1*

Rationale for Study of Problem *2*

Statement of Purpose *3*

Literature/Research Background *3*

Research Questions *5*

Design and Methodology *6*

Summary *7*

Definition of Terms *9*

Chapter 2 Literature Review 12

Introduction *12*

Data-Driven Decision-Making *13*

Formative Assessment – Using Data to Make Decisions *19*

Classroom Assessment Literacy *24*

Summary *29*

Chapter 3 Methodology 31

Introduction *31*

Research Questions *32*

Research Design *32*

District Profile and Teacher Population *33*

Sampling *34*

Participant Characteristics *36*

Instrumentation *40*

Administration Procedure *43*

Variables *44*

Data Analysis *46*

Delimitations *48*

Chapter 4 Results 49

Overview *49*

To What Extent do Teachers Use Summative Assessments in a Formative Way? *49*

What is the level of teacher assessment literacy in a large suburban school district? 60

What is the relationship between teachers' reported use of summative assessment in a formative way and their assessment literacy level? *70*

What changes in instructional practice result from teachers' use of assessment data?77

Summary *79*

Chapter 5 Conclusion and Recommendations 82

Overview *82*

Significant Findings *83*

Discussion *92*

Recommendations *98*

Limitations *102*

Conclusions *103*

List of References 107

Appendices 112

Appendix A - Email Permission to use Classroom Assessment Inventory 113

Appendix B

Survey to Measure the Extent to which Teachers Use Student Assessment Data to Inform Their Instructional Practice 114

Appendix C - Email Notification for Participants 125

Appendix D - Table D: Research questions and related data analysis method 127

Appendix E - Histograms of Assessment Literacy Scores 128

Appendix F - *Post Hoc* Analyses 130

Vita 134

Table of Tables

Table 1 *Frequency Distribution of Participant Characteristics* 37

Table 2 *Frequency Distribution of Sample by Year's Teaching Experience* 38

Table 3 *Frequency Distribution of Sample by Year's in Current Teaching Assignment* . 39

Table 4 *Frequency Distribution of Sample by Content Area* 39

Table 5 *Frequency Distribution of Specific Types of Assessments Administered by Teachers* 51

Table 6 *Frequency Distribution of Methods of Data Analysis* 55

Table 7 *Frequency Distribution of Data Analysis for Specific Assessments* 56

Table 8 *Comparison of Percentage Administration of Specific Types of Assessment and Percentage of Analysis of Assessment Data* 58

Table 9 *Frequency Distribution of Data Analysis Method: Disaggregated by AYP Subgroups* 60

Table 10 *Mean Composite Scores for Assessment Literacy Standards* 62

Table 11 *Comparison of Classroom Assessment Literacy Mean Scores for School Level* 63

Table 12 *MANOVA for School Level* 64

Table 13 *Comparison of Classroom Assessment Literacy Means for Years Teaching Experience* 65

Table 14 *MANOVA for Years Teaching Experience* 66

Table 15 *Comparison of Classroom Assessment Literacy Means by Degrees Attained* .. 67

Table 16 *MANOVA for Degrees Attained* 68

Table 17 *Classroom Assessment Literacy Mean for English, Mathematics, and Science Teachers* 69

Table 18 *Classroom Assessment Literacy Mean for English, Mathematics, and Science Teachers* 69

Table 19 *MANOVA for Primary Teaching Responsibility* 70

Table 20 *Classroom Assessment Literacy Mean by Type of Assessment Analyzed* 73

Table 21 *MANOVA for Released SOL Assessment Items Analyzed* 74

Table 22 *Classroom Assessment Literacy Mean by Method of Data Analysis* 76

Table 23 *MANOVA for Methods of Data Analysis: Examining Mode* 77

Table 24 *Frequency Distribution for Changes Made in Instructional Practice Based on Student Assessment Data* 78

Table 25 *Comparison of Classroom Assessment Literacy Inventory Scores* 89

Table of Figures

Figure 1. Comparison of most frequently administered assessments by school level….52

Abstract

A DESCRIPTIVE STUDY OF TEACHERS' INSTRUCTIONAL USE OF ASSESSMENT DATA

Nancy R. Hoover

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Virginia Commonwealth University 2009

Director: Dr. Lisa Abrams
Foundations of Education
School of Education

The overarching question for this study is: to what extent are teachers using summative assessment data in a formative way? A survey research design study was implemented to address this question. A web-based survey was administered to elementary, middle, and high school teachers in a large, suburban school division in central Virginia. The survey data were used to determine the frequency with which teachers administered specific types of summative assessments, analyzed student summative assessment data, made changes in their instructional practice as a result of their analysis, and the level of teachers' assessment literacy.

The results of this study suggest teachers are administering a variety of summative

assessments, with varying frequencies, throughout the year and analyzing data on a regular basis. Teachers' formative use of summative assessment data is most often demonstrated through analysis using central tendency statistics. Disaggregating data by content standards or student subgroups is not as frequently attempted. Regardless of the methods of data analysis, an overwhelming majority of teachers reported using assessment data results to evaluate their instructional practice and make changes to enhance student learning.

The assessment literacy level of teachers did not appear to have any influence on the extent to which they use summative assessments in a formative way. However, assessment literacy scores did differ across teacher characteristics. High school teachers had a higher assessment literacy score than elementary school teachers, and teachers with graduate degrees scored higher than those with a bachelor's degree. Experience mattered as well; more experienced teachers had a higher assessment literacy score than beginning teachers. Finally, science and mathematics teachers had a higher assessment literacy score than elementary teachers.

The findings of this study give building administrators and staff development leaders insight into current instructional practices of teachers. Additionally, a general measure of assessment literacy establishes a baseline from which educational leaders can develop future training to raise the assessment literacy of teachers

Chapter 1

Introduction

Statement of Problem

As two teachers commiserate at the end of a particular grading period, one is heard saying to the other, "I am not sure when I am supposed to teach anymore…it seems all we do with students these days is test and then test some more. Between my own tests I give to determine students' grades, and the benchmark tests we have to give for the division, and then the state-mandated tests, there is hardly anytime for me to cover the material needed for all these tests. I wish we would go back to the days when all I had to do was teach. I do not see how all this testing is helpful. From the looks of my students' grades on the latest benchmark, it is clear all this testing is taking its toll on learning!" As her colleague nods in agreement, both sigh and go back to work.

This is a typical scenario that plays out in schools across the nation as teachers feel the pressure of state and federal accountability models to document student learning. Studies indicate that assessment can be a powerful tool to enhance learning, not just measure achievement (Black & Wiliam, 1998b; Murnane, Sharkey & Boudett, 2005; Black, Harrison, Lee, Marshall & Wiliam, 2004). The benefit from assessments is not solely the measure of achievement, but in how the assessment data are used. The pressure to document achievement through mandated testing can be a struggle for teachers if they fail to recognize the formative nature of assessment and instead consider assessment separate from instruction.

This, then, raises a series of questions: To what extent are teachers using the information gained from analyzing student assessment data in formative ways? What instructional changes are teachers making as a result of analyzing student assessment data? Do teachers' levels of assessment literacy influence their use of assessment data? How are teachers analyzing assessment data and is there room for improvement?

Rationale for Study of Problem

Schools are charged with ensuring that all students learn. Today's state and federal accountability mandates bring added pressure to make sure schools are doing all they can to help students achieve. The current accountability model holds schools and students responsible for learning with school accreditation and student graduation hanging in the balance. Each state has developed some form of an accountability model. A primary instrument for accountability measures is a summative assessment in the form of end-of-course tests for selected content areas.

In an effort to gauge students' learning leading up to the state testing, teachers are often required to give periodic benchmark tests, work in professional learning communities to develop common assessments, and document instruction and learning through analysis of assessment data. However, rather than distract from teaching and learning, as the two hypothetical teachers mentioned earlier, another intended goal of testing can be to enhance learning. In other words, teaching in today's accountability climate not only includes assessment as a way to document student learning, it requires teachers to use assessment as a way of determining what to teach. When testing is approached as assessment for learning, learning can be enhanced (Black & Wiliam, 1998a; DuFour, 2004; Bernhardt, 2005; Streifer & Schumann, 2005; Stiggins, 2008). For this reason, it is important to determine the way teachers are currently using assessment data to formulate their instructional strategies to support and enhance student learning.

Statement of Purpose

The purpose of this study was to determine the extent to which teachers use summative assessment data in a formative way and how this formative use of assessment data influenced their instructional practices. This study proposed to measure the types of summative assessments teachers administer as well the frequency of teachers' administration and analysis of student assessment data. Additionally, this study proposed to examine teacher perceptions of their use of student assessment data and how it changed their instructional practice. Finally, this study measured teachers' classroom assessment literacy to see if any relationship existed between teachers' assessment literacy and their use of assessment data to inform instructional practice.

Literature/Research Background

The use of data to inform instructional practice is frequently referred to as Data-Driven Decision-Making (DDDM). As a model for improving student learning, DDDM is a multi-faceted approach that requires several key elements to be successful. Under the umbrella of DDDM fall five essential concepts (McLeod, 2005): 1) access to quality data, 2) measurable instructional goals, 3) frequent formative assessments, 4) professional learning communities, and 5) knowledge to take instructional action. For this study, two aspects of DDDM were examined: the use of formative assessment and teachers' knowledge to take instructional action.

Stiggins (2004) characterizes two types of assessments: assessments of learning and assessments for learning. Ideally, teachers employ various forms of assessments to find an instructional balance that enables them to evaluate how students are learning and then measure what students have learned (Stiggins, 2004). Assessments of learning are summative assessments such as teacher-generated tests administered at the end of an instructional unit,

common departmental assessments designed to ensure equitable and adequate student learning across different instructors, division benchmark tests to gauge student achievement, and state end-of-course tests or nationally norm-referenced tests administered to document student achievement.

Assessments for learning could be considered formative assessments, assessments teachers used to gauge student learning and to gain information to adapt and adjust instruction according to students' needs. It is not the assessment itself that is classified as summative or formative but rather how the information is used from that assessment (Black & Wiliam, 2003). When information obtained from various assessment tools is used to adapt and meet the instructional needs of students, then formative assessment is present in the classroom.

Rather than two different assessment methods, assessments typically considered a summative measure of student learning can be used in a formative way. For example, Bernhardt (2000) suggests the power of data analysis comes at the intersection of various data sets. When teachers consider disaggregated summative assessment data compared with student subgroups or instructional content, and then plan instruction and evaluate the effectiveness of their instruction, they are taking a summative assessment and using it in a formative fashion (Bernhardt, 2000). This study examined some ways in which teachers used summative-type assessments in a formative way by first examining how frequently they administered summative type assessments and what instructional changes they made as a result of their analysis of student assessment data.

The second aspect of DDDM considered for this study was teachers' knowledge to take instructional action. Without sufficient knowledge of classroom assessment, effective analyses of assessment data and subsequent action could be undermined. An evaluation of teacher assessment literacy provided insight into the ability of teachers' to use student assessment data to

inform their instructional practice. This was very timely as teachers' assessment literacy has not always been at the forefront of educational research despite the assumption that teachers spend a considerable amount of time assessing students' learning. It has only been in the last two decades that standards for teacher competence in educational assessment of students were even developed (Plake, 1993). Consequently, this study measured teachers' assessment literacy to see if any relationship existed between teachers' use of summative assessment in a formative way and their level of assessment literacy.

Research Questions

The overarching research question that guided this study was: to what extent do teachers use summative assessment data to inform their instructional decisions? To address these questions, several research questions were developed to evaluate the types and frequency of summative assessment data and data analysis methods used by teachers, how these analyses influenced their instructional practice, as well as an evaluation of teachers' assessment literacy. Four research questions were asked to measure the extent to which teachers use summative assessment data to inform their instructional practice:

1. To what extent do teachers use summative assessments in a formative way?
2. What is the level of teacher assessment literacy in a large suburban school district?
3. What is the relationship between teachers' reported use of summative assessment in a formative way and their assessment literacy level?
4. What changes in their instructional practice result from teachers' use of assessment data?

Design and Methodology

This study employed a survey research design. A stratified random sample of 1500 teachers from a large, suburban school district in central Virginia was surveyed at the end of the 2008-2009 academic year. Limiting the survey to one specific district ensured that all teachers were operating under the same division expectations with respect to district level policies and expectations for the administration of summative assessments. This division had made a concerted effort the past two years to implement professional learning communities with an emphasis on collecting and sharing assessment data to document student achievement and plan instruction to meet the needs of students. Along with professional learning communities, this division set the expectation that teachers would develop and administer common departmental assessments to help gauge student learning. Division benchmark tests were developed for those grade-levels and courses associated with a state-mandated, end-of-course assessment. Benchmark assessments were required for all mathematics courses associated with an end-of-course assessment, and for those schools who failed to make adequate yearly progress in language arts, division benchmark assessments were also required. All other benchmark assessments were required at the principals' discretion. Additionally, teachers in this division had access to an electronic gradebook program that provided various statistical analyses if teachers wished to analyze data.

An email invitation to participate in the survey was sent to 500 teachers at each level: elementary, middle, and high school. The survey consisted of three main sections. The first section gathered demographic data to determine characteristics of the sample. Besides gender and ethnicity data, the survey solicited responses with respect to school level taught, years teaching experience, years in current teaching assignment, degree attained, and primary teaching

responsibility. The second section gathered frequency data regarding specific types of summative assessments administered as well as the methods and frequency of data analysis. This section also evaluated types of changes teachers had made in their instructional practice as a result of their analysis of assessment data. Part three of the survey used the Classroom Assessment Literacy Inventory (Plake, 1993) to measure teachers' knowledge of assessment.

The survey was administered using a web-based survey program, Survey Monkey. Respondents were provided with a link to access the survey and a two-week window to respond. Once accessed, the program blocked all collection of email or IP addresses thus protecting the anonymity of respondents. A follow-up email was sent one week into the two-week window to thank teachers for responding, or if they had not accessed the survey, to remind them to do so.

Data were analyzed using frequency distributions to give a description of the types and frequency of summative assessments administered as well as the methods and frequency of data analysis. A composite score for each assessment standard included in the Classroom Assessment Literacy Inventory was calculated and mean scores were compared for various teachers' characteristics: school level, years' experience, degree attained, and primary teaching responsibility.

The use of summative assessment in a formative way was operationalized as the measure of teachers' self-reported frequency of data analysis and data analyses methods for specific types of summative assessments. Multivariate analyses of variance were conducted to determine if any relationships existed between teachers' use of summative assessment in a formative way and their level of assessment literacy.

Summary

The need to ensure that all students master content material and demonstrate mastery on

state-level assessments increases the need for teachers to be much more savvy assessment data consumers. When teachers have a solid foundation in assessment literacy and use summative assessment data in a formative way the effect can be significant, especially for low-attaining learners (Black & Wiliam, 1998b). Examining data-driven decision making (DDDM) helps deepen the understanding of what is happening in today's classroom and gives educational leaders some insight to what can be done to strengthen teacher practices.

As federal and state legislation bring pressures to bear on schools to increase learning for all students, it behooves educational leaders to examine the extent to which teachers are using data obtained from various assessments in the classroom to inform their instructional practices and subsequently improve student learning. Additionally, educational leaders need to understand teachers' assessment literacy levels to increase the efficacy of instructional practices enhanced by teachers' use of assessment data.

The findings of this study will help school leaders identify and target specific needs for future professional development with respect to analyzing and using data to make instructional decisions. Like the hypothetical teachers mentioned at the beginning of this chapter, teachers are working hard yet perhaps are overlooking a powerful instructional tool. Rather than regard assessments as a distraction from learning, perhaps teachers will find assessments can enhance learning.

By determining the types and frequency of assessment data teachers are currently using, school leaders might gain an insight to current practice and identify areas of need. Teacher assessment literacy levels could have a direct impact on future staff development. Teacher characteristics, such as school level taught, year's experience, degrees attained, or primary teaching responsibility, could influence the motivation and effective use of student assessment

data. Any relationship found between teachers' assessment literacy level and their use of assessment data in a formative way could provide insight to future staff development or organizational structure that would enhance collegial collaboration to promote and enhance teachers' use of student assessment data.

The use of assessment data is an emerging field in education today. Increasing state and federal standards requires school leaders to tap all potential benefits when it comes to student learning. Measuring teachers' current use of assessment data to inform their instructional practice, along with their assessment literacy level, is a first step to reaping what may be an untapped potential in every classroom: a teacher who can effectively use assessment data to improve their instruction thus increasing student learning at all levels.

Definition of Terms

Assessment. Any student or teacher activity that yields information regarding what the student has learned.

Assessment literacy. Knowledge of assessment as set by the Standards of Teacher Competency in Educational Assessment of Students. Assessment literacy can be defined as the process of obtaining information to judge the effectiveness of instruction, the adequacy of the curriculum, to give feedback to students about their strengths, weaknesses and progress, and to make decisions about instruction ("Standards for Teacher Competence in Educational Assessment of Students," 1990).

Adequate Yearly Progress (AYP). Designation determined by federal legislation No Child Left Behind that requires documented academic benchmarks that include, but not limited to, academic achievement for various subgroups of students who may be considered at risk for low achievement.

AYP Subgroups. Subgroups of students identified by the federal legislation No Child Left Behind.

Data-Driven Decision-Making. "A system of teaching and management practices that gets better information about students into the hands of classroom teachers" (McLeod, 2005, p.1).

Departmental common assessments. Assessments generated through collaboration among teachers within a given department. Assessments that are the intended to measure student learning based on common instructional goals developed through professional learning communities.

Disaggregated assessment data. Assessment results that are broken into similar categories, such as specific topics or student ethnic groups, intended to compare results within the category.

Division benchmark assessment. A division-generated assessment intended to gauge student mastery of standards of learning for specific content-related concepts. Typically administered on a quarterly basis.

Formative assessment. The specific use of student assessment data to modify or plan instruction.

Professional learning community. A small group of teachers with either a common content or a common group of students, who meet regularly to share and discuss student performance and instructional strategies.

Statewide Standards of Learning assessment. State-generated assessments intended to measure student mastery of specific content. Used for accountability purposes.

Nationally norm-referenced assessment. Nationally developed test that measures and

compares individual student performance against all other students within same grade level.

Chapter 2

Literature Review

Introduction

This literature review draws on research that addresses three primary areas of interest. First, what does Data-Driven Decision-Making look like in practice and what specific structures should schools have in place to foster and enhance teachers' use of student assessment data to inform and influence their instructional decisions will be explored. Secondly, the use of formative assessment by teachers is examined to determine the effect this practice has on student achievement. In conjunction with formative assessment, teachers' knowledge of classroom assessment is explored. Since four of the seven Standards for Teacher Competence in Educational Assessment of Students are related to teachers' choice and analysis of student assessment data, assessment literacy is closely tied to the instructional use of assessment data. To that extent, this literature review examines all three concepts to determine current educational best-practice for each, and how these concepts interact in order to more effectively evaluate the findings from this study.

This study examined the extent to which teachers use assessment data to inform their instructional practices and how their level of assessment literacy may inform this process. The current educational reform movement could be traced back to the seminal report, A Nation At Risk, by the United States National Commission on Excellence in Education (National Commission on Excellence in Education, 1983). What started a few decades ago as a call to

action for educational reform has increased in magnitude to the extent that today's educational climate is one of accountability through the use of mandated state testing. A quarter century's worth of research seems to indicate that the educational community is now examining the science of teaching as much as it has examined the art of teaching in the past (Wilson, 2007). Data-Driven Decision making, incorporating teachers' use of formative assessment, the increasing demands on teachers to evaluate and interpret students' assessment data, and how this might raise student achievement, makes this topic especially relevant in today's educational, high-stakes, accountability climate.

Data-Driven Decision-Making

The phrase, Data-Driven Decision-Making (DDDM), has become increasingly heard in discussions regarding how to improve student achievement in schools. Many educational researchers have devoted much of their efforts to define DDDM and determine the factors that enhance and enable effective DDDM (Bernhardt, 2004; McLeod, 2005, Streifer & Schumann, 2005; Murnane, Sharkey & Boudett, 2005; Lachat & Smith, 2005; Wayman, Midgley & Stringfield, 2005; Brunner, et al., 2005; Mandinach, Honey & Light, 2006). The growing body of research suggests that the topic is becoming increasingly important as educators navigate the ever-rising accountability standards.

Bernhardt (2000; 2004) takes a systems approach to data analysis examining the types of data that can be collected and how cross-sections of data can be revealing in tracking and planning for student achievement. For data-driven decision-making to be most effective, schools need to collect and organize data with respect to demographics; student, parent, and teacher perceptions; school processes such as specific academic programs; and student achievement. The added value of analysis comes in the intersections of these data to address specific questions

teachers, administrators, and division personnel might have. For example, looking at cross-sections of demographic and student learning data, achievement gaps can be identified and appropriate interventions taken. Expanding data analysis to look at demographics, student learning, and school processes can identify those programs that are especially effective with certain populations while targeting for review those that are not (Bernhardt, 2000). Appropriate disaggregation of these various banks of data allows division personnel, administrators, and teachers to allocate resources and monitor results more effectively.

Bernhardt (2000) notes there are challenges to the effective use of data for school improvement. According to Bernhardt, teachers and administrators may feel formal data analysis is not necessary as they are in the building everyday and are aware of which students are having trouble and which are not. Lack of training, limited access to data, and low assessment literacy levels can impede the effective use of data. Finally, there is the real possibility the reluctance to analyze data is born out of fear. Teachers' worry that data analysis may reveal evidence of incompetence or be used as an evaluative tool can undermine the effective use of data within the classroom to improve student achievement (Bernhardt, 2000).

Mandinach, Honey, and Light (2006) developed a theoretical framework for DDDM that assumes all individuals, regardless of their position within a school system, have a need that requires the use of data. District personnel use data differently than building administrators and building administrators have a different purpose for data than teachers. However all have a need to collect and analyze data to make informed decisions. Mandinach, et al., depicts a DDDM continuum that moves from the collection and organization of data to information gleaned from data. This information then is synthesized into knowledge that informs instructional decisions, from here decisions lead to instructional actions. The impact of the instructional decision and

resulting action is then measured through collection of additional data and the process starts over again (Mandinach, Honey & Light, 2006).

As noted by Mandinach, et al. (2006), school issues can affect the use of data. The proliferation of technology tools to collect and organize data has left many schools data rich but information poor. The challenge comes in taking the myriad sources of data and turning data into actionable knowledge. Though this theoretical framework for DDDM addresses the flow of data to action, it does not necessarily identify what that action might be (Mandinach, Honey & Light, 2006).

McLeod (2005) approaches DDDM from a teacher's perspective. He defines DDDM as a "system of teaching and management practices that gets better information about students into the hands of teachers." (McLeod, 2005, p.1) However, for DDDM to be most effective, teachers need to shift focus from delivery of instruction to achievement of results. In his white paper explaining factors that contribute to effective data-driven decision-making, McLeod expounds on what he considers five essential concepts for effective DDDM. Teachers must have access to all types of assessment data to help them improve their instructional practice. Additionally, teachers need a solid foundation in assessment literacy. Access alone is useless without the knowledge to appropriately interpret data and communicate the results to principals, students, and parents. Secondly, teachers must use assessment information to set measurable goals for student achievement. McLeod recommends school personnel employ SMART goals; goals that are specific, measurable, attainable, results-oriented and time-bound. The ability to use data to identify and set SMART goals is critical to DDDM (McLeod, 2005).

Third on the list of essential concepts deals with teachers' abilities to administer and analyze assessment in ways that are formative. Frequent formative assessments provide

benchmarks for teachers to evaluate progress towards instructional goals. Research has shown the ability to plan, implement, and analyze formative assessments is a powerful tool to raise student achievement. The use of formative assessment has been found to raise student achievement. Black, et al., found the average effect size for the use of formative assessment in the classroom was approximately 0.3 standard deviations. If these results could be produced across a school, it would raise a school in the lower quartile to well above the national average (Black, Harrison, Lee, Marshall & Wiliam, 2004; Wiliam, Lee, Harrison & Black, 2004).

Fourth on McLeod's list of essential concepts is the establishment of professional learning communities. The power of formative assessment is increased when teachers move from teaching and analyzing data in isolation to a collaborative culture where teachers plan and support each other. Professional learning communities provide opportunities for teachers to learn from each other. Finally, the power of data is diminished unless its analysis leads to action. McLeod's final essential concept for DDDM calls for teachers to use the information obtained from data analysis to develop focused instructional interventions to improve student learning (McLeod, 2005).

Lachat and Smith (2005) conducted a qualitative study of teachers' use of data in five low-performing urban high schools that were undergoing reform in order to raise student achievement. Though the findings were limited, Lachat and Smith found implementing the use of data to positively affect student achievement is possible but the study revealed some significant challenges. First and foremost, efficient access to quality data required schools to examine ways in which data was warehoused and accessed. High levels of student mobility can exacerbate efforts to streamline accurate and timely data retrieval. Additionally, better communication and understanding between those responsible for producing data reports and

those requesting data reports is necessary. Without direct access to databases, teachers must have knowledge of what types of data are useful for their specific need in order to request student data for instructional purposes (Lachat & Smith, 2005).

Their study also found the ability to disaggregate data was quickly realized as a particular benefit. Teachers' preconceived notions were confronted upon close examination of disaggregated data. For example, teachers at one school felt low attendance issues were the cause of poor student performance. Examining disaggregated data, comparing the performance of students with low attendance compared to those with acceptable attendance, revealed significant deficiencies in both groups. This prompted teachers to examine their instructional processes rather than blame attendance for poor student performance (Lachat & Smith, 2005). "The data thus confirmed the school had two problems – student attendance and quality of instruction. Reviewing the data and eliminating teacher assumptions that the problem was only an attendance issue allowed for more productive discussions about the content and quality of instruction" (Lachat & Smith, 2005, p.342).

As McLeod (2005) noted in his framework for DDDM, Lachat and Smith (2005) also found a collaborative culture enhanced teachers' use of data. Collaboration, in conjunction with examination of data, allowed focused questions to emerge that motivated teachers to look past the data and analyze what the data meant for their school. Finally, it was found that effective leadership that modeled the appropriate use of data was essential for a systemic change in attitudes and uses of data within the school (Lachat & Smith, 2005).

A case study of a low-performing middle school in California illustrated the powerful impact of teachers' accessing and analyzing disaggregated data and using the results to make instructional decisions. Torch Middle School found the use of data to make instructional

decisions resulted in a phenomenal jump in students' achievement levels (Nelson & Eddy, 2008). Based on California's Academic Performance Index (API), with a scale of 200 to 1000, Torch Middle School was given a score of 435 for the 2000-2001 academic year. Five years later, in 2006, after a concerted effort to use data to make instructional decisions, Torch's API score improved significantly with a score of 719 (Nelson & Eddy, 2008). Though the study cites other factors that may have contributed to the improvement, administrators' and teachers' use of data changed the climate of the school and its expectations for students. This study illustrates the potential for improvement when quality data is available, teachers are knowledgeable data consumers, and the school structure promotes collaboration that fosters effective evaluation practices (Nelson & Eddy, 2008).

The power of data in helping teachers make instructional decisions to positively affect student achievement is not without its challenges (Wayman, 2005). The ability to efficiently organize and access data is limited by the technology available within school divisions. To effectively implement DDDM, schools must have the technology infrastructure to support data storage and efficient retrieval (Sharkey & Murnane, 2003; Wayman, 2005). Other studies cite the need for teacher collaboration and a systematic plan for a school division to implement DDDM (Wayman, Midgley & Stringfield, 2005; Wayman, Cho & Johnson, 2007; Brunner, et al., 2005; Streifer & Shumann, 2005). Teachers' knowledge of appropriate assessment practices and analyses, as well as the opportunity and motivation to act on the data, are also challenges that can undermine DDDM no matter how sophisticated the technology (Sharkey & Murnane, 2003). Data-driven decision-making is a complex process that defies a one-size-fits-all application. However, as these studies have shown, it has the potential to enhance instruction and raise student achievement.

Formative Assessment – Using Data to Make Decisions

Teaching requires constant decision-making. Teachers continually make instructional decisions about things such as what to teach, how to teach, how long to teach, or what questions to ask (McMillan, 2001). However, the extent to which teachers collect and gather assessment data and use this information to make instructional decisions is less well known. Regardless of the technology available or the existence of a collaborative culture, if teachers are not examining data and making instructional decisions based on data, little benefit from any model of data-driven decision-making will be fully realized.

In the late 1960s, Bloom described what he called mastery-learning theory (Bloom, 1968). Bloom intentionally used assessment data in a formative way by identifying weak areas of learning for targeted instruction. He found 80% of the students who experienced this strategy of assessment and instruction earned an A on the final exam, compared with just 20% the previous year. The following year the improvement was even greater; results showed 90% of the students participating achieved an A on the final exam (Bloom, 1968). Guskey (1990), using the framework of Bloom's mastery learning theory, reported the results of a study where teachers emphasized student mastery learning with a group of mildly, learning-disabled and at-risk elementary school students. Through the use of frequent formative assessments, the mainstream teacher and the special education teacher were able to adjust instruction to address students' instructional needs. The findings showed those students in the treatment group realized a 13.64% increase on the state mastery achievement test whereas those students in the non-treatment group only gained 3.89%. Teachers reported the benefit went beyond increased student achievement. Student engagement was increased and behavior issues were reduced. Teachers' use of data and subsequent adjustment of instruction provided a very real benefit to

enhance student achievement (Guskey, et.al., 1990).

In 1998, Black and Wiliam (1998a) referred to the classroom as a black box – an area of unknown content. Putting much effort into analysis of external inputs such as students, teachers, management rules and requirements and measuring outputs in terms of student achievement, no one was asking exactly what do teachers do in the classroom to enhance student learning? Black and Wiliam's meta-analysis of research regarding the effect of teacher assessment practices in the classroom revealed exciting possibilities for raising student achievement through classroom assessment (Black & Wiliam, 1998a; 1998b). Most promising was teachers' use of assessment data to inform their instructional practice. Assessments were defined as activities "undertaken by teachers – and by their students in assessing themselves – that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes formative assessment when the evidence is actually used to adapt the teaching to meet the student needs" (Black & Wiliam, 1998a, p.140). Looking into the black box of education, the classroom, Black and Wiliam sought to determine exactly how teachers used assessment data to positively affect student learning.

Classroom assessments can be classified into two categories. Summative assessments are those assessments that are given at the end of a chapter or unit and serve the purpose of certifying what students know or evaluating the effectiveness of a curriculum. Formative assessments are those assessments intended to help students and/or teachers evaluate what has been learned and determine what else needs to be done in order to meet the current instructional objective. The terms summative and formative refer more to the function of a particular assessment than the type of assessment (Black & Wiliam, 2003).

Having examined over two hundred studies that indicated the use of formative

assessment can improve student learning, Black and Wiliam (2004) then asked the question: What does formative assessment look like in the classroom and is there room for improvement? To answer these two questions, Black and Wiliam had twenty-four teachers participate in a mini-experiment whereby each teacher determined what data were available to them and developed an output measure for his/her class. A comparable classroom provided a basis for comparison. Teachers concentrated on developing questioning skills that included sufficient wait time to allow students to formulate their answers. Feedback on student work was given special attention and, where possible, comments for improvement were issued rather than grades. Opportunities for peer and self-assessment allowed students to evaluate how their work was meeting the instructional goals. Summative tests were used in a formative fashion, having students evaluate what they did and did not understand and develop their own strategies to clarify and remediate concepts needing reinforcement (Black, Harrison, Lee, Marshall & Wiliam, 2004). The study was limited by the inability to have two identical test groups for each teacher participant. Teacher attrition and the inability to control confounding variables also limited the findings. After accounting for these variables, the average effect size was found to be approximately 0.3 standard deviations. This indicated that if these results could be produced across a school, it would raise a school in the lower quartile to well above the national average (Black, Harrison, Lee, Marshall & Wiliam, 2004; Wiliam, Lee, Harrison & Black, 2004).

Other studies have shown positive results from teachers' use of data. The Grow Network Study examined how educators within New York City public school system used a web-based reporting system to gather data to inform their instructional decisions. Using structured interviews, ethnographic research, and surveys, researchers found teachers utilized the Grow Reports to make decisions about their instructional practices in four specific areas: 1) prioritizing

instruction time and effort to meet the needs of diverse learners, 2) communicating student achievement with administrators, teachers, parents, and students, 3) prompting teachers' reflection on instruction and shaping their professional development, and 4) encouraging students to be more self-directed by giving data to students (Brunner, et al., 2005). Of particular interest is the finding that over 70% of teachers reported the Grow Reports prompted them to be more self-reflective and critically evaluate the effectiveness of their instruction. Some reported that prior to the Grow Report, they assumed they were sufficiently addressing all the learning standards but after seeing a report of their student assessment data they recognized areas that needed improvement. New teachers found the Grow Report especially helpful, giving them the ability to determine what needed their instructional attention (Brunner, et al., 2005).

To be effective, formative assessments must occur frequently. Concern has been raised that adding assessments into the instructional plan in an effort to generate data to ostensibly raise student achievement has a counter effect – increased assessments mean less time to teach and thus lower student achievement. Yeh (2006) found these concerns to be unfounded. His study examined the use of a rapid assessment system for K-12 math and reading in one Texas school division. The rapid assessment system is designed to give administrators, teachers, and students' feedback regarding student progress. Over 87% of the teachers participating reported the rapid feedback made them more effective, allowing them to make immediate instructional adjustments. The constant flow of assessment data gave teachers common benchmarks to gauge student progress and prompted more collaboration as teachers discussed what worked and did not work in their classroom. Increased teacher confidence in instructional decisions provided more motivation to engage students in higher-order thinking. Rather than having a deleterious effect on student achievement, frequent testing with quick feedback allowed teachers to raise not

only student achievement on standardized tests but also the level of instruction and student thinking in their day-to-day classroom activities (Yeh, 2006).

Petersen (2007) studied three different schools across the country that made a concerted effort to use data to inform instruction. All three schools reported a positive effect. Serving populations that would be considered challenging, teachers at each school found consulting data helped inform their instructional practice resulting in raised student test scores. One elementary school realized an increase in the pass rate for math scores from 65% in 1994 to 98% in 2002 (Petersen, 2007). Even with increased state standards, the school has maintained a passing rate in the 90s. A small private preparatory school in Connecticut, serving kindergarten through eighth grade, found establishing a data-collection culture paid huge dividends in student achievement. Within the first year of the drive to use data, the percentage of kindergarten and first-grade students reading at or above grade level rose from 26% to 96% (Petersen, 2007). A charter school in California found increased collaboration among teachers significantly raised student achievement. With 53% of the student population classified as English-language learners, students' proficiency in language arts rose from 17 percent in 2002 to 32% in 2005, math proficiency increased from 23% to 53% during the same time frame (Petersen, 2007).

These studies have shown that teachers' use of data is varied and the effects are broad. From informal observations, to quick checks for understanding, to effective questioning employing wait time as an instructional strategy, to disaggregating data and collaborating with colleagues in a professional learning community, teachers' data use can take many forms. The literature provides a clear picture of how teachers are using data and guides this study by helping identify what specific activities and attitudes to examine when trying to determine the extent to which teachers are using data to inform instructional decisions.

Classroom Assessment Literacy

Assessment literacy standards for educators were not developed and published until 1990 ("Standards for Teacher Competence in Educational Assessment of Students," 1990). Through the joint efforts of the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association, seven competencies for teachers' knowledge of assessment were identified. These standards cover a range of teacher assessment activities that occur before, during and after instruction, making decisions based on assessment information, as well as collaborative activities with other educational professionals. Based on the standards for teacher competency in assessment, teachers should be:

1. Skilled in choosing appropriate assessment methods to make instructional decisions.
2. Skilled in developing appropriate assessment methods for instructional decisions.
3. Skilled in administering, scoring and interpreting assessments, both externally produced and teacher-generated assessments.
4. Skilled in using assessment data to make decisions about student learning, to plan lessons, develop curriculum, and for school improvement.
5. Skilled in developing valid grading procedures that use student assessments.
6. Skilled in communicating assessment results to students, parents, the community, and other educators.
7. Skilled in recognizing unethical, illegal, and inappropriate assessment methods and uses of assessment information

("Standards for Teacher Competence in Educational Assessment of Students," 1990). With the onset of state and federal accountability models that demand high learning

levels for all students, it has become evident that the old models of assessing student learning are outdated (Stiggins, 2008). To enhance learning, assessment data must be pre-planned and used throughout students' learning cycles. As Stiggins (2008) suggests in his recent assessment manifesto, teachers, administrators, and school divisions must adopt a balanced approach to assessment, employing all the uses of assessment data to its full advantage. This requires teachers to have a basic knowledge of assessment practices. Considering the extended time it takes to move from the development of theory to implementation of practice, it is no surprise that teacher assessment literacy is still a topic of debate and concern almost two decades after the publication of the assessment literacy standards. Stiggins (2008) use of the rather strong term manifesto in his report, *Assessment manifesto: A call for the development of balanced assessment systems*, shows the topic of assessment literacy to still be a matter of concern for educational leaders. The ensuing two decades after the publication of the assessment literacy standards has seen misguided reform efforts. Many reform efforts have focused on the development and administration of standardized tests rather than the appropriate interpretation of those tests (Stiggins, 2008).

Illinois embarked on an ambitious reform movement to connect student assessment to curriculum, instruction, and school improvement. Starting in 1993, efforts to improve student achievement through various assessment reform initiatives failed to find traction, in part due to lack of assessment literacy among educators. The focus on large-scale assessments virtually ignored the development of valid and reliable classroom assessments, all dependent upon teachers' assessment literacy (Vogel, Rau, Baker & Ashby, 2006). After three iterations of reform efforts, Illinois adopted the Standards-Aligned Curriculum Initiative (SAC). One outcome of the SAC initiative was the realization that though curriculum aligned with state

standards was essential, if teachers lacked knowledge to understand and interpret complex data reports, the full value of SAC would not be realized (Vogel, Rau, Baker & Ashby, 2006). Although the content of the curriculum may have changed through SAC, instructional practices have remained the same. Researchers evaluating the reform movement in Illinois recommended an increased focus on professional development for teachers specifically designed to increase their assessment literacy (Vogel, Rau, Baker & Ashby, 2006).

Murnane, Sharkey, and Boudett (2005) found the same issues with teacher assessment literacy when working with teams of teachers from ten different Boston public schools. They found that along with limited time and support to use data to improve instructional practice, teachers and administrators also lacked expertise and an understanding of the potential that lies in the use of student assessment data. In their work with teachers, Murnane, et al., found three approaches to the use of assessment data. An instrument approach uses data to make decisions about which students have mastered the material and which have not. For example, this type decision might determine who require summer school and who does not. A symbolic approach could be used to justify a decision that had already been made, such as citing assessment data to support a specific program or activity for students. The third approach, a conceptual approach, uses assessment data to evaluate what students know, what they can do, and how effective instruction is. Murnane and his colleagues found the conceptual approach to the most valuable and the most underutilized of the three. This approach requires assessment literacy and underscores the need for such knowledge to make effective instructional decisions based on assessment data (Sharkey & Murnane, 2003; Murnane, Sharkey & Boudett, 2005).

A study by Mertler (2003) evaluated the assessment literacy of preservice and inservice teachers. Using the Classroom Assessment Inventory (Plake, 1993), based on the Standards for

Teacher Competence in the Educational Assessment of Students ("Standards for Teacher Competence in Educational Assessment of Students," 1990), Mertler found inservice teachers scored slightly higher on assessment literacy levels compared to preservice teachers. For preservice teachers, Mertler found the composite scores for each standard ranged from 2.06 to 3.25, out of a maximum possible score of 5. For inservice teachers, scores ranged from 2.06 to 3.95 (Mertler, 2005). The results of this study mirrored results from a national study conducted ten years prior where researchers found inservice teachers' assessment literacy scores ranged from 2.70 to 3.96 (Plake & Impara, 1993).

For a teacher, assessment would seem the natural complement to teaching. However, the use of assessments in and of itself is will not ensure all students will learn. Teachers' attitudes and knowledge of the use of assessment can impact student learning (Husman, Brem & Duggan, 2005). Students' perceptions of their teachers with respect to teachers' emphasis on mastery objectives, learning to learn, versus performance objectives, learning to score well on a test, has a direct effect on the students' achievement. In Husman, Brem, and Duggan's (2005) study, they found that when students perceived their teacher had a performance orientation, students were motivated to do well on the test and approached learning as simply a means to do well on the test. If students perceived their teacher had a mastery orientation, students adopted a mastery orientation and were motivated to learn the material, not just do well on the test. Furthermore, the study showed a mastery orientation also accompanied an increase in performance orientation enhancing student learning as well as student performance on assessments (Husman, Brem & Duggan, 2005).

In 1999, the Nebraska State Department of Education developed a teaching endorsement in classroom assessment. After completing the program of study, teachers reported positive

benefits for themselves and their students as a result of their additional training. Teachers became more cognizant of the need to integrate assessment into their instruction rather than separate assessment from instruction. One participant reported an increased "awareness of assessment as a part of instruction, not an add-on" (Lukin, Bandalos, Eckhout & Mickelson, 2004. p.28). The Nebraska program found the full value of assessment and analysis is diminished unless it is appropriately used to make instructional decisions designed to improve student learning (Lukin, Bandalos, Eckhout & Mickelson, 2004).

As a follow-up to the endorsement in classroom assessment, Lincoln Public Schools also implemented Assessment Literacy Learning Teams, based on Stiggins' learning team approach. Small teams of teachers, plus other educational professionals, met regularly to discuss key assessment concepts. Participation was voluntary and teams were implemented at selected schools. The preliminary outcome indicated a positive effect on teacher confidence in making assessment and instructional decisions. Additionally, some students reported a more positive belief and attitude towards learning. These results, however, are somewhat limited because data was only collected at one high school. Despite the limited study, the researchers concluded there appears to be a positive relationship between teachers' assessment literacy and effective use of student assessment data to inform instructional practice (Lukin, Bandalos, Eckhout & Mickelson, 2004).

Assessment literacy has slowly moved to the forefront of educational discourse as high-stakes testing has increased the pressure on teachers to produce results. Accountability has increased the demand for teachers to be savvy data generators, planning appropriate assessment for instructional purposes. No longer able to rely on the one traditionally trained, division psychometrician to perform data analysis, teachers must be assessment literate to support

learning in their classroom (Lukin, Bandalos, Eckhout & Mickelson, 2004).

Summary

Teachers' use of data to inform their instructional practice is an essential part of an effective data-driven decision-making model. This use of assessment data, classified as formative assessment, has promising results for raising student achievement, especially for low-attaining students (Black & Wiliam, 1998). However, formative assessment by its very definition requires data be used to provide information to students and teachers and teachers have the knowledge to put the data to good use to improve teaching and learning (*Classroom Assessment and the National Science Education Standards*, 2001; Black & Wiliam, 1998a). Data-Driven Decision-Making is the link between these two concepts: formative assessment and teacher assessment literacy. Effective use of data to inform instructional practice requires not just specific steps of data-driven decision-making, but the effective use of formative assessment and the assessment literacy to know how to interpret data and use the analysis to make effective instructional decisions.

The research is rich with studies and journal articles regarding the advantage for student achievement when educators use data to make instructional decisions. Additionally, there is a decade of research indicating the need for improvement in assessment literacy for teachers. Interestingly, no research was found that linked these two concepts together. It would make sense that in light of the advantage of formative assessment, a link between the effective use of assessment data and teacher assessment literacy level be examined. As two essential components of data-driven decision-making, this study proposes to look at both concepts to determine any relationship.

The literature points to the value of these concepts: formative assessment, teacher

assessment literacy, and data-driven decision-making. The gap in the literature, especially linking all three together, shows there is still some work to be done. Bringing this line of inquiry together, this study sought to determine if a relationship existed that might amplify the clear benefits of having teachers knowledgeable in assessment literacy, collecting and analyzing quality formative assessment data for all students, and the educational practice of data-driven decision-making that put all to effective use.

Chapter 3

Methodology

Introduction

The purpose of this study was to determine the extent to which teachers use assessment data to inform their instructional practices. Specifically this study examined the types and frequencies of assessments administered and the data analyses performed by teachers, and how their use of assessment data changed their instructional practice. Teacher's classroom assessment literacy was also measured to examine any potential relationship between teachers' knowledge of assessment and their use of assessment data to inform their instructional practice. The study provided a descriptive baseline for a major, suburban school division with respect to teachers' use of data, their assessment literacy level, and a comparison of these factors against specific teacher characteristics that could influence their use of data.

With the advent of test-based accountability models, school divisions need to make the most effective use of student assessment data in order to support student learning to meet state and federal accountability standards (Stiggins, 2008). Though much work has been done to determine the necessary components of Data-Driven Decision-Making (DDDM), and the readiness of divisions to implement DDDM (Bernhardt, 2000; 2004; Murnane, Sharkey & Boudett, 2005; McLeod, 2005), the need still exists to determine the extent to which teachers are using data to make instructional decisions. This research describes the current landscape of two aspects of DDDM, the use of formative assessment and knowledge to take instructional action,

for a large, suburban school division in central Virginia. The results of this study identify teacher practices with respect to the administration and analysis of specific types of summative assessments, methods of data analysis, and teachers' instructional changes due to data analysis. Additionally, the measure of teacher assessment literacy provides a baseline that allows a comparison among teacher characteristics and the use of formative assessment.

Research Questions

The over-arching question for this study asked to what extent are teachers using summative assessment data to inform their instructional decisions? To address this broad question, several primary questions were developed to guide the research:

1. To what extent do teachers use summative assessments in a formative way?
2. What is the level of teacher assessment literacy in a large suburban school district?
3. What is the relationship between teachers' reported use of summative assessment in a formative way and their assessment literacy level?
4. What changes in instructional practice result from teachers' use of assessment data?

Assessment literacy data were examined with respect to teacher characteristics that might have a bearing on teachers' use of summative assessment to inform their instructional practice: school level, years' experience, degree attained, and primary teaching responsibility.

Research Design

This study employed a non-experimental, descriptive design and used survey research methods. Survey research is versatile in addressing questions with many facets such as those posited in this study. Of particular value is the generalizability of the findings to larger, similar

populations (McMillan, 2004, p. 195). Dillman (2000) noted the strength of survey research as, "The ability to estimate with considerable precision the percentage of a population that has a particular attribute by obtaining data from only a small fraction of the total population is what distinguishes surveys from all other research methods" (p. 9).

A web-based survey allowed efficient distribution of the survey to the target population. The use of email to administer surveys has been shown to produce comparable response rates to paper-mail surveys and quicker response times (Schaefer & Dillman, 1998). One concern with regard to use of email to conduct surveys was coverage error (Dillman, 2000). However, in this study the target population was K-12 teachers in a large suburban school division. Consequently, all members of the target population had confirmed email addresses thus limiting coverage error.

This study involved K-12, core-academic teachers in large, central Virginia suburban school division. The survey was administered via email contact during the fourth quarter of the academic year. By that time teachers had completed seven months of instructional time with their current students. With state mandated testing quickly approaching, as well as the school year drawing to a close, teachers had most likely used all the tools in their instructional arsenal to ensure their students were prepared. With the school year fresh in their minds, they were able to answer survey questions regarding their use of assessment data throughout the year. To maximize response rates, care was taken to assure the timing of the survey did not conflict with the run up to the state-testing window, otherwise teachers might have felt too over-burdened to respond.

District Profile and Teacher Population

This study focused on core-content teachers in a large suburban school division in central Virginia. The division had 64 schools and approximately 58,500 students enrolled. Student

demographics showed approximately 61% of the student population was white, 27% black, 7% Hispanic, and 3% Asian. Approximately 23% of the student population received free or reduced lunch at the elementary and middle school level. High schools in this division did not participate in federal lunch program so no data is available for that level ([School Division] Profile[1]). This school division met state and federal accountability standards for the 2007-2008 academic year ("School Division Report Card: [School Division][1]") which indicated teachers had found success in the classroom.

According to the National Center for Educational Statistics, in 2006-2007 this school division employed a total of 4547 teachers, 106 at the pre-kindergarten and kindergarten level, 1384 at the elementary level, and 3057 at the secondary level ("District Information: [School Division][1]"). The sample was sufficiently large to ensure adequate sample size of core content teachers at each school level for this study.

Sampling

A stratified, random sample of all K-12, core-content teachers were surveyed for this study. The sample was stratified by school level (elementary, middle, and high) and primary teaching responsibility; only elementary classroom teachers and middle and high school core-content teachers were included. A list of all elementary classroom, and middle and high school core-content teachers was obtained from the school division. Fifteen-hundred teachers were randomly selected, 500 from each school level: elementary, middle, and high school. Teachers were organized by school level and a random number generator was used to assign each teacher within each school level a unique number. The three lists were sorted numerically and the first 500 teachers on each list were selected for participation.

To determine adequate sample size in survey research, Sapsford (2007) recommends a

[1] School division unnamed to maintain confidentially

minimum of 40 cases, with 100 or more cases preferred, per level of the independent variables to be analyzed. This study had four independent variables: school level taught (e.g., elementary, middle, high), years teaching experience (e.g., 0-3 years, 4-10 years, 11+ years), degree attained (e.g., bachelor's, graduate), and primary teaching responsibility (e.g., elementary classroom teacher, English, mathematics, science, social studies, other). The levels of the independent variables totaled 14, necessitating a sample size of at least 1400. The sample for this study included 1500 teachers, exceeding Sapford's guidelines.

Another consideration with respect to survey research was response rate. Research on response rates for web-based versus mail surveys have produced mixed results (Cobanoglu, Warde & Moreo, 2001; Shannon & Bradshaw, 2002; Kaplowitz, Hadlock & Levine, 2004). Cobanoglu, Warde, and Moreo (2001) found the response rate for web-based surveys to be 44.2% while mail surveys had a response rate of 26.3%. However, Shannon and Bradshaw (2002) found a much lower response rate for web surveys, 33.3%, compared to mail surveys, 66.7% (Shannon & Bradshaw, 2002). Kaplowitz, Hadlock and Levine (2004) found the response rate for both mail and web-based surveys to be comparable, especially when similar pre-notice and reminder procedures were followed; however, it should be noted that even then, the response rate for their study was approximately 30.0% for web-based surveys and 31.5% for mailed surveys. Dillman (2000) reports a comparable response rate for mail and web-based surveys but cautions that coverage error is high if the sample does not have adequate and equal access to computers (Dillman, 2000). Since the target population for this survey was public school teachers who all had established email addresses with equal access to computers, it was felt a 35.0% to 40.0% response rate was a reasonable expectation.

Mitchell and Jolly (2007) suggest for populations of 1000, a sample size of 278 is

required for a 95% confidence level with a 5% sampling error (p. 235). For populations of 2000, a sample size of 322 is needed for same confidence level and sampling error estimate (Mitchell & Jolley, 2007, p.235). Consequently, for this survey design, a response by approximately 300 teachers was desired. By randomly selecting a stratified sample of 1500, and anticipating response rate of 35.0% to 40.0%, it was concluded the sample for this survey design would yield an adequate sample size for analyses at the acceptable confidence level and sampling error. The response rate for this study was 43.7% (n = 656). Fifty respondents completed less than half the items in Part II of the survey, lowering the rate of usable responses to 40.4%. Item-to-item response rates for Part II varied from 42.6% (n = 639) to 39.3% (n = 590). For Part III of the survey, the Classroom Assessment Literacy Inventory, only those respondents who completed all items were included for analysis, resulting in a response rate of 29.5% (n = 442). Even with this decrease in sample size, the number of responses was sufficient for analysis at a 95% confidence level.

Finally, though all teachers had the opportunity to use student assessment data to inform their instructional practice, this study focused only on those teachers responsible for core-academic instruction, elementary classroom teachers, and core-content teachers at the middle and high school level. The ever increasing need to monitor and constantly improve students' performance in core-academic areas, thus increasing the possibilities for core-academic teachers to use summative assessments in a formative way, was the rationale for this decision.

Participant Characteristics

The population for this study was 81.8% female and 18.2% male. The ethnicity distribution for the population was 9.3% black, not of Hispanic origin, 88.3% white, and 2.4% of

ethnic origin other than black or white (School Division Data)[2]. Table 1 summarizes the participant characteristics and provides a comparison to the population. Of those who responded, 80.6% were females, 19.4% were males; 31.0 % were elementary teachers, 32.5% were middle school teachers, and 36.8% were high school teachers. The ethnicity for the sample was 6.8% black, , not of Hispanic origin, 89.5% white, and 3.7% of ethnic origin other than black or white.

Table 1

Comparison of Frequency Distribution of Sample to Population

	Sample				Population			
Demographics	Total *n*(%)	Elementary *n*(%)	Middle *n*(%)	High *n*(%)	Total *N*(%)	Elementary *N*(%)	Middle *N*(%)	High *N*(%)
Gender								
Total	644[1] (100)	200 (31.0)	209 (32.5)	237 (36.8)	4005 (100)	1763 (44.0)	956 (23.9)	1286 (32.1)
Female	519 (80.6)	193 (29.9)	172 (26.7)	154 (23.9)	3274 (81.8)	1666 (41.6)	768 (19.2)	840 (21.0)
Male	125 (19.4)	7 (1.1)	37 (5.7)	81 (12.6)	730 (18.2)	97 (2.4)	188 (4.7)	445 (11.1)
Ethnicity								
Total	647[1] (100)	204 (31.5)	210 (32.5)	233 (36.0)	4005 (100)	1763 (44.0)	956 (23.9)	1286 (32.1)
Black, not of Hispanic origin	44 (6.8)	13 (2.0)	15 (2.3)	16 (2.5)	372 (9.3)	131 (3.3)	100 (2.5)	141 (3.5)
White	579 (89.5)	183 (28.3)	188 (29.1)	208 (32.1)	3536 (88.3)	1591 (39.7)	833 (20.8)	1112 (27.8)
Other	24 (3.7)	8 (1.2)	7 (1.1)	9 (1.4)	97 (2.4)	41 (1.0)	23 (.6)	33 (.8)

[1] n differs due to varying item response rates

In addition to school level, years teaching experience and years in current assignment were measured. These demographic data showed a slight inverse trend. Over half of the participants, 51.4%, reported 11+ years teaching experience while 35.7% reported four to ten

[2] School division unnamed to maintain confidentiality

year's experience, and 12.9% respondents were beginning teachers with zero to three year's experience (see Table 2 and Table 3). Conversely, only 20.5% of teachers reported they have been in their current assignment for 11+ years. The remaining participants were essentially evenly divided with 40.2% reporting they have been in their current teaching assignment for four to ten years and 39.4% reported being at their current assignment three years or less. Though this school division has an experienced teaching staff, it appears to be less common for teachers to remain in their current teaching assignment for more than 11+ years.

Table 2

Frequency Distribution of Sample by Year's Teaching Experience

	Elementary *n*(%)	Middle *n*(%)	High *n*(%)	Total *n*(%)
0-3 years	17 (2.6)	33 (5.1)	34 (5.2)	84 (12.8)
4-10 years	79 (12.1)	76 (11.7)	78 (12.0)	233 (35.7)
11+ years	107 (16.4)	104 (16.0)	124 (19.0)	335 (51.3)
Total Respondents	203 (31.1)	213 (32.7)	236 (36.2)	652 (100)

Table 3

Frequency Distribution of Sample by Year's in Current Teaching Assignment

	Elementary *n*(%)	Middle *n*(%)	High n(%)	Total *n*(%)
0-3 years	89 (13.6)	91 (13.9)	78 (11.9)	258 (39.4)
4-10 years	79 (12.1)	83 (12.7)	101 (15.4)	263 (40.2)
11+ years	37 (5.6)	40 (6.1)	57 (8.7)	134 (20.5)
Total Respondents	205 (31.3)	214 (32.7)	236 (36.0)	655 (100)

Table 4 summarizes the degrees attained by participants disaggregated by primary teaching responsibility. Data showed a slight majority of respondents, 354 (54.2%), had earned a bachelors degree, while 299 (45.8%) had earned a graduate degree. Less than 3%, 16 (2.4%) respondents reported a primary classroom teaching assignment other than an elementary classroom or core content class. The remaining respondents were divided between the elementary classroom and the core content areas for middle and high school.

Table 4

Frequency Distribution of Sample by Content Area

Content Area	Bachelors *n*(%)	Graduate *n*(%)	Total *n*(%)
English	60 (9.2)	79 (12.1)	139 (21.3)
Mathematics	82 (12.6)	49 (7.5)	131 (20.1)
Science	49 (7.5)	48 (11.5)	97 (14.9)
Social Studies	31 (4.8)	43 (6.9)	74 (11.3)
Elementary	124 (19.0)	71 (10.9)	195 (29.9)
Other	7 (1.1)	9 (1.4)	16 (2.5)
Total respondents	353 (54.1)	299 (45.9)	652 (100)

Instrumentation

A locally developed instrument was administered to measure the extent to which teachers used summative assessment data to inform their instructional practice. The locally developed survey instrument was based on research conducted in the areas of formative assessment and data-driven decision-making. Anderegg's (2007) study, *Classrooms and Schools Analyzing Student Data: A Study of Educational Practice* (Anderegg, 2007) was used to guide the instrument development. Additional guidance was found through McLeod's (2005) work on data driven decision making (McLeod, 2005). In conjunction with the locally developed instrument, the Teacher Assessment Literacy Questionnaire, developed by Plake and Impara (1993) was administered to measure teachers' knowledge of classroom assessment (see Appendix A for email permission to use Teacher Assessment Literacy Questionnaire).

Part I of the survey gathered teacher demographic information. Part II included a series of questions to measure frequency and types of assessment administered, the frequency and methods of data analysis, and changes teachers had made in their instructional practice as a result of data analysis. Part III of the survey included the Classroom Assessment Literacy Questionnaire. The entire survey can be found in Appendix B.

This study was designed to examine how teachers used specific summative assessments: teacher-generated tests, common departmental tests, benchmark testing, state-mandated end-of-course tests, and nationally norm-referenced tests, in a formative way. Black and Wiliam (1998a), in their meta-analysis of research on formative assessment, found that teachers' use of formative assessment significantly enhanced student achievement. Formative assessment is defined as the use of assessment data to help students and teachers evaluate what has been learned and what instructional practices should be implemented to enhance and strengthen

learning (Black & Wiliam, 2003). Formative assessment can take many forms; the term formative is more a function of the use of assessment data than the type of assessment itself (Black & Wiliam, 2003). To determine the extent to which teachers were using summative assessments in a formative way, survey questions were asked to measure the frequency teachers administered specific summative assessments and the frequency and methods of data analysis conducted with these summative assessments.

In addition to measuring the types and frequencies of summative assessments, the survey also measured the methods and frequency of data analysis teachers might use: mean, mode, median scores, and item analysis, as well as disaggregated data by Adequate Yearly Progress (AYP) subgroups and SOL standards. Central tendency analyses and item analysis were included as a method of data analysis most likely familiar to teachers, easily calculated, and so perhaps frequently used. Survey questions specific to disaggregating assessment data were developed for the potential analyses information these methods could yield for teachers. As Bernhardt's (2004) research suggests, the power of analysis comes from the intersection of two types of data and the resulting instructional changes that result from such analysis. For example, comparing mean scores on an assessment by AYP subgroups can inform a teacher of the efficacy of instruction not just for the class but for subgroups within the class. Due to the beneficial potential for this specific type of data analysis, and the heightened need to monitor the program of AYP subgroups, survey items solicited information to find out of the extent to which teachers employed disaggregated data as a type of data analysis method. Responses to these survey items, indicating frequency and methods of assessment data analysis, gave a descriptive picture of the extent that teachers were analyzing assessment data and thereby using summative assessment data in a formative way.

The survey also measured changes teachers made in their instructional practice as a result of assessment data analysis. As Black and Wiliam (2003) described in their research, teachers use of formative assessment is only useful when teachers use the information gained to implement changes in their instructional practice to enhance and strengthen learning. This survey provided a description of what specific instructional changes teachers reported as a result of their analysis of assessment data.

McLeod (2005) emphasizes five essential concepts necessary for effective DDDM. Of these five concepts, two were the focus for this study: the use of formative assessment and the knowledge to take instructional action. As mentioned, Part II of this survey was designed to evaluate the ways in which teachers used summative assessments in a formative way. Part III of the survey addressed the need for teachers' to have the requisite knowledge to take instructional action. The third part of the survey included an established instrument, the Classroom Assessment Literacy Inventory (CALI), designed to measure teachers' knowledge of educational assessment as specified in the Standards for Teacher Competence in Educational Assessment of Students (Plake, 1993). The first four Standards are related to the use of assessment data to make instructional decisions; only items from the CALI associated with thee standards were included in the instrument.

Prior to administering the survey, a pilot survey was conducted with selected teachers at each school level. Participants in the pilot survey were emailed a copy of the survey and asked to evaluate the overall length and clarity of the survey questions. Participants in the pilot study expressed concern with the length of the survey. As a result, questions were streamlined wherever possible, and the CALI was reduced just to measure knowledge of the first four standards. Pilot survey respondents reported difficulty with some questions within the CALI.

However, those questions were not altered as they were part of an established instrument.

The validity of the CALI had been established by the author when the instrument was originally developed. Survey items associated with this instrument were evaluated by educational professionals with established credentials in areas related to teacher assessment literacy (Plake, 1993). The locally developed instrument used for this research was informally evaluated by members of the pilot survey and university professors for construct validity.

Administration Procedure

This study used a web-based survey program to administer the survey instrument to teachers in a large, suburban school division. The web-based program, SurveyMonkey, was used to collect data for this research. This program has received positive reviews and affords the flexibility to customize a survey while also addressing many of the issues addressed in Dillman's tailored design method (Gordan, 2002). The survey URL link was provided to the participants in the invitational email. Design options were selected to enhance the ease of use for the participant.

Dillman's (2000) tailored design method was followed to maximize response rate and reduce survey error. An email was sent to each member of the sample inviting them to participate in the survey. Potential participants were clearly explained the purpose of the study and what was expected. An introductory email message explained the purpose of the study and the potential benefit for the respondent (see Appendix C). Participants were assured their responses would remain confidential and secure. This point was made in the written email invitation to all participates that the web-based survey program was set up to block all records of email and IP addresses of respondents. The survey items started with the simplest questions first, followed by items in a matrix design to minimize reading while enhancing participants

ability to mark items efficiently (Dillman, 2000).

In keeping with Dillman's tailored design method, selected teachers were sent an email message explaining the purpose of the study and a web-link to access the survey as pre-notification. Teachers had a two-week window of time in which to respond. At the one-week time interval, an email was sent to all participants reminding them to complete the survey if they have not already (see Appendix C for the invitation and reminder emails).

Variables

Independent variables. This study was designed to describe how teachers are using assessment data to inform their instructional practice. Consequently, two of the four research questions are descriptive in nature and do not include independent variables as part of the analyses. For example, the first research question asked to what extent do teachers use summative assessments in a formative way? A variety of data were collected to address this question. For example, frequency analyses were conducted to determine the most common types of summative assessments administered by teachers: teacher-generated, common-departmental, division benchmark, released SOL items, and nationally norm-referenced assessments. The types and frequency of the methods of data analysis were also considered and included: calculating mean, mode, and standard deviation, item analysis, disaggregating by AYP subgroups, and disaggregating by individual SOL standards.

For the second question, what is the level of teacher assessment literacy in a large suburban school division, the independent variables were teacher characteristics: school level, years of experience, degrees attained, and primary teaching responsibility. School level was divided into three levels: elementary, middle, and high. Years of teaching experience were organized by categories of 0-3 years, 4-10 years, and 11+ years. These groupings reflected

different stages in teachers' experience and would help provide insight if knowledge of assessment varied across teachers' careers. Degrees attained were divided into just two categories: bachelor's and graduate. By considering degrees attained, the possible relationship between teachers' advanced degrees and their increased knowledge of assessment data could be explored. Finally, for primary teaching responsibility the levels for this variable included elementary, English, mathematics, science or social studies. A classroom category titled "other" was included for those teachers for whatever reason did not fall within the five main categories.

The third research question provided a framework to determine the relationship between teachers' reported use of summative assessment in a formative way and their assessment literacy level. The formative use of summative assessments was operationalized to reflect the frequency with which teachers analyzed specific types of summative assessments as well as the frequency with which they used specific methods of data analysis. Frequency data for these two independent variables were divided into three levels: greater than four times a year, two to four times a year, and less than two times a year. Assessment literacy scores were compared across frequencies for each independent variable to determine if there was any relationship between those teachers who analyzed assessments more frequently and their level of assessment literacy.

The final question asked what changes in teachers' instructional practice result from their use of assessment data and provided for descriptive results The frequency with which teachers reported specific changes in instructional practice such as: changing pacing of future instruction, regrouping students, differentiating instruction, reteaching, or remediate and retest was examined.

Dependent variables. For those questions designed to elicit frequency data, the dependent variables were the frequencies that indicated how often within the course of the school

year teachers administered specific summative assessments or used specific methods of data analysis. This dependent variable was divided into six frequency levels: weekly, monthly, quarterly, each semester, annually, and never.

To explore any relationship between teachers' use of summative assessment in a formative way and teachers' assessment literacy, the dependent variable was the score for each standard measured by the Classroom Assessment Literacy Inventory. To address the last research question, a dichotomous response of yes or no was used as the dependent variable to measure the changes teachers' made in their instructional practice as a result of their analysis of summative assessment data.

Data Analysis

The primary research questions in this study asked to what extent are teachers using summative assessment data to inform their instructional decisions? Four research questions were asked to address this broad question.

Research Question 1: To what extent do teachers use summative assessments in a formative way? For the first research question a frequency distribution was generated to examine the extent to which teachers administered specific summative assessments. Additionally, frequency distributions were calculated to gain an understanding of how often and what methods of data analysis teachers used. These distributions gave a descriptive overview of the survey responses.

Research Question 2: What is the level of teacher assessment literacy in a large suburban school district? Responses for the Classroom Assessment Literacy Inventory were evaluated for the correct answer. Respondents' answers were recoded so a value of one indicated a correct answer and a value of zero an incorrect answer. A composite score for each

standard was calculated by the totaling the score for each standard. For example, if a teacher had answered all questions correctly, the composite score would be a five. The score decreased for each incorrect answer within the five questions for a given standard. Mean scores were calculated and compared for each level of four independent variables: school level, years' experience, degree attained, and primary teaching responsibility.

A multiple analysis of variance (MANOVA) was conducted, at a confidence level of 95%, to compare composite scores across the levels for each independent variable. A Bonferroni post hoc analysis was conducted if the overall results of the MANOVA indicated a significant difference was found. The one exception was the analysis comparing the composite scores for degree attained. With only two levels of this independent variable, a post hoc analysis was not necessary.

Research Question 3: What is the relationship between teachers' reported use of summative assessment in a formative way and their assessment literacy level? The use of summative assessment in a formative way was operationalized as teachers' responses indicating the methods and frequency of data analysis methods. Self-reported frequencies of data analysis methods were considered a measure of the use of summative assessments in a formative way. Additionally, any reported analysis of specific summative assessments was also considered a measure of the use of summative assessments in a formative way. A MANOVA was conducted to determine if there were any significant differences in assessment literacy scores and the frequency with which teachers reported analyzing summative assessment data. When significance was found, a Bonferroni *post hoc* analysis was performed to determine the exact nature of the differences and to examine if those teachers with a higher frequency of data analysis perhaps had a higher assessment literacy score.

Research Question 4: What changes in instructional practice result from teachers' use of assessment data? A frequency distribution was conducted run to examine the ways teachers reported their instructional practice had changed as a result of their analysis of assessment data. Teachers' open-ended item responses were evaluated for insight to changes teachers made to their instructional practice as a result of assessment data analysis. See Appendix D for a chart listing research questions and related data analysis methods.

Delimitations

This survey was administered within the same large, suburban school division. As public school teachers, all participants were familiar with the accountability measures mandated by state and federal legislation. Teachers had been exposed to the division's emphasis on improving student achievement through the use of benchmark testing so the use of student assessment data was an established expectation for this division. Although the use of only one school division limited the generalizability of the findings, having all respondents from the same school division provided a commonality that minimized errors due to different definitions and expectations of assessment policies or procedures.

Only elementary classroom teachers and core-content middle and high school teachers were selected for this study. Limiting the study to these teachers ensured all respondents had an opportunity to administer summative assessments throughout the academic year. In addition, these categories of teachers might be more likely to examine assessment data to evaluate student learning as their primary teaching responsibilities are more closely tied accountability measures.

Chapter 4

Results

Overview

The purpose of this study was to examine the extent to which teachers use summative assessment data to inform their instructional practice. To address this issue, four research questions were developed to examine teachers' assessment literacy, the types and frequency of summative assessment administered and methods of data analysis, and how these analyses influence teachers' instructional practice.

To answer the questions, a survey was conducted of teachers in a large, suburban school district in central Virginia using a web-based instrument. A locally developed instrument was used to measure the frequency and types of assessment administered by teachers, as well as frequency and methods of data analysis, and any changes in instructional practice respondents made as a result of data analysis. The locally developed instrument for this study was guided by Anderegg's (2007) study, *Classrooms and Schools Analyzing Student Data; A Study of Educational Practice* (Anderegg, 2007) and McLeod's (2005) work in data driven decision making. This study also included an established instrument, the Classroom Assessment Literacy Inventory (Plake, 1993), to evaluate the assessment literacy level of respondents.

To What Extent do Teachers Use Summative Assessments in a Formative Way?

Frequency of assessments. To answer the first research question, to what extent do teachers use summative assessments in a formative way, it was necessary to determine the

frequency with which the summative assessments targeted in this study were administered. The frequency of different types of summative assessments measured in this study included: teacher-generated assessments, common departmental assessments, division benchmark assessments, released statewide standards of learning assessments, and nationally norm-referenced assessments.

As seen in Table 5, results indicated teacher-generated assessments were the most frequently administered type of assessment with 84.5% teachers reporting at least a weekly administration. Teachers also reported administering common departmental assessments on a regular basis albeit not as frequent as their own, teacher-generated assessments. Approximately 25.0% of teachers administered common departmental assessments on a weekly or monthly basis; one-third reported giving a common departmental assessment on a quarterly basis.

The next most frequent assessment administration was division benchmark tests. Results showed 80.0% of teachers reported administering of this summative assessment on a quarterly basis. Benchmark tests are typically scheduled for quarterly administration to help teachers gauge students learning over a longer period of time. It stands to reason this type of assessment would be most frequently administered on a quarterly basis. Approximately two-thirds a (64.3%) of teachers reported never administering a nationally norm-referenced assessment.

Table 5

Frequency Distribution of Specific Types of Assessments Administered by Teachers

	Weekly *n* (%)	Monthly *n* (%)	Quarterly *n* (%)	Each Semester *n* (%)	Annually *n* (%)	Never *n* (%)	*n*[1]
Teacher-generated assessments	516 (84.5)	79 (12.9)	7 (1.1)	3 (0.5)	1 (0.2)	5 (0.8)	611
Departmental common assessments	143 (23.5)	171 (28.1)	197 (32.3)	28 (4.6)	12 (2.0)	58 (9.5)	609
Division Benchmark assessments	5 (0.8)	18 (13.2)	487 (80.0)	39 (6.4)	6 (1.0)	54 (8.9)	609
Released Statewide Standards of Learning (SOLs) assessment/items	80 (13.2)	80 (13.2)	111 (18.3)	91 (15.0)	151 (24.9)	94 (15.5)	607
Nationally norm-referenced assessments	13 (2.2)	26 (4.3)	54 (9.0)	42 (7.0)	79 (13.2)	386 (64.3)	600

[1] n differs due to varying item response rates

The frequency of summative assessments administered was compared across school level to see if any one type of assessment was more utilized at one level than another. For each type of summative assessment, the most frequent administration category was considered for this analysis, with the exception of common-departmental assessments. This type of summative assessment was reported with very similar frequencies for weekly, monthly, and quarterly administration. As a result all three of these frequency categories were used.

As seen in Figure 1, the frequency of teacher-generated assessments on a weekly basis was evenly distributed among school levels. On a weekly and monthly basis, more elementary

and middle school teachers reported utilizing common departmental assessments than high school teachers, while high school teachers reported a higher administration on a quarterly basis. Quarterly administration of Division Benchmark assessments were fairly evenly distributed across all three school levels. The use of SOL released items was slightly higher for middle and high school teachers than for elementary school teachers. Nationally norm-referenced assessment administration was not included in this graphic as only 13.2% of teachers reported administering a nationally norm-referenced assessment annually; an overwhelming majority of teachers (64.3%) reported never administering a nationally norm-referenced assessment. However, of the 13.2% of teachers that reported administering a nationally norm-referenced test annually, 43.0% were elementary, 29.0% were middle, and 27.8% were high school teachers.

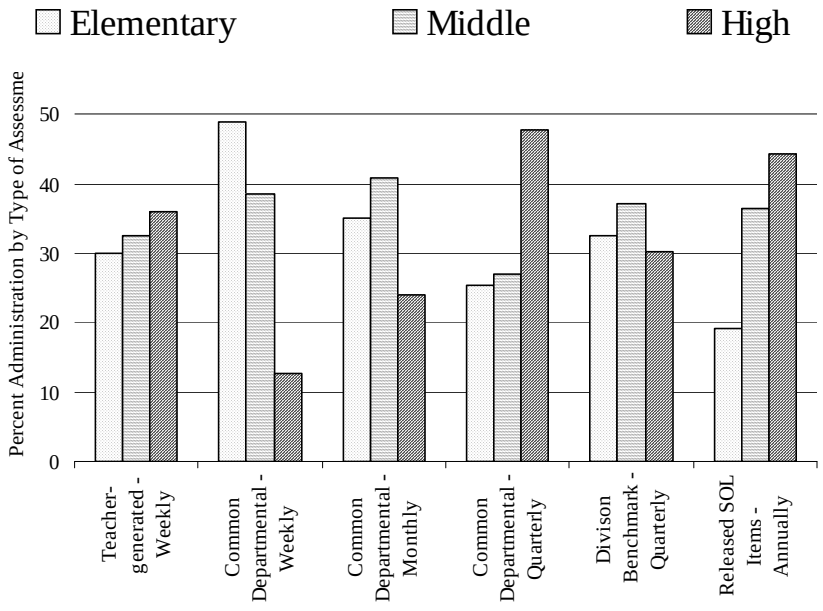


Figure 1. Comparison of most frequently administered assessments by school level for weekly teacher-generated assessment (n = 519), weekly common departmental assessments (n = 143), monthly common-departmental assessments (n = 171), quarterly common-departmental assessments (n=197), quarterly division benchmark assessments (n = 487), and annually released SOL items (n = 151).

Methods and frequency of data analysis. Teachers were surveyed to determine the methods and frequency of data analysis they performed on student assessment data. Methods of data analysis included central tendency calculations teachers might easily employ: calculating class mean, mode, or standard deviations. Additionally, teachers reported the frequency with which they disaggregated assessment data by Adequate Yearly Progress (AYP) subgroups and individual SOL standards. Finally, frequency data with respect to assessment item analysis was collected.

Table 6 shows the frequency distribution for methods of data analysis. Teachers most frequently reported analyzing student assessment data by examining the mean (36.7%), mode (34.1%), and standard deviation (41.7%) on a weekly basis; very few reported analysis of these types more frequently than on a quarterly basis. Teachers' reported frequency with regards to item analysis revealed the same trend. Of those teachers who reported that they analyzed individual items (30.4%), 22.3% reported doing so on a monthly basis, and 32.6% performed these types of analyses quarterly.

Almost one-third of teachers reported (32.8%) they examined student assessment data disaggregated by individual SOL standards quarterly during the school year. Additionally, 12.4% respondents reported disaggregating assessment data by SOL standards weekly and 16.4% reported doing so on a monthly basis. Conversely, almost one-third (29.2%) reported never disaggregating student assessment data by individual SOL standards. While 80.0% of teachers reported they administered a benchmark test (see Table 5) on a quarterly basis, less than one-third of the teachers reported disaggregating assessment data by individual SOL standards. Since benchmark testing is designed to help teachers gauge how students have mastered specific SOL standards, it is surprising that noticeably fewer reported disaggregating data by SOL

standards.

Disaggregation of assessment data by AYP subgroups showed a different distribution compared to the other aforementioned data analysis methods. Of those teachers responding, 29.2% reported disaggregating student assessment data by AYP subgroups on a quarterly basis. Ten percent or less reported doing the same on a more frequent basis. Approximately the same percentage of teachers reported disaggregating data by AYP subgroups at least once a semester (9.1%) or on an annual basis (13.8%). Just under one-third of respondents (30.6%) reported never analyzing students' assessment data by AYP subgroups at any time in the year. With the increased pressure to demonstrate academic achievement for AYP subgroups, it is surprising that almost one-third of the respondents never utilized this method of data analysis.

Table 6

Frequency Distribution of Methods of Data Analysis

	Weekly *n* (%)	Monthly *n* (%)	Quarterly *n* (%)	Each Semester *n* (%)	Annually *n* (%)	Never *n* (%)	*n*[1]
Examine mean	232 (36.7)	140 (22.1)	183 (28.9)	17 (2.7)	4 (0.4)	57 (9.0)	633
Examine Mode	216 (34.1)	161 (25.4)	153 (24.2)	14 (2.2)	3 (0.5)	86 (13.6)	633
Examine Standard Deviation	263 (41.7)	128 (20.3)	149 (23.6)	14 (2.2)	8 (1.3)	69 (10.9)	631
Disaggregate by AYP subgroups	44 (7.1)	64 (10.3)	182 (29.2)	57 (9.1)	86 (13.8)	191 (30.6)	624
Disaggregate by individual SOL standards	78 (12.4)	103 (16.4)	206 (32.8)	45 (7.2)	106 (16.9)	90 (14.3)	628
Item analysis	192 (30.4)	141 (22.3)	206 (32.6)	26 (4.1)	24 (3.8)	43 (6.8)	632

[1] n differs due to varying item response rates

Frequency of Data Analysis for Specific Assessments. To further examine the extent to which teachers use summative assessment in a formative way, survey items measured the frequency of the specific types of assessments analyzed (see Table 7). Teacher-generated assessments were the most frequently reported type of assessment analyzed with 77.3% of teachers reporting they analyze the results of their own classroom assessments on a weekly basis. A smaller percentage of respondents, roughly one-third, reported analyzing the results of departmental common assessments on a monthly or quarterly basis. Division Benchmark assessments were analyzed most frequently on a quarterly basis by 77.6% of the responding

teachers. Almost 65% teachers reported never analyzing nationally norm-referenced assessments.

Table 7

Frequency Distribution of Data Analysis for Specific Assessments

	Weekly *n* (%)	Monthly *n* (%)	Quarterly *n* (%)	Each Semester *n* (%)	Annually *n* (%)	Never *n* (%)	*n*[1]
Teacher-generated assessments	470 (77.3)	103 (17.4)	12 (2.0)	0 (0.0)	2 (0.3)	18 (3.0)	608
Departmental common assessments	120 (19.9)	187 (31.0)	188 (31.2)	30 (5.0)	11 (1.8)	67 (11.1)	603
Division Benchmark assessments	5 (0.8)	23 (3.8)	471 (77.6)	37 (6.1)	8 (1.3)	63 (10.4)	607
Released Statewide Standards of Learning (SOLs) assessment/items	64 (10.6)	78 (13.0)	109 (18.1)	91 (15.1)	162 (27.0)	97 (16.1)	601
Nationally norm-referenced assessments	18 (3.0)	24 (4.0)	47 (7.9)	41 (6.9)	83 (13.9)	383 (64.3)	596

[1] n differs due to varying response rates

A comparison of the frequency of administration of specific types of assessments with the frequency of analysis of these assessments was made to see if there was any parallel between the frequency teachers administer summative assessments and the frequency they analyze summative assessments. This comparison (see Table 8) showed similarities that indicated at

least the same percentage of teachers who reported administering a specific type of summative assessment is similar to the percentage of teachers who reported analyzing each type of summative assessment. This data indicates that teachers are analyzing summative assessment data; exactly how they are analyzing data is explored further in this study.

Table 8

Comparison of Percentage Administration of Specific Types of Assessment and Percentage of Analysis of Assessment Data

	Weekly		Monthly		Quarterly		Each Semester		Annually		Never	
	administered	analyzed	administered	analyzed	administered	analyzed	administered	analyzed	administered	analyzed	administered	analyzed
Teacher generated assessment	84.6	77.3	12.9	17.4	1.1	2.0	0.5	0.0	0.2	0.3	0.8	3.0
Departmental common assessment	23.5	19.9	28.1	31.0	32.3	31.2	4.6	5.0	2.0	1.8	9.5	11.1
Division Benchmark assessment	0.8	0.8	13.2	3.8	80.0	77.6	6.4	6.1	1.0	1.3	8.9	10.4
Released SOL assessment items	13.2	10.6	13.2	13.0	18.3	18.1	15.0	15.1	24.9	27.0	15.5	16.1
Nationally norm-referenced assessments	2.2	3.0	4.3	4.0	9.0	7.6	7.0	6.9	13.2	13.9	64.3	64.3

As Bernhardt (2000) noted, the power of data analysis can be found in the intersections of different variables such as student achievement and student subgroups. Based on this assertion, and federal requirements (No Child Left Behind Act, 2001), specific attention was given to data analysis method of disaggregation of data by AYP (Adequate Yearly Progress) subgroups (see Table 9). As noted in Table 6, when asked how frequently teachers disaggregated assessment data by AYP subgroups, 29.2% reported performing this analysis on a quarterly basis but 30.6% reported never employing this method of data analysis. An additional survey item measured the types of assessment teachers most frequently analyzed by disaggregating results according to AYP subgroups (see Table 9). The most frequent type of assessment was Benchmark tests, with 41.7% reporting using this method on a quarterly basis. Frequencies for all other types of assessments showed approximately half of all respondents rarely or never used this form of data analysis, the one exception being nationally norm-referenced assessments. Almost three-quarters of the respondents, 73.2%, reported never using this method of data analysis.

Table 9

Frequency Distribution of Data Analysis Method: Disaggregated by AYP Subgroups

	Weekly *n* (%)	Monthly *n* (%)	Quarterly *n* (%)	Each Semester *n* (%)	Annually *n* (%)	Never *n* (%)	*n*[1]
Teacher-generated assessments	106 (18.2)	57 (9.8)	61 (10.4)	25 (4.3)	25 (4.3)	310 (53.1)	584
Departmental common assessments	38 (6.5)	60 (10.3)	121 (20.8)	39 (6.7)	25 (4.3)	298 (51.3)	581
Division Benchmark assessments	7 (1.2)	17 (2.9)	243 (41.7)	48 (8.2)	27 (4.6)	241 (41.3)	583
Released Statewide Standards of Learning (SOLs) assessment/items	21 (3.6)	27 (4.7)	83 (14.3)	66 (11.4)	131 (22.6)	252 (43.4)	580
Nationally norm-referenced assessments	8 (1.4)	19 (3.3)	37 (6.4)	30 (5.2)	60 (10.5)	420 (73.2)	574

[1] n differs due to varying item response rates

What is the level of teacher assessment literacy in a large suburban school district?

Standards for assessment literacy for educators were developed through the joint efforts of the American Federation of Teachers, the National Council on Measurement in Education and the National Education Association ("Standards for Teacher Competence in Educational Assessment of Students," 1990). Seven standards were developed that describe a range of teacher skill sets associated with making instructional decisions and collaborative activities with other professionals. For this study, the first four of the seven assessment literacy standards were measured. These four standards are related to teachers' use of assessment data to make

instructional decisions:

- Standard 1: Choosing - Teachers should be skilled in choosing appropriate assessment methods to make instructional decisions.
- Standard 2: Developing -Teachers should be skilled in developing appropriate assessment methods for instructional decisions.
- Standard 3: Interpreting - Teachers should be skilled in administering, scoring, and interpreting assessments, both externally produced and teacher-generated assessments.
- Standard 4: Decisions - Teachers should be skilled in using assessment data to make decisions about student learning, to plan lessons, develop curriculum and for school improvement.

("Standards for Teacher Competence in Educational Assessment of Students," 1990).

Teachers' responses were evaluated for correctness, with a correct response being assigned a value of one and an incorrect response assigned a value of zero. This procedure mirrors the same procedure used by Plake, et al. (1993) in the national administration of the CALI in the early 1990s. A composite score for each standard was derived by tallying the total number of correct responses for the five questions. If a teacher responded correctly to all five questions for a given standard, the composite score for that standard would be a five. Four out of five correct responses would be a composite score of 4, and so on. Means approaching 5 indicated a greater knowledge for each specific standard. If a teacher missed all five questions, the composite score was zero. Due to the limited number of items comprising the standards, only those respondents who had complete data on each standard were analyzed for this study. Of the 656 respondents who started the survey, only 442 completed all questions associated with the

CALI. Table 10 shows the mean composite score for all respondents for the four standards measured. Standard composite scores ranged from 3.03 to 3.71.

Table 10

Mean Composite Scores for Assessment Literacy Standards

Standard	*n*	$\bar{x}$	*SD*
Choosing	442	3.24	1.018
Developing	442	3.03	.799
Interpreting	442	3.71	.959
Decisions	442	3.15	1.061

The mean standard scores ranged from3.03 to 3.71 suggesting that average levels of assessment literacy were fairly similar; however the standard deviation within each standard suggests greater variability within groups, especially for Standard 1 and Standard 4. This variability indicates that despite a mean score of 3.03 to 3.71, the variability reveals a wide range of knowledge within the sample. Whereas for Standard 2 and Standard 3 a clear majority of respondents scored near the mean, for the remaining standards, the distribution of scores was more evenly divided suggesting a wider knowledge gap than considering the means alone might indicate. See appendix F for a visual display of the score variability for each standard.

Composite scores were disaggregated according to each independent variable: school level, years' experience, degrees attained, and content taught. Means were compared for each level of each independent variable to determine if there was significant variation in teachers' knowledge about assessment literacy according to varying characteristics.

School level: elementary, middle, and high. The means for the first four standards measured by the Classroom Assessment Literacy Inventory showed a slight variation according

to school level taught for all four standards. Table 11 summarizes the mean scores for each standard by school level.

Table 11

Comparison of Classroom Assessment Literacy Mean Scores for School Level

Standard	Elementary			Middle			High		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Choosing	123	2.97	1.008	156	3.22	.994	163	3.47	1.002
Developing	123	3.12	.845	156	3.09	.754	163	2.91	.797
Interpreting	123	3.80	1.086	156	3.69	.833	163	3.66	.971
Decisions	123	3.16	1.082	156	3.08	1.069	163	3.21	1.041

A MANOVA was conducted to determine if there was a significant difference in mean assessment literacy scores according to school level. As shown in Table 12, the results of the MANOVA indicated a significant difference among school levels for Standard 1, choosing appropriate methods to make instructional decisions. A Bonferroni *post hoc* analysis (see Appendix F, Table 1) indicated the mean score for elementary teachers was statistically different from that of high school teachers. Elementary teachers' average composite score was 2.97 compared to 3.47 for high school teachers. Though these differences are statistically significant, this does not mean the scores are significantly different from a practical standpoint.

Table 12

MANOVA for School Level

Standard		*df*	*F*	*p*
Choosing	Between Groups	2	8.746	.000*
	Within Groups	439		
	Total	441		
Developing	Between Groups	2	2.747	.065
	Within Groups	439		
	Total	441		
Interpreting	Between Groups	2	.802	.449
	Within Groups	439		
	Total	441		
Decisions	Between Groups	2	.680	.507
	Within Groups	439		
	Total	441		

** $p < .05$*

Years of Experience. Mean composite scores for each assessment literacy standard were compared for the different categories of years of teaching experience. This independent variable was divided into three levels: 0-3 years' experience, 4-10 years experience, and 11+ years' experience. Table 13 summarizes the mean scores for each category of years teaching experience by assessment literacy standard. As shown, teachers with the most experience tended to have higher means in each of the four standards than their less-experienced counterparts.

Table 13

Comparison of Classroom Assessment Literacy Means for Years Teaching Experience

Standard	0-3 years			4-10 years			11+ years		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Choosing	48	3.17	.975	173	3.10	1.051	217	3.38	.989
Developing	48	2.92	.767	173	3.02	.821	217	3.05	.795
Interpreting	48	3.37	1.044	173	3.05	1.008	217	3.79	.888
Decisions	48	2.65	1.021	173	3.03	1.029	217	3.24	1.071

To determine if there was any significant difference among the levels of teacher experience, a MANOVA was conducted. The results of the MANOVA showed a significant difference according to teaching experience for Standards 1, 3, and 4 (see Table 14). A Bonferroni *post hoc* analysis revealed the specific difference between the three levels of this independent variable (see Appendix F, Table 2).

For Standard 1, choosing appropriate assessment methods to make instructional decisions, a significant difference was found between teachers with 4-10 years' experience and those with 11+ years' experience. Teachers with 11+ years experience had a mean composite score of 3.38, while teachers with 4-10 years' experience had a mean composite score of 3.10. The most experienced teachers (11+ years teaching experience) scored significantly higher than teachers with 0-3 years' teaching experience) as well as teachers with 4-10 years experience teachers on Standard 3 which focused on measuring teachers' skill in administering, scoring, and interpreting assessments. Standard 4 measured teachers' knowledge to use assessment data to

make decisions about students' learning, to plan lessons, to develop curriculum and for school improvement. On this standard, teachers with 11+ years of experience had a mean composite score of 3.24 compared to beginning teachers' score, 2.65, and teachers with 4-10 years experience scored, 3.03.

Table 14

MANOVA for Years Teaching Experience

Standard		*df*	*F*	*p*
Choosing	Between Groups	2	3.817	.023*
	Within Groups	435		
	Total	437		
Developing	Between Groups	2	.549	.578
	Within Groups	435		
	Total	437		
Interpreting	Between Groups	2	3.702	.025*
	Within Groups	435		
	Total	437		
Decisions	Between Groups	2	6.358	.002*
	Within Groups	435		
	Total	437		

** $p < .05$*

Degrees attained. Teachers' assessment literacy was also analyzed by degrees attained. There were two levels of this independent variable: bachelors and graduate degrees. Table 15 compares mean scores for each standard by degrees attained. For Standards 1, 3, and 4, the mean score for teachers with a bachelor's degree was lower than the mean score for those with a graduate degree. That trend was reversed for Standard 2. A MANOVA was conducted to

determine if any of the differences observed were statistically significant. Due to the two levels of this independent variable, no post hoc analysis was required.

With a 95% confidence level, the results of the MANOVA showed a significant difference between the scores for Standard 4 only; there was no significant difference between mean scores for Standard 1, 2, and 3 (see Table 16). For Standard 4, which targeted with teachers' knowledge to use assessment data to make instructional decisions, teachers with a bachelor's degree had a mean score of 3.02 while teachers with a graduate degree had a mean score of 3.28.

Table 15

Comparison of Classroom Assessment Literacy Means by Degrees Attained

	Bachelor			Graduate		
	n	*M*	*SD*	*N*	*M*	*SD*
Standard 1	223	3.23	.954	217	3.25	1.087
Standard 2	223	3.08	.755	217	2.97	.844
Standard 3	223	3.63	1.018	217	3.78	.896
Standard 4	223	3.02	1.042	217	3.28	1.071

Table 16

MANOVA for Degrees Attained

Standard		*df*	*F*	*p*
Choosing	Between Groups	1	.043	.835
	Within Groups	438		
	Total	439		
Developing	Between Groups	1	2.018	.156
	Within Groups	438		
	Total	439		
Interpreting	Between Groups	1	2.564	.110
	Within Groups	438		
	Total	439		
Decisions	Between Groups	1	6.596	.011*
	Within Groups	438		
	Total	439		

** $p < .05$*

Primary teaching responsibility. Respondents identified their primary teaching responsibility as either English, mathematics, science, social studies, elementary, or other. Tables 17 and 18 compare the mean scores for the four standards according to teaching responsibility. The results of the MANOVA showed a significant difference between teachers' assessment literacy knowledge for Standard 1, choosing appropriate assessment methods to make instructional decision (see Table 18).

The results of the Bonferroni *post hoc* analysis showed a significant difference for Standard 1. Elementary teachers (2.98) had a mean significantly lower than both mathematics (3.43) and science (3.48) teachers (see Appendix F, Table 3).

Table 17

Classroom Assessment Literacy Mean for English, Mathematics, and Science Teachers

Standard	English			Mathematics			Science		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Choosing	92	3.28	.918	92	3.45	1.020	71	3.48	.984
Developing	92	2.93	.708	92	3.10	.826	71	3.07	.704
Interpreting	92	3.72	.869	92	3.54	.954	71	3.72	.937
Decisions	92	2.98	1.109	92	3.20	1.102	71	3.23	1.017

Table 18

Classroom Assessment Literacy Mean for English, Mathematics, and Science Teachers

Standard	Social Studies			Elementary			Other		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Choosing	61	3.16	1.113	118	2.98	.996	8	2.75	1.018
Developing	61	2.84	.860	118	3.13	.853	8	2.88	.799
Interpreting	61	3.69	.867	118	3.81	1.088	8	3.75	.959
Decisions	61	3.20	.928	118	3.17	1.096	8	3.38	1.061

Table 19

MANOVA for Primary Teaching Responsibility

Standard		*df*	*F*	*p*
Choosing	Between Groups	5	3.610	.003*
	Within Groups	436		
	Total	441		
Developing	Between Groups	5	1.566	.168
	Within Groups	436		
	Total	441		
Interpreting	Between Groups	5	.814	.540
	Within Groups	436		
	Total	441		
Decisions	Between Groups	5	.688	.633
	Within Groups	436		
	Total	441		

** $p < .05$*

What is the relationship between teachers' reported use of summative assessment in a formative way and their assessment literacy level?

Teachers' use of summative assessment in a formative way was compared to their scores on the Classroom Assessment Literacy Inventory to determine if teachers' assessment literacy level had any relationship with the frequency with which they used of summative assessment data in a formative way. For this study, teachers' self-reported frequencies of data analysis methods and analysis of summative assessments were considered a measure of the use of summative assessments in a formative way. The assumption was made that the analysis of specific types of summative assessments, as well as the utilization of specific methods of data

analysis, resulted in teachers critically analyzing the results and making instructional decisions. It should be noted that the first set of teacher activities (analysis and use of methods of data analysis) does not necessarily mean the latter (formative use of data analysis) occurs.

Initially respondents were asked to report the frequency with which they analyzed data from different types of assessments: teacher-generated assessments, departmental common assessments, division benchmark assessments, released SOL assessment items, and nationally norm-referenced assessments. They were also asked to report the frequency with which the used specific methods of data analysis: examining means, examining mode, evaluating standard deviation, disaggregated by AYP subgroups, disaggregated by SOL standards, and item analysis.

For both survey items the response options were: weekly, monthly, quarterly, each semester, annually, and never. New categorical variables were constructed from these two survey items in order to create a variable with sufficient cells sizes. The number of teachers who reported either a weekly or monthly analysis of a specific assessment or method of data analysis were combined and recoded to reflect an analysis frequency of greater than four times a year. Those reported frequencies of analyses on a quarterly basis or each semester were combined to reflect an analysis frequency of two to four times a year. The remaining categories, annually or never, were combined to reflect analysis frequency of less than 2 times a year.

Teacher analysis of summative assessments. Teachers were asked to report the frequency with which they analyzed specific types of summative assessments. Only teacher-generated assessments, common departmental assessments, division benchmark assessments, and released SOL assessment items were examined because of the high frequency with which teachers reported analyzing data from these assessments. Excluded from the analysis were responses regarding the analysis of norm-referenced assessments. With 64.3% of respondents

reporting they never analyzed nationally norm-referenced tests (see Table 7), these data were not considered relevant as this analysis considered the frequency of analysis. Table 20 summarizes the mean assessment literacy scores of teachers, disaggregated by the frequency category for each type of summative assessment analyzed.

A series of four separate MANOVAs were conducted and showed no significant difference in the mean assessment literacy scores by the frequency with which teachers reported analyzing data from these assessments with the exception of Standard 4. Standard 4 evaluates teachers' knowledge to use assessment data to make instructional decisions about student learning, planning lessons, developing curriculum and school improvement. For this standard, a significant difference was found for those teachers who reported they analyzed released SOL assessment items. Table 21 shows the results of the MANOVA for teachers' reported frequency of analysis for released SOL assessment items.

A Bonferroni *post hoc* analysis (see Appendix F, Table 4) showed teachers who reported analyzing released SOL assessment items more than four times a year had a mean score of 3.28 while those that reported a frequency of 2 to 4 times a year had a mean score of 2.97. Teachers who reported analyzing released SOL assessment items less than two times a year (3.30) also had a mean score higher than those who reported performing the same analysis 2 to 4 times a year (2.97).

Table 20

Classroom Assessment Literacy Mean by Type of Assessment Analyzed

Standard	Analyze > 4 times per year			Analyze 2-4 times per year			Analyze less than 2 times per year		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Choosing									
Teacher-generated assessments	383	3.25	1.006	7	3.00	.813	13	3.00	1.291
Common departmental assessments	205	3.17	1.027	142	3.27	.990	56	3.36	1.017
Division benchmark assessments	18	3.00	.970	363	3.23	1.015	22	3.41	1.008
Released SOL assessment items	96	3.23	.978	147	3.20	1.031	160	3.27	1.020
Developing									
Teacher-generated assessments	383	3.02	.804	7	3.14	.690	13	3.00	.707
Common departmental assessments	205	3.02	.813	142	2.99	.794	56	3.36	.755
Division benchmark assessments	18	2.89	1.023	363	3.02	.795	22	3.14	.640
Released SOL assessment items	96	3.07	.798	147	2.95	.747	160	3.07	.840
Interpreting									
Teacher-generated assessments	383	3.68	.959	7	3.57	.976	13	3.85	.899
Common departmental assessments	205	3.69	.944	142	3.60	1.004	56	3.36	.862
Division benchmark assessments	18	3.50	1.249	363	3.68	.953	22	3.82	.733
Released SOL assessment items	96	3.60	.934	147	3.64	.986	160	3.77	.940
Decisions									
Teacher-generated assessments	383	2.17	1.057	7	2.71	.756	13	3.38	1.121
Common departmental assessments	205	2.17	1.094	142	3.21	.988	56	3.36	1.083
Division benchmark assessments	18	3.28	1.127	363	3.17	1.049	22	3.18	1.140
Released SOL assessment items	96	3.30	1.106	147	2.97	.925	160	3.28	1.111

Table 21

MANOVA for Released SOL Assessment Items Analyzed

Standards		*df*	*F*	*p*
Choosing				
	Between Groups	2	.448	.639
	Within Groups	431		
	Total	433		
Developing				
	Between Groups	2	1.907	.150
	Within Groups	431		
	Total	433		
Interpreting				
	Between Groups	2	1.628	.170
	Within Groups	431		
	Total	433		
Decisions				
	Between Groups	2	4.207	.015*
	Within Groups	431		
	Total	433		

*$p < .5$

Methods of data analysis. Teachers' reported frequency of using different methods of data analysis was examined to determine if approaches varied according to the level of assessment literacy. Teachers were asked how frequently they analyzed summative assessment data using the following methods of data analysis: examine the mean, the mode, the standard deviation, disaggregate assessment data by AYP subgroups, disaggregate assessment data by SOL standards, and conduct an item analysis. Table 22 shows a comparison of the mean assessment literacy standard scores for each method of data analysis.

A series of six separate MANOVAs were conducted and showed no significant difference between any of the methods of data analysis and teachers' assessment literacy scores with the exception of examining the mode for summative assessment data. For this data analysis method, a significant difference in assessment literacy scores was found for Standard 3: teachers should

be skilled in administering, scoring, and interpreting assessments. Table 23 summarizes the results of the MANOVA for this method of data analysis.

A Bonferroni *post hoc* analysis was performed and showed there was a significant difference in scores between those teachers who report using this method of data analysis more than four times per year and those that reported using this method less than 2 times per year (see Appendix F, Table 5). Teachers who examined mode data more than four times a year scored .343 points lower on standard three than those who reported using this method of data analysis less than twice a year.

Table 22

Classroom Assessment Literacy Mean by Method of Data Analysis

Standard	Analyze > 4 times per year			Analyze 2-4 times per year			Analyze less than 2 times per year		
	n	M	SD	n	M	SD	n	M	SD
Choosing									
Examine Mean	254	3.25	1.036	134	3.25	.994	44	3.07	1.021
Examine Mode	262	3.19	1.059	113	3.26	.989	57	3.26	.899
Calculate Standard Deviation	270	3.23	1.017	113	3.23	1.027	49	3.22	1.046
Disaggregate by AYP Subgroups	69	3.14	1.033	179	3.13	1.028	184	3.36	.999
Disaggregate by Released SOL items	126	3.10	.987	175	3.23	1.025	131	3.36	1.038
Item Analysis	236	3.17	1.101	151	3.30	1.033	45	3.29	1.036
Developing									
Examine Mean	254	3.05	.796	134	2.99	.822	44	2.98	.792
Examine Mode	262	3.02	.797	113	2.98	.790	57	3.11	.859
Calculate Standard Deviation	270	3.07	.801	113	2.97	.829	49	2.90	.743
Disaggregate by AYP Subgroups	69	2.94	.820	179	3.02	.831	184	3.06	.769
Disaggregate by Released SOL items	126	3.02	.754	175	3.03	.746	131	3.02	.894
Item Analysis	236	3.08	.707	151	3.01	.883	45	2.78	.951
Interpreting									
Examine Mean	254	3.70	.959	134	3.69	1.013	44	3.68	.857
Examine Mode	262	3.62	1.009	113	3.74	.952	57	3.96	.706
Calculate Standard Deviation	270	3.72	.942	113	3.64	1.027	49	3.73	.953
Disaggregate by AYP Subgroups	69	3.70	.880	179	3.63	1.043	184	3.77	.928
Disaggregate by Released SOL items	126	3.61	.938	175	3.70	1.008	131	3.79	.905
Item Analysis	236	3.68	.980	151	3.74	.962	45	3.67	.896
Decisions									
Examine Mean	254	3.17	1.063	134	3.03	1.062	44	3.41	.972
Examine Mode	262	3.07	1.084	113	3.28	.995	57	3.26	1.027
Calculate Standard Deviation	270	3.10	1.091	113	3.23	1.027	49	3.24	.925
Disaggregate by AYP Subgroups	69	3.07	1.075	179	3.15	1.052	184	3.18	1.059
Disaggregate by Released SOL items	126	3.11	1.126	175	3.13	1.026	131	3.22	1.032
Item Analysis	236	3.18	1.100	151	3.09	1.026	45	3.22	.927

Table 23

MANOVA for Methods of Data Analysis: Examining Mode

Standard		*df*	*F*	*p*
Choosing				
	Between Groups	2	.771	.463
	Within Groups	438		
	Total	440		
Developing				
	Between Groups	2	.388	.678
	Within Groups	438		
	Total	440		
Interpreting				
	Between Groups	2	3.188	.042*
	Within Groups	438		
	Total	440		
Decisions				
	Between Groups	2	25.0	.131
	Within Groups	438		
	Total	440		

* $p < .05$

What changes in instructional practice result from teachers' use of assessment data?

To answer this question, respondents were asked to what changes they made in their instructional practice based on student assessment data. Respondents were given a specific choice of possible changes: change pacing, regroup students, differentiate instruction to remediate and/or enhance learning, re-teach concepts, remediate and retest. Teachers also had a choice to note that constraints of pacing prevented re-teaching but consideration is given to instructional strategies for the next unit based on student performance on previous unit. Finally, teachers could provide a narrative response if they made any changes not listed. Teachers were asked to describe any changes they had made in their instructional practice not included in the survey item. This open response permitted teachers to elaborate on the ways they had changed their instructional practice as a result of assessment data analysis.

Frequency distributions showed that a vast majority of respondents reported making some change in their instructional practice based on student assessment data (see Table 24).

However, 64.3% reported that pacing prevented re-teaching concepts indicating that any changes resulting from evaluation of student assessment data was done for subsequent instruction rather than returning to the previous unit. This result seems inconsistent with 93.5% of respondents who reported they do re-teach topics/concepts based on student assessment data. However, these results may indicate that teachers do not consider topics/concepts as discrete pieces of information but rather part of a spiraling curriculum that allows teachers to revisit a particular topic/concept in the next unit to strengthen students' understanding.

Table 24
Frequency Distribution for Changes Made in Instructional Practice Based on Student Assessment Data

	Yes *n* (%)	No *n* (%)	*n*
Changing pacing of future instruction.	560 (91.7)	51 (8.3)	611
Regroup students within the classroom.	487 (80.4)	119 (19.6)	606
Differentiate instruction to remediate and/or enhance learning.	587 (96.4)	22 (3.6)	609
Re-teach topics/concepts	571 (93.5)	40 (6.5)	611
Remediate and re-test for specific unit	501 (82.7)	105 (17.3)	606
Pacing prevents re-teaching but consideration is given to instructional strategies for next unit.	373 (64.3)	207 (35.7)	580
I have changed my instructional practice in ways not listed here.	231 (42.9)	308 (57.1)	539

The open-ended item allowed teachers to list specific strategies and changes they may have made to their instructional practice not listed in the survey response options. Many teachers discussed ways they remediate students, either individually or in small groups. One

respondent wrote," If there are a few children who have not mastered a concept then I provide a tutor (volunteer) to work with those children until they have reached mastery." Many teachers also differentiated their instruction to accommodate the various learning styles of their students. As noted by one respondent:

> If students do not demonstrate understanding (based on assessment data), I use different teaching methods to re-teach material if time permits or to change the way all material is taught in future lessons. I will often give small "quizzes" to determine how my students learn material (visual, audial *[sic]*, kinesthenic *[sic]*, etc.,) to determine the best ways to introduce material, but, typically, I try to segment my classes into 20 minute segments that address all forms of learning so that, if a student does not understand the material one way, it is still presented in another way in the same class to boost student comprehension.

Respondents also indicated they used assessment data to make instructional decisions. One teacher wrote, "This year I have actually assessed the results of a particular SOL skill where the majority of my class had difficulty grasping and decided to reteach the skill because of it's *[sic]* importance." However, the frustration with frequent testing and added pressure to ensure all students achieve at higher and higher levels was also evident in the narrative responses. One respondent commented:

> I feel like I am testing and evaluating more than I am teaching and planning. I am only one person and often find it difficult to meet the needs of all - especially when outside factors (home life and support) play such a major role in a student's life when at the elementary level.

Summary

This study sought to determine the extent to which teachers use summative assessment data in a formative way. Frequency data showed teachers administer summative

assessments on a regular basis. Additionally, teachers' self-reported frequency of data analyses from summative assessment seems to parallel the administration of such assessments. However, when it came to identifying and measuring the frequency of specific methods of data analysis, it was rare to find more than one third of respondents utilizing any of the methods measured in this study. Only one third of the respondents reported examining central tendency data on a regular basis. Whereas approximately 33.0% teachers reported they disaggregated summative assessment data by AYP subgroups and/or SOL standards, another 30.0% reported never disaggregating summative assessment data by AYP subgroups.

Another aspect of effective use of data to make instructional decisions is knowledge of assessment. The Classroom Assessment Literacy Inventory was included in this study to evaluate the overall knowledge of assessment. Overall mean scores ranged from 3.03 to 3.71, on a 5 point scale. Though significant differences were found between the mean assessment literacy scores among various teacher characteristics (school level, years teaching experience, degrees attained, and primary teaching responsibility), the statistical difference did not necessarily suggest a practical one. Differences in mean scores for the assessment literacy inventory, whenever significant, were always less than half a point on a scale of 0 to 5. Standard deviations for mean scores for each standard did indicate a wide variability within each standard. Although teachers answered correctly, on average, three out of the five questions per standard, the high standard deviation indicates many teachers a knowledge gap with many teachers scoring high on a particular standard, but many teachers scoring low as well.

Scores for the individual assessment literacy standards were compared to the frequency of teachers' self-reported analysis of summative assessments. With a few exceptions, no significant difference was found between assessment literacy scores of those that were frequently analyzing summative assessment data and those who reported rarely

analyzing summative assessment data. As before, when any differences were found, they were so small as to question the practical difference even those there was a statistical difference.

Teachers overwhelmingly reported making changes to their instructional practice as a result of analyzing assessment data. Over 90.0% of respondents reported either changing the pacing of instruction, differentiating instruction, or re-teaching topics/concepts as a result of their analysis of assessment data. Narrative responses revealed the ways many teachers were implementing various changes to their instruction but also demonstrated teachers' frustration with the need to assess so frequently.

The results of this study suggest that teachers are considering summative assessment data when planning instruction. However, teachers reported analyzing summative assessment data with a higher frequency than they reported utilizing specific methods of data analysis. It is not clear if there are other types of analytical approaches, other than those mentioned in this study, that teachers are using or if teachers consider their own intuitive ability to gauge student learning through informal evaluation of assessment data a method of data analysis. The results also suggest that teachers employ a variety of instructional practices in response to summative assessment data.

Teachers' knowledge of assessment literacy did not have a strong influence on their assessment data analysis practices. However, in general there was some evidence that more experienced teachers do have a slightly higher assessment literacy score than their less experienced counterparts. Although not necessarily a strong statistical difference, the trend might indicate that seasoned teachers develop a keener awareness of what constitutes quality assessment and how best to use information from assessments to inform their instructional practice.

Chapter 5

Conclusion and Recommendations

Overview

The overarching question for this study asked to what extent are teachers using summative assessment data in a formative way? A non-experimental, descriptive study using a survey research design was conducted to address this question. Using elementary, middle, and high school teachers in a large, suburban school district in central Virginia, a web-based survey was conducted to determine the frequency with which teachers administered specific types of summative assessments, the frequency with which teachers analyzed student assessment data, the level of teachers' assessment literacy, and the changes teachers made in their instructional practice as a result of their analysis of summative assessment data. Four research questions guided this study:

1. To what extent do teachers use summative assessments in a formative way?
2. What is the level of teacher assessment literacy in a large suburban school district?
3. What is the relationship between teachers reported use of summative assessment in a formative way and their assessment literacy level?
4. What changes in instructional practice result from teachers' use of assessment data?

To address these questions, measures were developed to determine the frequency and types of summative assessment teachers administered, the frequency of specific methods of assessment data analysis, the frequency with which teachers analyzed summative

assessments, the types of changes teachers made in their instructional practice as a result of data analysis, their classroom assessment literacy level and the relationship between the use of summative assessments in a formative way and assessment literacy.

Significant Findings

Frequency and types of summative assessments administered. Not surprisingly, teachers administered teacher-generated tests the most frequently. Almost all of respondents reported administering this type of assessment on a weekly or monthly basis. Practically the same number of respondents reported administering departmental common assessment; however while some reported administering this type of assessment on a weekly basis, a majority administered departmental common assessments on a monthly to quarterly basis. In this division, all teachers were expected to participate in professional learning communities, and common departmental assessments were encouraged for all content areas. The high percentage of teachers reporting administering this type of assessment indicates teachers are working to implement this particular aspect of professional learning communities. Division benchmark tests were the next most frequently administered assessment, with most respondents using this assessment on a quarterly basis.

The use of benchmark tests appears to be widespread in this division with 80.0% of respondents administering this test on a quarterly basis and approximately 10.0% reported administering benchmark assessments at other points in the year. This frequency was high considering that even though all respondents were classroom or core-content teachers, not all respondents necessarily taught a course with a state-mandated, end-of-course test. Consequently, not all respondents would necessarily have been required to administer a benchmark assessment. Benchmark tests were only developed for grade-levels and courses associated with state-mandated, end-of-course assessments.

This division required all third through eighth grade mathematics teachers, as well as all Algebra I, Geometry, and Algebra II teachers, to administer a mathematics benchmark test each quarter. Additionally, if a school failed to meet AYP requirements for language arts, language arts teachers for grade levels and courses associated with state-mandated, end-of-course assessment were also required to administer a benchmark test. All other benchmark administrations were left to the discretion of the principal. That approximately 90.0% of teachers reported administering this type of summative assessment at some point in the academic year indicates benchmark assessments are used above what is minimally required by the division.

Frequency distributions for released SOL assessment items indicated teachers were utilizing this type of summative assessment throughout the year. As SOL items are released, teachers can incorporate these questions into their regular assessments. Additionally, many division benchmark assessments incorporate released SOL items. A quarter of the respondents reported administering released SOL assessment items annually, while approximately one-sixth of respondents reported they administered this type of assessment for each frequency category: weekly, monthly, quarterly, or each semester. Just over 15.0% of teachers reported they never administered released SOL items any time during the academic year.

Whereas it appears the use of released SOL items was not as frequently used by teachers on a quarterly basis, the more equally distributed administration of this type of assessment across the academic year indicates teachers find these assessment items valuable and useful in some way and use these items throughout the year rather than reserve them for a less frequent assessment. For those who reported never administering this type of assessment, it is possible for a core content teacher to not necessarily teach a course with a state-mandated, end-of-course, assessment. For example, physics, calculus, and trigonometry

teachers would all be considered core-content teachers but these courses are not associated with an end-of-course assessment, thus perhaps accounting for the approximately 15.0% who reported never using this type of assessment. Nationally norm-referenced assessments were not administered frequently at all. Over two thirds of the respondents reported never administering this type of assessment.

Frequency and methods of data analysis. A comparison of the frequency for the types of assessments administered and teachers' self-reported frequency of data analysis for specific types of assessments showed similar results. When respondents were asked how frequently they analyzed data from specific types of summative assessments, practically the same percentage reported analyzing as reported administering specific types of summative assessments.

The most frequent data analysis method was calculating central tendency statistics. An overwhelming majority, 80.0% to 90.0%, used this type of data analysis on a weekly to quarterly basis. Teachers in this division use an electronic gradebook program that can be set up to calculate central tendency data for individual assessments or overall student averages. The ease with which teachers can obtain this data may account for the high frequency and high percentage of teachers using these methods of data analysis.

The same percentages were noted for item analysis of assessment data. Interestingly, this method of data analysis is not available through the electronic gradebook program used by teachers so in all likelihood this method of data analysis would have had to be manually evaluated; however, for those teachers who administer an online version of the division benchmark assessment, reports can be generated that provide teachers with an item analysis for the assessment. However, the benchmark analyses would only be done on a quarterly basis and over half the respondents reported performing this method of data analysis on a weekly or monthly basis. The high percentage of teachers conducting an item analysis on a

weekly to quarterly basis implies teachers are engaging in some critical analysis of assessments to document students' performance.

The same pattern does not hold for disaggregating data by SOL standards. Teachers used this method of data analysis most frequently on a quarterly basis; however even then only about one third reported using this method. On average, about 15.0% of respondents note using this method on a weekly or monthly basis. It is notable that the percentage of teachers reporting the administration of released SOL items on a weekly and monthly basis compares very closely to those that reported they disaggregated assessment data by individual SOL standards with the same frequency. These results suggest that teachers are taking advantage of the released SOL items and disaggregating by SOL standards, ostensibly to determine how their students performed for a given standard.

Teachers reported most frequently disaggregating assessment data by SOL standard on a quarterly basis. In this division, the benchmark assessment items are disaggregated by SOL standards for those teachers who administer the test online. For this survey item, almost one-third reported disaggregating data by SOL standard, however this falls very short of the 80% who reported administering a division benchmark assessment which could indicate that though many teachers are administering the division's benchmark test, far fewer are analyzing these assessments to determine students' mastery of specific SOL standards. These teachers may not be administering the online version of the division benchmark, or if they are, choosing to ignore any disaggregate data by SOL standards.

As for disaggregating data by AYP subgroups, the results showed an overall decline with approximately 10% of teachers reporting utilizing this method of data analysis on a weekly, monthly, each semester or annually. However, the percentage jumped to 29.2% when considering respondents who reported performing this method of data analysis on a quarterly basis. It should be noted that mathematics courses have the greatest impact on AYP

status and this division has required mathematics' benchmark assessment data be reported by AYP subgroups. Approximately twenty percent of respondents were mathematics teachers and another 20.0% were English teachers. These teachers alone might account for the 29.2% reporting disaggregating by AYP subgroups. For all other teachers, the results of assessment data, disaggregated by AYP subgroups, might not have much significance for these teachers and not as frequently used, if used at all.

Although all methods of data analysis were used by most teachers at some point during the school year, many respondents reported never using any specific method of data analysis. Disaggregating data by AYP subgroups was the most underutilized. When asked about the frequency with which they analyzed summative assessments by disaggregating data by AYP subgroups, on average 45.0% to 50.0% of respondents reported never using this method of data analysis. Of course, as noted prior, not every teacher feels the pressure to disaggregate by AYP subgroups as it is mathematics and language arts courses that considered for a school's AYP status.

The results of this survey indicate that teachers do employ methods of data analyses. Teachers appear to have access to efficient retrieval of central tendency data and conduct item analysis on a regular basis; however, fewer teachers disaggregate data by SOL standards and even fewer by AYP subgroups. The decreased use of more advanced methods of data analysis, such as the examination of disaggregated data, could indicate that teachers do not feel the need to more critically analyze assessment data unless obligated by some external means.

Changes in instructional practice as a result of analysis of assessment data. Practically all teachers reported some change in their instructional practice as a result of their analysis of assessment data. Most indicated they based decisions about future instruction on results of past instruction. In other words, they changed the pace of instruction or regrouped

students to better suit their instructional goals and differentiated instruction to enhance student learning. Although 80.0% to 90.0% reported they re-teach concepts and/or remediate and retest for specific units, approximately 65.0% of respondents reported pacing prevents re-teaching concepts. The frequency data indicated a discrepancy between these two survey items; however, teachers' narrative responses indicated the scope and sequence of curriculum allowed them to reteach and remediate concepts in subsequent units rather than go back and reteach previous concepts.

Teachers' classroom assessment literacy. The Classroom Assessment Literacy Inventory (Plake, 1993) was used to measure teachers' assessment literacy levels. This inventory is based on the Standards for Teacher Competence in Educational Assessments of Students ("Standard for Teacher Competence in Educational Assessment of Students," 1990). Only the first four standards were measured for this study as these standards were most related to the use of assessment data to make decisions.

Composite scores were calculated for each standard measured. There were five questions per standard; composite scores were a reflection of the number of correct responses for each standard. A score of five indicated a teacher got all five questions for a particular standard correct, a score of zero indicated the teacher missed all five questions. Mean scores for each standard ranged from 3.03 to 3.71 and were slightly lower than those reported by Plake, et al., (1993). Table 25 shows the comparison of Plake's data from the 1993 administration of the Classroom Assessment Literacy Inventory.

The research suggests that teachers need a foundation in assessment literacy to effectively use assessment data to enhance student achievement through DDDM (Black, et al., 1998a; Sharkey & Murnane, 2003; Lachat & Smith, 2005; McLeod, 2005; Stiggins, 2008). Federal and state accountability mandates have created an emphasis on documenting student learning though the use of summative assessments. However, today's accountability

movement has not been accompanied by an increase in knowledge of assessment. The results of the Classroom Assessment Literacy Inventory suggest teachers' knowledge of classroom assessment has not changed substantially since the early 1990s, in fact a slight decline in assessment literacy scores was found. The variability within each standard has not narrowed over the past two decades as well. Teachers in this study showed a wide range of knowledge for each standard, with scores varying across the scale. For Standard 1 and Standard 4, almost as many teachers scored below the mean as scored above the mean indicating a wider knowledge gap for these two assessment standards than for Standard 2 and Standard 3.

Table 25

Comparison of Classroom Assessment Literacy Inventory Scores

Standard	Current Study			Plake (p.12, 1993)		
	n	*M*	*SD*	*n*	*M*	*SD*
Choosing	442	3.24	1.02	555	3.46	.93
Developing	442	3.03	.80	555	3.22	.80
Interpreting	442	3.71	.96	555	3.96	.90
Decisions	442	3.15	1.06	555	3.40	1.11

Mean assessment literacy scores were compared, at a 95% confidence level, across teacher characteristics to determine if any particular characteristic was associated with higher or lower assessment literacy levels. When comparing school levels, a significant difference was found between elementary teachers and middle and high school teachers. Middle and high school teachers scored significantly higher than elementary teachers on Standard 1, choosing appropriate assessment methods for instructional decisions. Years of experience were also examined and teachers with more experience scored significantly higher on Standards 1, Standard 3 (interpreting), and Standard 4 (decisions) than did beginning teachers.

Results according to degrees attained also showed a significant difference in assessment literacy scores. Teachers with a graduate degree scored significantly higher on Standard 1 than those with a bachelor's degree. Finally, mean scores were compared among teachers by content area taught. A significant difference was found between science and mathematics teachers compared to elementary school teachers on Standard 1; science and mathematics teachers scored higher than elementary teachers.

Initial interpretation might suggest elementary teachers have some room to improve in terms of classroom assessment literacy; however, it should be noted that even when a significant difference was found, the mean scores rarely differed more than 10% calling into questions the practical significance despite finding a statistical significance. The difference in scores could be attributed to the formative nature of elementary school instruction. Elementary teachers may rely more heavily on day in and day out formative assessments, naturally adjusting and reteaching concepts as needed. This might negate the need for a more formal analysis of assessment data, giving elementary teachers less experience with analysis of assessment data than secondary teachers who more typically would need to document mastery learning of their content. Another explanation could be attributed to the content of mathematics and science teachers are more aligned with the classroom assessment inventory and these teachers are more accustom to working with data.

However, it should be noted that over the past 20 years, despite the advent of accountability models that rely increasingly summative assessments for accountability purposes, assessment literacy scores have not increased and in fact might show a downward trend when considered in the context of this study. With the increased reliance on assessments data to measure and document student learning, it is surprising teachers' knowledge in assessment has not increased when comparing the results to Plake's (1993) national administration of the same inventory in the early 1990s.

The relationship between the use of summative assessments in a formative way and assessment literacy. For this study, frequency data for the types of summative assessments analyzed, as well as the methods of data analyses were considered a measure of teachers' formative use of summative assessments. Assessment literacy scores were compared across frequency data to determine if there was a relationship between assessment literacy and the formative use of summative assessments. For these analyses, the assumption was made that the frequency of analysis of specific types of summative assessments, as well as the utilization of specific methods of data analysis, were necessary for the formative use of summative assessment data. However, it should be noted that the indirect measure of the formative use of summative assessment does not mean the first set of teacher activities (analysis and use of methods of data analysis) automatically guarantees the latter (formative use of data analysis) occurs. This analysis was predicated on the assumption that having the data would naturally prompt the teacher to critically evaluate it and respond accordingly.

The data showed that no significant difference in mean assessment literacy scores regardless of the frequency with which teachers reported analyzing summative assessments. When considering the frequency with which teachers used specific methods of data analysis, the results were contradictory. Only the frequency for using the mode method of data analysis showed any significant difference for assessment literacy scores. In this instance, those teachers who reported using the mode method of data analysis less than twice a year had a statistically significant higher mean than those teachers who reported using the method of data analysis more than four times a year. Although a statistically significant difference was found, the practical implications are limited. Of all the central tendency analyses teachers could use, examining the mode would seem the least informative when evaluating student learning. Perhaps this explains the lower score on the assessment literacy standards for those who seemed to rely on this method of data analysis more frequently.

Discussion

To what extent do teachers use summative assessments in a formative way?

McLeod (2005) defines data driven decision making (DDDM) as a "system of teaching and management practice that gets better information about students into the hands of teachers." (McLeod, 2005, p.1). Though McLeod notes there are multiple essential components to DDDM, one of these is teachers' use of formative assessment which was a focus of the present study.

Black and Wiliam defined formative assessments as activities that provide feedback to teachers that allow them to adjust their instruction to meet the needs of their students (Black & Wiliam, 1998a). Though formative assessments can take many forms, it is not the assessment itself that determines its formative nature but rather it is the way assessment results are used that makes it formative (Black & Wiliam, 2003).

This study explored teachers' use of specific summative assessments and how they used theses assessments in a formative way. Stiggins (2004) characterized summative assessments as assessments of learning, a means of documenting student mastery at the end of instruction. He described formative assessment as assessment of learning, a means of evaluating how students are learning (Stiggins, 2004). However, as Black and Wiliam (2003) suggested, summative assessments can be formative in nature if the data are analyzed in ways that provide teachers with information that can be used to change their instructional practice to enhance student learning.

The results from this study showed teachers are indeed administering summative assessments on a regular basis. Teacher generated assessments, along with common departmental assessments and division benchmark assessments were all administered on a weekly to quarterly basis. The use of released SOL assessment items was also utilized although less frequently. The amount of summative assessments administered showed

sufficient frequency to provide ample data for analysis. Teachers' reported frequency of data analysis for specific summative assessments was used as a measure of the formative use of these summative assessments. Survey results suggest that analyses of specific summative assessment data paralleled teachers' reported administration of the same assessments. The extent to which teachers reported analyzing data from summative assessments served as an indicator of the formative use of assessment data. It would appear that teachers are administering assessments frequently and making attempts to use summative assessments in a formative way.

This study suggested teachers employed some form of summative assessment in a formative way, teachers' use of one particular method of data analysis was examined in greater detail: according to disaggregating student assessment data by AYP subgroups. As McLeod (2005) suggests, teachers must use data in a formative way to reap the benefits of DDDM. Bernhardt (2000, 2004) notes the power of data analysis comes from the intersection of data sets that can reveal the effectiveness of instructional strategies and student learning (Bernhardt, 2000, 2004). For example, examining central tendency scores can provide an overall picture of student achievement on summative assessments. However, disaggregating the data by AYP subgroups can enhance the information to show the efficacy of instructional strategies for more subsets of students.

The results of this study demonstrated teachers' overwhelming use of central tendency data as a means of data analysis. However, easy calculation and access to central tendency data does not mean the significance of the data is well understood or considered when evaluating student learning. Additionally, teachers reported analyzing individual assessment items on a regular basis. However, less frequently used was the data analysis method of disaggregating data. According to Bernhardt (2000) disaggregating data to examine the intersections of various data sets can be very informative. For every summative

assessment listed, practically half the respondents reported never analyzing summative assessment data by AYP subgroups. While not all teachers bear the pressure brought on by federal accountability measures, it is surprising that in the heightened awareness of accountability and the need to document learning for all students, more teachers do not employ the one method that, according to Bernhardt (2000, 2004) could have an the greatest impact on impact on their schools ability to document learning for all – disaggregation by relevant subgroups and standards. This appears to be an opportunity lost to evaluate effective instruction across all content areas for those most critical subgroups as designated by federal standards.

What is the level of teacher assessment literacy in a large suburban school district? As federal and state mandates increasingly raise the bar for student achievement, teachers can no longer rely on the old models of assessing student learning. Assessments must be pre-planned, assessment data analyzed, and instructional action taken throughout the learning cycle to enhance student learning. In order to take the full advantage of the information assessment data analysis has to offer, teachers must be knowledgeable in assessment practices (Stiggins, 2008).

Attempts to implement plans to raise student achievement through the use of increased summative assessments have been misguided when teachers' knowledge of assessment was ignored (Vogel, Rau, Baker & Ashby, 2006; Murnane, Sharkey, & Boudett, 2005). Studies also found that when teachers are more knowledgeable of appropriate and effective assessment practices, they have more confidence in their efficacy as an instructor and students develop a more positive attitude towards learning (Lukin, Bandalos, Eckhout & Mickelson, 2004).

For this study, teachers' knowledge of classroom assessments was measured using the Classroom Assessment Literacy Inventory, first developed and administered in 1993 (Plake,

Impara & Fager, 1993). Plake's survey was based on the Standards for Teacher Competence in Educational Assessments of Students ("Standard for Teacher Competence in Educational Assessment of Students," 1990). Only the first four of the seven standards were measured for this study as these standards were related to the use of assessment data to make decisions. In their original study, Plake, et al., (1993) found assessment literacy scores for the first four standards ranged from 3.22 to 3.96. When ranked, Plake found the highest score for Standard 3, associated with administering, scoring and interpreting assessments. The scores were followed, in order, by Standard 1 (choosing appropriate assessment methods to make instructional decisions), then Standard 4 (using assessment to make decision about student learning), and then Standard 2 (developing appropriate assessment methods for instructional decisions. This study found teachers' assessment literacy scores overall to be approximately two-tenths lower than those found by Plake, but interestingly the variability within standards followed the same order. Respondents scored the highest (3.71) for Standard 3 (Interpreting), followed by Standard 1 (Choosing, 3.25), Standard 4 (Decision, 3.15) and then Standard 2 (Developing, 3.03). The variability within each standard does suggest there is somewhat of a knowledge gap.

When compared to Plake, et al. study, teachers' overall knowledge seems to have remained static in the ensuing years since the development of the standards for teacher competence in educational assessment of students. Teachers still show the same strengths and variability of knowledge indicating there is still room for teacher growth even twenty years after the establishment of these standards. Table 25 shows the comparison of the results from the 1993 study (Plake, Impara & Fager, 1993, p.12) and this study.

Assessment literacy data were also compared across various teacher characteristics to determine if any particular group of teachers possessed a higher knowledge level than another. In general, secondary school teachers scored higher than elementary school teachers

on the standards. Not surprisingly, teachers with more experience had a higher knowledge of particular standards than those with less experience. The same trend was evident between teachers with graduate degrees than those with bachelor's degree. And in terms of primary teaching responsibility, science and mathematics teachers scored higher for a particular standard than elementary school teachers. However, all differences were limited to a specific standard or two and no one group emerged as considerably more knowledgeable than another. The standard deviation for each measure does indicate that although a strong significant difference was not necessarily found between teacher characteristics, there is difference in assessment knowledge among the sample indicating there is still room for growth for teachers in assessment literacy.

What is the relationship between teachers' reported use of summative assessments in a formative way and their assessment literacy level? The third research question investigated the potential relationship between the use of summative assessment data in a formative way and teachers' assessment literacy level. As McLeod (2005) noted, effective DDDM requires teacher to not only use formative assessment in the classroom but also posses the knowledge to develop instructional interventions to improve student learning (McLeod, 2005). In his 1960s study, Bloom (1968) found teachers who used formative assessment data in a knowledgeable way had great gains in student achievement (Bloom, 1968). Mandinach, Honey, and Light (2006) developed a theoretical framework for DDDM. This framework illustrates the cyclical nature of DDDM starts with the collection of assessment data and ends with teachers taking instructional action to improve student learning. At that point, the cycle starts over with new data, evaluation of the data and again, instructional action to improve student achievement (Mandinach, Honey, and Light, 2006).

The question then becomes: do those teachers who report the most frequent formative use of summative assessment data have a greater knowledge of classroom assessment

literacy? The results of this study revealed no significant difference among assessment literacy scores for those teachers who reported more frequent formative use of summative assessments and those who rarely or never reported using summative assessments in a formative way. The limitations with respect to the measure of the formative use of summative assessment warrant caution when considering these results.

What changes in instructional practice result from teachers' use of assessment data? Effective DDDM for teachers will not exist unless instructional practice is evaluated. Mandinach, Honey, and Light (2006) identify instructional action as part of the DDDM process but do not necessarily define what that action might be (Mandinach, Honey, & Light, 2006). Nelson and Eddy (2008) found the power of teachers' actions as a result of use of assessment data to make instructional decisions. Studies by Bloom (1968) and Guskey (1990) both demonstrated that teacher action, as a result of the analysis of assessment data, had a profound effect on student achievement (Bloom, 1968; Guskey, 1990). Furthermore, increased student achievement had a secondary effect, student engagement increased and behavior problems decreased (Guskey, 1990). The literature suggests that when teachers act on student assessment data, their instruction can be more effective.

This study measured specific ways teachers changed their instructional practice as a result of their analysis of student assessment data. Primarily teachers reported some form of differentiating instruction, changed the pace of instruction, or regrouped or remediated students as needed. Narrative responses found the same pattern with teachers reporting the various ways in which they had differentiated instruction or changed pace of instruction. An overwhelming number of teachers reported some form of remediation for individual students or groups of students. Similar to the findings of Bloom (1968) and Guskey (1990), teachers in this study appear to be putting assessment data to use through the various instructional interventions such as remediation and differentiation. With 80% to 96% of teachers reporting

some change in instructional practice as a result of analysis of student assessment data, Mandinach's, *et al* (2006) theoretical framework seems to be at work in this division. Teachers are analyzing summative assessments, evaluating and changing their instructional practice and starting the process over by once again analyzing summative assessments and going through the process again.

Recommendations

Implications for practice. There is no doubt that teachers are developing and administering summative assessments on a regular basis. There is also evidence that teachers are using central tendency data such as score means, mode, and standard deviation, to evaluate student progress according to this research. Additionally, teachers are making an effort to evaluate their assessments by conducting an item analysis on a regular basis. However, the more effective means of data analysis as indicated by Bernhardt (2000, 2004), disaggregating data by specific variables, is left untouched by practically half the teachers who responded.

This raises questions of why? It would behoove school divisions to explore the reasons teachers shy away from disaggregating assessment data and work to remove any obstacles to make this method of data analysis more easily accessible for teachers. Examining the technology infrastructure is key to promoting teachers formative use of summative assessment data. Whereas teachers may be very comfortable estimating the mean for a particular assessment, and informally evaluate the mode and standard deviation, more advanced technology is necessary to efficiently disaggregate data.

In addition to sufficient technological support, a collaborative culture is necessary to support advanced data analysis methods such as disaggregating assessment data (McLeod, 2005; Wayman, 2005; Wayman, Midgley & Stringfield, 2005; Wayman, Cho & Johnson, 2007). Building leaders would do well to look for ways to foster increased collaboration.

Disaggregated data, or any assessment data for that matter, is better examined in a collaborative culture. It is through the collaboration with colleagues teachers exchange instructional strategies and interventions to help raise student achievement. The benefit of data analysis is weakened if instructional action is not taken; collaboration gives teachers the support to take action.

The findings of this study suggest classroom assessment literacy has not improved since the early 1990s when assessment literacy emerged as an educational concern ("Standards for Teacher Competence in Educational Assessment of Students," 1990). The increase in assessment responsibility has not necessarily been accompanied by an increase in assessment literacy for teachers. As Vogel, et al. (2006) found in Illinois, reform efforts to raise student achievement through the increase of state mandated assessments did not gain any ground until educational leaders addressed the issue of teachers' knowledge to understand and interpret more advanced data sets (Vogel, Rau, Baker, & Ashby, 2006).

School division personnel would be prudent to give careful consideration to the level of assessment literacy of their teachers. Staff development to enhance teachers' understanding of the benefit of the various methods of data analysis might fall short if not also accompanied by training to enhance teacher assessment literacy. Nebraska went so far as to offer an endorsement in classroom assessment (Lukin, Bandalos, Eckhout, & Mickelson, 2004). Murnane, Sharkey, and Boudett (2005) found three approaches commonly used when evaluating data. The instrument approach and the symbolic approach are used more for programmatic decisions whereas the conceptual approach is related to the use of assessment data to make instructional decisions. Of these three, the conceptual approach was the most underutilized (Murnane, Sharkey, & Boudett, 2005). School divisions should consider building programs and staff development that fosters and supports the use of assessment data to make instructional decisions. As the literature suggests, the most effective

use of assessment data demands a strong foundation in assessment literacy and thus underscores the need to promote an increased assessment literacy level among teachers (Sharkey & Murnane, 2003; Murnane, Sharkey & Boudett, 2005).

For further study. It is not clear from this study if the issue is technology support or teacher knowledge, or a combination of both, that promotes more advanced methods of data analysis, such as disaggregating data by content standards or student subgroups. This study does confirm that data analysis is occurring in today's classrooms; however, it does not address the question of why some methods of data analysis are more frequently utilized than others. Further research is needed to determine how teachers approach data analysis and whether the issue is access to the technology tools to efficiently retrieve and analyze data or if teachers' knowledge of data analysis that prevents more advanced analyses by teachers.

A more precise measure of teacher classroom assessment literacy would be a welcome addition to educational research. As Plake, et al., (1993) noted in when analyzing the results from the national administration of the CALI in the early 1990s, the criterion-referenced nature of this instrument warrants caution when trying to interpret results based on subscales. Despite the limitations of the CALI, the trend noted between the initial administration of this instrument and the results from this study indicate that despite almost 20 years of attention, the field of classroom assessment literacy is still a matter of concern for educators and one the needs further study.

The failure to find a relationship between teachers' formative use of summative assessments and their assessment literacy level does not mean one does not exist. The use of summative assessment in a formative way was measured using frequency of analysis data, a limitation that could mask a relationship. Although an indication of teachers' consideration of the formative assessment value of summative assessment data, the limitation of this measure makes interpretation of these results difficult to evaluate. The limitations of this

study's measure of the use of summative assessments in a formative way might not adequately describe all teachers are doing with assessment data. It stands to reason that those teachers who use summative assessments to gain insight into student learning and design appropriate student interventions would have a heightened knowledge of the effective use of assessment data. A more precise and reliable measure of the use of summative assessment in a formative way provide valuable insight into teacher practice and any relationship between their practice and their assessment literacy level.

The relationship between teachers' formative use of assessment data and teachers' assessment literacy level is an area of research that needs more study. Future research could provide a more descriptive study of how teachers interpret and use summative assessment data. A mixed-methods design, utilizing not just survey measures, but focus groups and interviews would provide the rich detail this study cannot. Coupled with a more detailed measure of teacher assessment literacy, perhaps a relationship could be more definitively established.

Teachers in this school division appear to be changing their instructional practices on the basis of assessment data, although it remains to be seen how much of this change is due to a formal analysis of assessment data or the more informal formative assessments of observation, questioning and an intuitive feeling. Further study might look more closely at teachers' planning process and determine how much planning is related to specific assessment data. McLeod (2005) cites the need to set SMART goals (specific, measurable, attainable, results-oriented, and time-bound) for effective DDDM and assessment data analysis is certainly a key component to this process.

Though this study does not specifically indicate how much of teachers' instructional change data is due to specific analysis of summative assessment, it is gratifying to see the

innovative and flexible approaches teacher employ to ensure all students in this division have an opportunity to learn.

Limitations

Several limitations of this study are associated with the nature of survey research. The first limitation included concerns regarding confidentiality of responses (McMillan, 2004). This was especially a concern with the use of an electronic web-based survey. Participants might be reluctant to respond out of concern that somehow their responses could be traced back to them through the initial email solicitation. Additionally, since Part III of the survey asked questions that clearly had only one right answer, participants might have felt intimidated that incorrect responses would reflect poorly on their professional knowledge and negatively reflect on their teaching ability.

This limitation was addressed through the setup of the web-based survey program. An email invitation with the link to the survey was used to invite participation rather than including the survey as an attachment or part of the invitation email and having participants email their response. Additionally, the web-based program was set up to block all collection of email and IP addresses so no information was kept that could have identified any of the respondents. Assurances of confidentiality were included in the cover email to help assuage participants concerns.

Additional limitations were those commonly found with self-administered surveys. The typically low return rate required a sufficient sample population to ensure adequate data for analysis (Mitchell & Jolley, 2007). Additionally, since survey response required motivation on the participants' part, there was the potential for a bias sample (Mitchell & Jolley, 2007) with only those with the greatest interest, and perhaps similar practices, responding. Both these limitations were addressed through the use of a sufficiently large sample size and the fact that all participants were practicing educators with at least a

professional interest in the topic.

One additional limitation in survey designs of this nature is the extent to which response are accurate. This could be failure of the participant to understand a particular question or lack of motivation on the part of the participant. In a self-administered survey there is no opportunity to ask for clarification and some questions may be more applicable to one school level than another. Self-administered surveys also prevent further exploration of a response, leaving some responses either inaccurate due to a misunderstanding by the participant or failure to elicit an accurate response due to poor wording (Dillman, 2000). The use of a pilot survey, with participant feedback, helped identify those survey items that were either confusing or not applicable for a particular school level.

Conclusions

This research provided a descriptive study of teachers' use of summative assessment data in formative way. The study also provided a measure of teachers' assessment literacy level for those assessment standards related to the use of assessment data to make instructional decisions. This descriptive study failed to establish a relationship between teachers' formative use of summative assessment in a formative way and their assessment literacy level. Despite not finding a relationship between teachers' assessment literacy and their use of assessment data to make instructional decisions, teachers nevertheless are employing various instructional interventions to enhance student learning, indicating some knowledge base but perhaps not one specifically measured in this study. In the narrative response item, teachers indicated they evaluate students' progress and proscribe instructional interventions according to student need. Differentiation and remediation were two common responses to teachers' evaluating of student assessment data. Recognizing the need to evaluate assessment data, and then taking instructional action to address student learning is one of the key components to McLeod's DDDM model (McLeod, 2005).

For building administrators, this study shows teachers are analyzing summative assessment but perhaps missing what Bernhardt (2000) suggests is one of the more powerful methods of data analysis: disaggregation by student characteristics (Bernhardt, 2000) such as those delineated by AYP subgroups. A review of the technology resources available for teachers to efficiently disaggregate data by student characteristics, as well as SOL standards, might be beneficial.

Furthermore, educational leaders might want to evaluate the strength of the collaborative culture within the school. Bernhardt (2000) notes teachers reliance on their own intuition and a concern about perception of their instructional competence when they compare their assessment data to colleagues' data often undermine teachers' willingness to disaggregate data. As Lachat and Smith (2005) found in their study, disaggregating data can reveal teachers' misconception about instruction and other factors that can affect student achievement. For one group of participants in Lachat and Smith's study, low student achievement had been attributed to attendance issues; however, when teachers looked at the disaggregated data, their perception was challenged. Comparing the performance of students with low attendance compared to those with acceptable attendance revealed significant deficiencies in both groups (Lachat & Smith, 2005). Collaboratively considering disaggregated assessment data prompts teachers to reflect on the effectiveness of instruction. Establishing a collaborative culture, where teachers feel the effect of a team effort rather than a solitary effort can further enhance an open discussion of assessment data and ways to share technique and strategies to raise student achievement.

Those responsible for staff development should consider ways to increase teachers' assessment literacy levels. It is surprising, twenty years after the establishment of standards for teacher competence in classroom assessment, in the context of this and Plake's study, teacher scores have not increased. Accountability models have increasingly become the

norm, and the days of depending on one traditionally trained, division psychometrician to perform data analysis are fast declining (Lukin, Bandalos, Eckhout & Mickelson, 2004). Every classroom teachers needs to serve as his/her own psychometrician. Staff development should focus on making every teacher more assessment literate to give teachers the skills to do the advanced analysis that the literature has shown to be beneficial for instructional decisions.

The educational community has made great strides in bringing the science of teaching on par with the art of teaching. However, the results of this study show that there is still progress to be made. Promoting assessment literacy while giving teachers the knowledge and tools they need to effectively organize and analyze summative assessment data will help all divisions reap the full benefit of mandated assessments. Teachers can make informed decisions about the effectiveness of their instruction, based on summative assessment data. Rather than looking at state and division mandated testing as something separate from instruction, teachers will begin to consider these assessments as an enhancement to instruction. If so, the full benefit will be felt not just by the division, the school and the teachers. The student will be the primary benefactor.

List of References

List of References

American Federation of Teachers, National Council on Measurement in Education, National Education Association. (1990). *Standards for Teacher Competence in Educational Assessment of Students* (http://www.unl.edu/buros/bimm/html/article3.html). .

Anderegg, C. C. (2007). *Classrooms and Schools Analyzing Student Data: A Study of Educational Practice.* Unpublished doctoral dissertation, Pepperdine University, Graduate School of Education and Psychology.

Bernhardt, V. L. (2000). Intersections: New routes open when one type of data crosses another. *Journal of Staff Development, 21*(1), 33-36.

Bernhardt, V. L. (2004, November/December). Continuous improvement: It takes more than test scores. *ACSA Leadership,* pp. 16-19.

Bernhardt, V. L. (2005, February). Data tools for school improvement. *Educational Leadership,* pp. 66-69.

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*, pp. 9-21.

Black, P., & Wiliam, D. (1998a, October). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, pp. 139-148.

Black, P., & Wiliam, D. (1998b). Assessment and Classroom Learning. *Assessment in Education, 5*(1), 7-74.

Black, P., & Wiliam, D. (2003, October). "In Praise of Educational Research': Formative assessment. *British Educational Research Journal, 29*(5), 623-637.

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*, pp. 9-21.

Bloom, B. (1968). Learning for mastery: Instruction and curriculum (Evaluation Comment). .

Brunner, C., Fasca, C., Heinze, J., Honey, M., Light, D., Mandinach, E., et al. (2005). Linking Data and Learning: The Grow Network Study. *Journal of Education for Students Placed at Risk, 10*(3), 214-267.

Center for Education. (2001). *Classroom Assessment and the National Science Education Standards* (National Academy of Sciences). Retrieved 20 August, 2007, from http://books.nap.edu/openbood.php?isbn=030906998X.

Cobanoglu, C., Warde, B., & Moreo, P. J. (2001, Quarter 4). A comparison of mail, fax, and web-based survey methods. *International Journal of Market Research, 43*, 441-452.

Dillman, D. A. (2000). *Mail and Internet Surveys.* New York, New York: John Wiley & Sons, Inc.

DuFour, R. (2004, May). What is a "professional learning community?". *Educational Leadership*, pp. 6-11.

Gordan, A. (2002). SurveyMonkey.com: Web-based survey and evaluation system - http://www.SurveyMonkey.com. *Internet and High Education, 5*, 83-87.

Guskey, T. R., & and others. (1990). Using mastery in the regular classroom to help learning disabled and at-risk students. (ED335840, pp. 1-18). .

Husman, J., Brem, S., & Duggan, M. (2005, Fall). Student goal orientation and formative assessment. *Academic Exchange Quarterly, 9*(3).

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly, 69*(1), 94-101.

Lachat, M. A., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk, 10*(3), 333-349.

Lukin, L. E., Bandalos, D. L., Eckhout, T., J., & Mickelson, K. (2004, Summer). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, pp. 26-32.

Mandinach, E. B., Honey, M., & Light, D. (2006, April). *A Theoretical Framework for Data-Driven Decision Making.* Presented at the AERA Annual Meeting, San Francisco, California.

McLeod, S. (2005, May). *Data-Driven Teachers* (UCEA Center for the Advanced Study of Technology Leadership in Education). www.microsoft.com/Education/ThoughtLeaders.aspx.

McMillan, J. H. (2001). *Essential Assessment Concepts for Teachers and Administrators.* Thousand Oaks, California: Corwin Press, Inc.

McMillan, J. C. (2004). *Educational Research: Fundamentals for the Consumer.* Boston, MA: Pearson Education, Inc.

Mertler, C. A. (2003). Patterns of response and nonresponse from teachers to traditional and web surveys. *Practical Assessment, Research and Evaluation, 8*(22). Retrieved 6 December 2008, from http://www.pareonline.net/getvn.asp?v=8&n=22.

Mitchell, M. L., & Jolley, J. M. (2007). *Research Design Explained.* Belmont, California: Thomson Higher Education.

Murnane, R. J., Sharkey, N. S., & Boudett, K. P. (2005). Using student-assessment results to improve instruction: Lessons from a workshop. *Journal of Education for Students Placed at Risk, 10*(3), 269-280.

National Center for Education Statistics. *District Information: [School Division]*. Retrieved 7 April 2009, from Institute of Education Services U.S. Department of Education

National Commission on Excellence in Education. (1983). *A Nation at Risk: The Imperative for Educational Reform* (Washington, D.C.: U.S. Department of Education). .

Nelson, M., & Eddy, R. M. (2008, Spring). Evaluative thinking and action in the classroom. In T. Berry & R.M. Eddy (eds). *Consequence of No Child Left Behind for Educational Evaluation. New Directions for Evaluation, 117,* 37-46.

No Child Left Behind Act of 2001. (2001). Retrieved 10 October 2009, from http://www.ed.gov/nclb.

Petersen, J. L. (2007, Winter). Learning: The brave new world of data-informed instruction.

Plake, B. S., & Impara, J. C. (1993, Winter). Assessment competencies of teachers A national survey. *Educational Measurement: Issues and Practice, 12*(4), 10-12, 39.

Sapsford, R. (2007). *Survey Research.* Thousand Oaks, California: SAGE Publications, Inc.

Schaefer, D. A., & Dillman, D. A. (1998). Development of a standard e-mail methodology: Results of an experiment. *Public Opinion Quarterly, 63*(3), 378-397.

Shannon, D., M., & Bradshaw, C. C. (2002). A comparison of response rate, response time, and costs of mail and electronic surveys. *The Journal of Experimental Education, 70*(2), 179-192.

Sharkey, N., & Murnane, R. J. (2003, November). Learning from student assessment results. *Educational Leadership.*

Stiggins, R. (2004, September). New assessment beliefs for a new school mission. *Phi Delta Kappan,* pp. 22-27.

Stiggins, R. (2008). *Assessment manifesto: A call for the development of balanced assessment systems.* Portland, Oregon: ETS Assessment Training Institute.

Streifer, P. A., Schumann. (2005). Using data mining to identify actionable information: Breaking new ground in data-driven decision making. *Journal of Education for Students Placed at Risk, 10*(3), 281-293.

Vogel, L. R., Rau, W. C., Baker, P. J., & Ashby, D. E. (2006). Bringing assessment literacy to the local school: A decade of reform initiatives in Illinois. *Journal of Education for Students Placed at Risk, 11*(1), 39-55.

Wayman, J. C. (2005). Involving teachers in data-driven decision making: Using computer data systems to support teacher inquiry and reflection. *Journal of Education for Students Placed at Risk, 10*(3), 295-308.

Wayman, J. C., Cho, V., & Johnson, M. T. (2007). The data-informed district: A district-wide evaluation of data used in the Narona County School District. Austin: The University of Texas.

Wayman, J. C., Midgley, S., & Stringfield, S. (2005). *Collaborative teams to support data-based decision making and instructional improvement.* Presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004, March). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*(1), 49-65.

Wilson, V. (Aug). (2007, August). Conversation. Virginia Commonwealth University.

Yeh, S. S. (2006, April). High-stakes testing: Can rapid assessment reduce the pressure? *Teachers College Record, 108*(4), 621-661.

Appendices

Appendix A

Email Permission to use Classroom Assessment Inventory

Attachments can contain viruses that may harm your computer. Attachments may not display correctly.

Nancy R Hoover

From: Barbara Plake [bplake@unlserve.unl.edu] **Sent:** Sun 3/15/2009 1:39 PM
To: Nancy R Hoover
Cc:
Subject: Re: Question about Teacher Classroom Assessment Literacy
Attachments: KEY Literacy test.doc(344B) Tcher Assmt Literacy 1&2.doc(52KB)

Nancy,

I've attached the assessment literacy questionnaire and the key. The questionnaire has two parts; part 1 is the test and part 2 asks for some demographic and perception questions. It is OK to use only part 1 if that fits your needs better. You have my permission to use this questionnaire in your research. I hope you will find it useful.

Barbara Plake

On 3/15/09 11:13 AM, "Nancy R Hoover" <Nancy_Hoover@ccpsnet.net> wrote:

> Good afternoon, Dr. Plake,
>
> I am a doctoral student at Virginia Commonwealth University, in Richmond,
> Virginia, working on my Ph.D. in Educational Leadership. I am currently
> working on my dissertation and intend to study the extent to which teachers
> use assessment data to inform their instructional practice. One thing that
> interests me is the possibility that teachers' assessment literacy has some
> impact on their use of assessment data to make instructional decisions.
>
> I have read with interest your work in developing an instrument to measure
> teachers' classroom assessment literacy - most specifically the article
> referenced below. I writing to ask if you might be able to share with me a
> copy of the survey and give me permission to use it, or use some form of the
> instrument, for my study. Also, if you have any additional information or
> advice regarding this topic to share, I would be very happy to hear from you.
>
> I know you must be very busy and I do appreciate any time and information you
> can send my way.
>
> The article that prompted me to email you:
>
> Plake, B. S., Impara, J. C., & Fager, J. J. (1993) Assessment competencies of
> teachers: A National survey. Educational
> Measurement: Issues and Practice, 12, 10-12.
>
>
> Thank you again,
>
> Nancy Hoover
> Pre-Engineering Specialty Center Coordinator
> Lloyd C. Bird High School
> Chesterfield, VA 23832
> 804-768-6110, ext. 165

Appendix B

Survey to Measure the Extent to which Teachers Use Student Assessment Data to Inform Their Instructional Practice

Opening page of survey

The survey consists of three main parts. Part I includes demographic data such as years teaching experience, school level assignment, etc. Part II asks questions regarding the types and frequencies of student assessments and assessment data analysis you use in your classroom, as well as your perception of factors that influence your analysis of student assessment data.

Finally, Part III asks questions directly related teachers' knowledge of educational assessment of students. Your response will provide valuable insight into the way analysis of student assessment data affects instruction. Your responses are completely confidential and secure. You may exit the survey at any time you no longer feel like participating.

Thank you for your participation in this study.

Part I: Demographic Information

1. Level of teaching assignment:
   - ☐ Elementary
   - ☐ Middle
   - ☐ High

2. Years of teaching experience:
   - ☐ 0-3 years
   - ☐ 4-10 years
   - ☐ 11+ years

3. Years in teaching in my current assignment:
   - ☐ 0-3 years
   - ☐ 4-10 years
   - ☐ 11+ years

4. Gender:
   - ☐ Male
   - ☐ Female

5. Ethnicity:
   - ☐ American Indian
   - ☐ Asian
   - ☐ Black, not of Hispanic origin
   - ☐ Hispanic
   - ☐ White
   - ☐ Pacific Islander
   - ☐ Other

6. Highest degree attained:
   - ☐ Bachelors
   - ☐ Masters
   - ☐ Doctorate

7. Content of primary teaching responsibility (mark all that apply):
   - ☐ English
   - ☐ Mathematics
   - ☐ Science
   - ☐ Social Studies
   - ☐ Elementary
   - ☐ Other

Part II: Teacher Practice

Methods and Frequency of Data Analysis

8. How often do you use the following methods to analyze student assessment data?	At least Weekly	At least Monthly	At least Quarterly (every 9 weeks)	Each Semester	Annually	Never
a) Examine the class average (Mean)	□	□	□	□	□	□
b) Look at frequency of assessment scores (Mode)	□	□	□	□	□	□
c) Look at the range of student scores (Standard Deviation)	□	□	□	□	□	□
d) Look at overall student performance disaggregated by AYP subgroups	□	□	□	□	□	□
e) Look at overall student performance disaggregated by individual SOL standards	□	□	□	□	□	□
f) Look at student scores question by question (item analysis)	□	□	□	□	□	□

Type of Student Assessments Administered

9. How often do you administer the following assessments in your class?	Weekly	Monthly	Quarterly	Each Semester	Annually	Never
a) Teacher-generated assessments	□	□	□	□	□	□
b) Departmental common assessments	□	□	□	□	□	□
c) Division Benchmark assessments	□	□	□	□	□	□
d) Released Statewide Standards of Learning (SOLs) assessment/items	□	□	□	□	□	□
e) Nationally norm-referenced assessments	□	□	□	□	□	□

Frequency of analysis by Type of Student Assessments Administered

10. How often do you analyze student test results from the following types of assessments in your class?	Weekly	Monthly	Quarterly	Each Semester	Annually	Never
a) Teacher-generated assessments	□	□	□	□	□	□
b) Departmental common assessments	□	□	□	□	□	□
c) Division benchmark assessments	□	□	□	□	□	□
d) Released Statewide Standards of Learning (SOLs) assessment/items	□	□	□	□	□	□
e) Nationally norm-referenced assessments	□	□	□	□	□	□

Frequency of specific types of analysis of students' assessments administered

11. How often do you disaggregate student test results to determine the performance of AYP subgroups in your class?	Weekly	Monthly	Quarterly	Each Semester	Annually	Never
a) Teacher-generated assessments	□	□	□	□	□	□
b) Departmental common assessments	□	□	□	□	□	□
c) Division benchmark assessments	□	□	□	□	□	□
d) Released Statewide Standards of Learning (SOLs) assessment/items	□	□	□	□	□	□
e) Nationally norm-referenced assessments	□	□	□	□	□	□

Instructional Practice

12. I have made the following changes to my instructional practice based on student assessment data:	Yes	No
a) Changing the pacing of future instruction	□	□
b) Regroup students within the classroom	□	□
c) Differentiate instruction to remediate and/or enhance learning.	□	□
d) Re-teach topics/concepts.	□	□
e) Remediate and re-test for specific unit.	□	□
f) Pacing prevents re-teaching but consideration is given to instructional strategies for next unit.	□	□
g) I have changed my instructional practice but in ways not listed here. i) Please note those changes: _____	□	□

Part III: Teacher Assessment Literacy Questionnaire

Please read each item carefully and select the response you think is the best one. If you think you know which is best, even if you are not positive, mark that response.

Choosing assessment methods.

13. What is the most important consideration in choosing a method for assessing student achievement?

    a. Ease of scoring the assessment.
    b. Ease of preparing the method of assessment.
    c. Accuracy of assessing attainment of instructional objectives.
    d. Acceptance by the school administration.

14. When scores from a standardized test are said to be reliable, what does it imply?

    a. Student scores from the test can be used for a large number of educational decisions.
    b. If a student retook the same test, he or she would get a similar score on each retake.
    c. The test score is a more valid measure than teacher judgments.
    d. The test score accurately reflects the content of instruction in classes where the test is administered.

15. Mrs. Bruce wished to assess her students' understanding of the method of problem solving she had been teaching. Which assessment strategy below would be most valid?

    a. Select a textbook that has a "teacher's guide" with a test developed by the authors.
    b. Develop an assessment consistent with an outline of what she has actually taught in the class.
    c. Select a standardized test that provides a score on problem solving skills.
    d. Select an instrument that measures students' attitudes about problem solving strategies.

16. What is the most effective use a teacher can make of an assessment strategy that requires students to show their work , e.g., the way they arrived at a solution to a problem or the logic used to arrive at a conclusion?

    a. Assigning grades for a unit of instruction on problem solving.
    b. Providing instructional feedback to individual students.
    c. Motivating students to attempt innovative ways to solve problems.
    d. None of the above.

17. Ms. Green, the principal, was evaluating the teaching performance of Mr. Wajesa, the fourth grade teacher. One of the things Ms. Green wanted to learn was if the students were being encouraged to use higher order thinking skills in the class. What documentation would be the most valid to help Ms. Green to make this decision?

    a. Mr. Wajesa's lesson plans.
    b. The state curriculum guides for fourth grade.
    c. Copies of Mr. Wajesa's unit tests or assessment strategies used to assign grades.
    d. Worksheets completed by Mr. Wajesa's students, but not used for grading.

Developing assessment methods

18. A teacher wants to document the validity of the scores from a classroom assessment strategy she plans to use for assigning grades on a class unit. What kind of information would provide the best evidence for this purpose?

    a. Have other teachers judge whether the assessment strategy covers what was taught.
    b. Match an outline of the instructional content to the content of the assessment strategy.
    c. Let students in the class indicate if they thought the assessment was valid.
    d. Ask parents if the assessment reflects important learning outcomes.

19. Which of the following actions would most likely increase the reliability of Mrs. Lockwood's multiple choice end-of-unit examination in physical science?

    a. Use a blueprint to develop the test questions.
    b. Change the test format to true-false questions.
    c. Add more items like those already in the test.
    d. Add an essay component.

20. Ms. Guardia wants to assess her students' skills in organizing ideas rather than just repeating facts. Which words should she use in formulating essay exercises to achieve this goal?

    a. compare, contrast, criticize
    b. identify, specify, list
    c. order, match, select
    d. define, recall, restate

21. Mr. Woodruff wanted his students to appreciate the literary works of Edgar Allen Poe. Which of his test items shown below will best measure his instructional goal?

    a. "Spoke the raven, nevermore." comes from which of Poe's works?
    b. True or False: Poe was an orphan and never knew his biological parents.
    c. Edgar Allen Poe wrote:
        1. Novels
        2. Short stories
        3. Poems
        4. All of the above.
    d. Discuss briefly your view of Poe's contribution to American literature.

22. Several students in Ms. Atwell's class got low scores on her end-of-unit test in doing multi-step story problems in mathematics. She wanted to know which students were having similar problems so she could group them for instruction. Which assessment strategy would be best for her to use for grouping students?

    a. Use the test provided in the "teacher's guide."
    b. Have the students take a test that has separate items for each step of the process.
    c. Look at the student's records and standardized test scores to see which topics the students had not performed well on previously.
    d. Give students story problems to complete and have them show their work.

Administering, scoring and interpreting assessment results.

23. Many teachers score classroom tests using a 100-point percent correct scale. In general, what does a student's score of 90 on such a scale mean?

    a. The student answered 90% of the items on this test correctly.
    b. The student knows 90% of the instructional content of the unit covered by this test.
    c. The student scored higher than 90% of all the students who took the test.
    d. The student scored 90% higher than the average student in the class.

24. Students in Mr. Jakman's science class are required to develop a model of the solar system as part of their end of unit grade. Which scoring procedure below will maximize the objectivity of assessing these student projects?

    a. When the models are turned in, Mr. Jakman identifies the most attractive models and gives them the highest grades, the next most attractive get a lower grade and so on.
    b. Mr. Jakman asks other teachers in the building to rate each project on a 5 point scale based on their quality.
    c. Before the projects are turned in, Mr. Jakman constructs a scoring key based on the critical features of the projects as identified by the highest performing students in the class.
    d. Before the projects are turned in, Mr. Jakman prepares a model or blueprint of the critical features of the product and assigns scoring weights to these features. The models with the highest scores receive the highest grade.

25. At the close of the first month of school, Mrs. Friend gives her fifth grade students a test she developed in social studies. Her test is modeled after a standardized social studies test. It presents passages and then asks questions related to understanding and problem definition. When the test was scored, she noticed that two of her students who had been performing well in their class assignments, scored a lot lower than other students. Of the following types of additional information which would be most helpful in interpreting the results of this test?

    a. The gender of the students.
    b. The age of the students.
    c. Reliability data for the standardized social studies test she used as the model.
    d. Reading comprehension scores for the students.

26. Frank, a beginning fifth grader, received a G. E. (grade equivalent score) of 8.0 on the Reading Comprehension subtest of a standardized test. This score should be interpreted to mean that Frank:

    a. can read and understand 8th grade reading level material.
    b. scored as well as a typical beginning 8th grader scored on this test.
    c. is performing in Reading Comprehension at the 8th grade level.
    d. will probably reach maximum performance in Reading Comprehension at the beginning of the 8th grade.

27. When the directions indicate each section of a standardized test is timed separately, which of the following is acceptable test-taking behavior?

    a. John finishes the vocabulary section early; he then rechecks many of his answers in that section.
    b. Mary finishes the vocabulary section early; she checks her answers on the previous test section.
    c. Jane finishes the vocabulary section early; she looks ahead at the next test section but does not mark her answer sheet for any of those items.
    d. Bob did not finish the vocabulary section; he continues to work on that section when the testing time is up.

Using assessment results for decision-making.

28. Ms. Camp is starting a new semester with a factoring unit in her Algebra I class. Before beginning the unit, she gives her students a test on the commutative, associative, and distributive properties of addition and multiplication. Which of the following is the most likely reason she gives this test to her students?

    a. The principal needs to report the results of this assessment to the state-testing director.
    b. Ms. Camp wants to give the students practice in taking tests early in the semester.
    c. Ms. Camp wants to check for prerequisite knowledge in her students before she begins the unit on factoring.
    d. Ms. Camp wants to measure growth in student achievement of these concepts, and scores on this test will serve as the students' knowledge baseline.

29. To evaluate the effectiveness of the mathematics program for her gifted first graders, Ms. Allen gave them a standardized mathematics test normed on third graders. To decide how well her students performed, Ms. Allen compared her students' scores to those of the third-grade norm group. Why is this an incorrect application of standardized test norms?

    a. The norms are not reliable for first graders.
    b. The norms are not valid for first graders.
    c. Third grade mathematics items are too difficult for first graders.
    d. The time limits are too short for first graders.

30. When planning classroom instruction for a unit on arithmetic operations with fractions, which of these types of information have more potential to be helpful?

norm-referenced information: describes each student's performance relative to a other students in a group (e.g. percentile ranks, stanines), or

criterion-referenced information: describes each student's performance in terms of status on specific learning outcomes (e.g., number of items correctly answered for each specific objective)

a. Norm-referenced information.
b. Criterion-referenced information.
c. Both types of information are equally useful in helping to plan for instruction.
d. Neither, test information is not useful in helping to plan instruction.

31. Students' scores on standardized tests are sometimes inconsistent with their performances on classroom assessments, e.g. teacher tests or other in-class activities. Which of the following is **NOT** a reasonable explanation for such discrepancies?

a. Some students freeze up on standardized tests, but they do fine on classroom assessments.
b. Students often take standardized test less seriously than they take classroom assessment.
c. Standardized tests measure only recall of information while classroom assessments measure more complex thinking.
d. Standardized tests may have less curriculum validity than classroom assessment

32. Elementary school teachers in the Baker School system collectively designed and developed a new curricula in Reading, Mathematics, and Science that is based on locally developed objectives and objectives in state curriculum guides. The new curricula were not matched directly to the content of the fourth grade standardized test. A newspaper reports the fourth grade students in Baker Public Schools are among the lowest scoring districts in the State Assessment Program. Which of the following would **invalidate** the comparison between Baker Public Schools and other schools in the state?

a. The curriculum objectives of the other districts may more closely match those of the State Assessment.
b. Other school systems did not design their curriculum to be consistent with the State Assessment test.
c. Instruction in Baker schools is poor.
d. Other school systems have different promotion policies than Baker.

Appendix C

Email Notification for Participants

Initial email.

Dear Colleague,

I am a long-time teacher in Chesterfield County and am currently the specialty center coordinator at Lloyd C. Bird High School; I am also a doctoral student in the Educational Leadership program at Virginia Commonwealth University. For my dissertation research, and in cooperation with Chesterfield County Public Schools, I am surveying randomly selected teachers at the elementary, middle, and high school levels. As one of those randomly selected teachers, I am hoping you will assist me by completing the attached survey. The survey should take approximately 15-20 minutes to complete. Your responses to this survey will be confidential and secure. In no way will your responses be identifiable to you or your school.

You may access the survey at **(insert active link to survey).**

The purpose of this study is to gain understanding of the relationship between student assessment and instruction. I seek to answer the question to what extent do teachers use student assessment data to inform their instructional practice? By participating in this study you will provide helpful insight to what types of student assessment data are most frequently used as well as what conditions foster the most effective use of student assessment data. The survey will remain open and active for the next two weeks, through **(insert date)**.

Thank you in advance for your participation. Your response will help provide valuable information to advance teacher practices in general and your division specifically. Should you have any questions regarding this survey, please feel free to contact me at 804-678-8419 or at nancyrhoover@comcast.net.

Sincerely,

Nancy Hoover
Doctoral Student, Educational Leadership
Virginia Commonwealth University
Pre-Engineering Specialty Center Coordinator
Lloyd C. Bird High School

Reminder email to be sent out one week after initial email.

Dear Colleague,

A week ago you were asked to complete a survey as part of a study to answer the question to what extent do teachers use students assessment data to inform their instructional practice? Your participation provides helpful insight to what types of student assessment data are most frequently used as well as what conditions foster the most effective use of students' assessment data.

If you have had a moment to complete the survey I would like to thank you. If you have not yet had a chance to access the survey, I hope you will be able to in the coming week. Your response will help provide valuable information to advance teacher practices in general and your division specifically. The survey will remain open and active until (insert date).

Should you have any questions regarding this survey, please feel free to contact me at 804-678-8419 or at nancyrhoover@comcast.net.

Sincerely,

Nancy Hoover
Doctoral Student, Educational Leadership
Virginia Commonwealth University
Pre-Engineering Specialty Center Coordinator
Lloyd C. Bird High School

Appendix D

Table D

Research questions and related data analysis method

Research Question	IV	DV	Data Analysis
To what extent do teachers use summative assessments in a formative way?		Frequency of administration and analysis of summative assessments	• Frequency distribution
What is the level of teacher assessment literacy in a large suburban school district?	Teacher characteristics	Classroom Assessment Literacy Inventory	• Mean scale score • MANOVA
What is the relationship between teachers' reported use of summative assessments in a formative way and their assessment literacy level?	Frequency of methods of data analysis Frequency of analysis of summative assessments	Teacher Assessment Literacy Scores for Standards 1, 2, 3, 4	• Frequency distribution • MANOVA
What changes in their instructional practice result from teachers' use of assessment data?		Instructional Practice	• Frequency distribution

Appendix E

Histograms of Assessment Literacy Scores

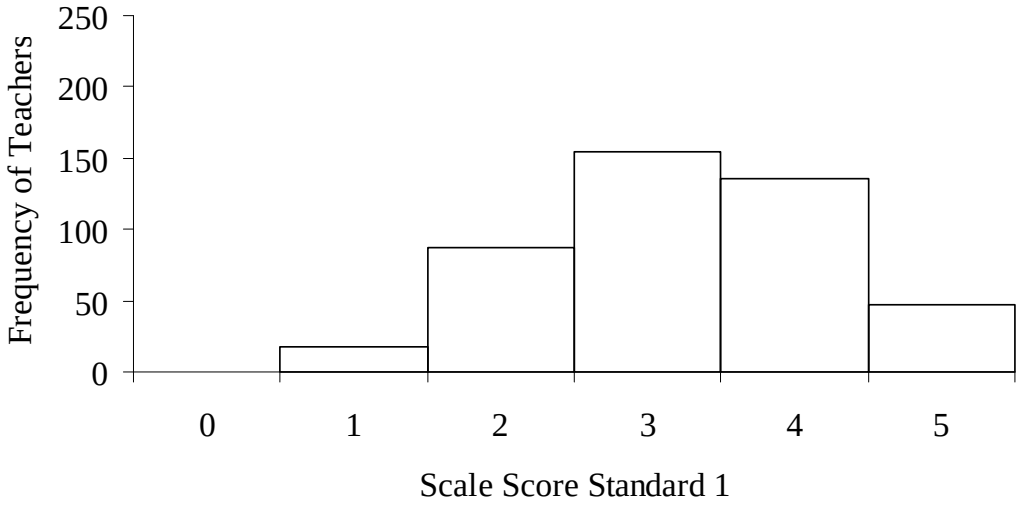


Figure E1. Distribution of Standard 1 Assessment Literacy Scores

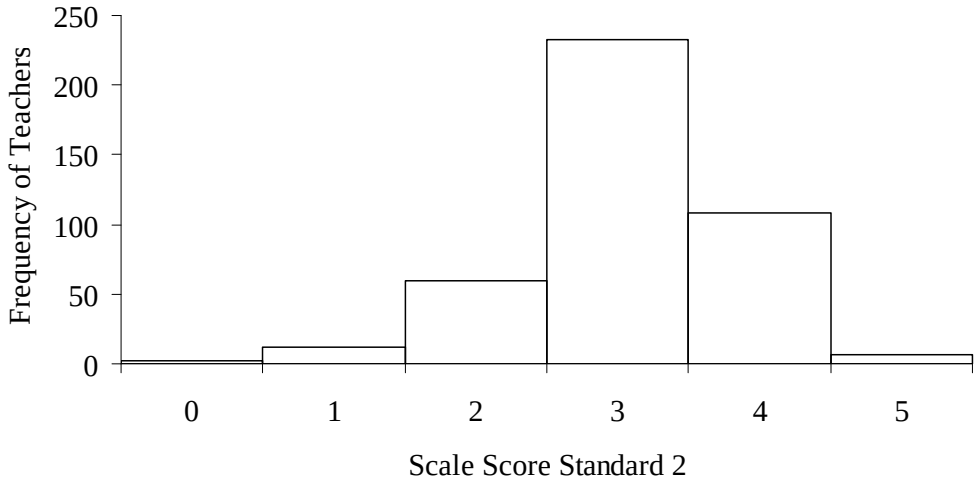


Figure E2. Distribution of Standard 2 Assessment Literacy Scores

Figure E3. Distribution of Standard 3 Assessment Literacy Scores

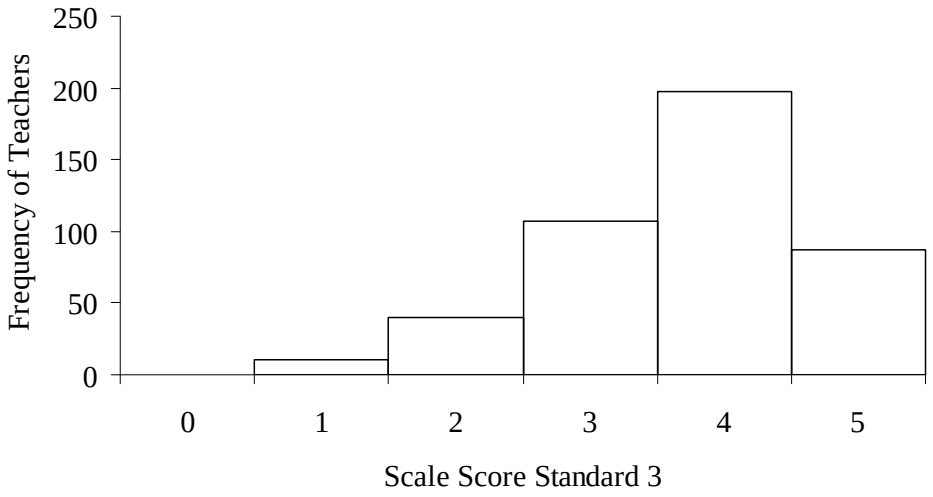
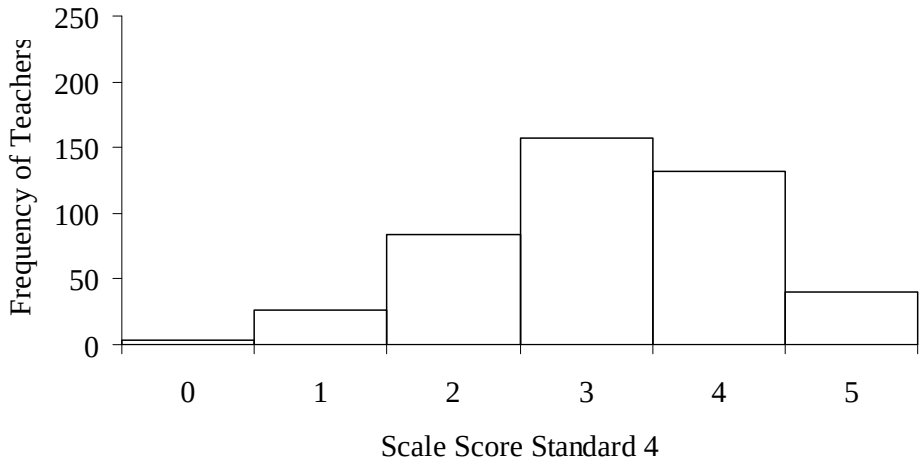


Figure E4. Distribution of Standard 4 Assessment Literacy Scores.



Appendix F

Post Hoc Analyses

Table 1

Bonferroni Post Hoc Analysis for Classroom Assessment Literacy Standards by School Level

	Standard 1		
	Diff	*SE*	*p*
Elementary*Middle	-.257	.121	.101
Elementary*High	-.499*	.120	.000
Middle*Elementary	.257	.121	.101
Middle*High	-.242	.112	.094
High*Elementary	.499*	.120	.000
High*Middle	.242	.112	.094

p < .05

Table 2

Post Hoc Analysis for Classroom Assessment Literacy Standards by Years Experience

	Standard 1			Standard 3			Standard 4		
	Diff	*SE*	*p*	*Diff*	*SE*	*p*	*Diff*	*SE*	*p*
0-3yrs* 4-10yrs	-.063	.165	1.000	-.319	.156	.124	-.522*	.171	.007
0-3yrs* 11+yrs	-.216	.162	.547	-.413*	.152	.021	-.594*	.167	.001
4-10yrs* 0-3yrs	.063	.165	1.000	.319	.156	.124	.522*	.171	.007
4-10yrs* 11+yrs	-.278*	.103	.022	-.094	.097	.998	-.072	.107	1.000
11+yrs* 0-3yrs	.216	.162	.547	.413*	.152	.021	.594*	.167	.001
11+yrs* 4-10yrs	.278*	.103	.022	.094	.097	.998	.072	.107	1.000

* $p < .05$

Table 3

Post Hoc Analysis for Primary Teaching Responsibility

	Standard 1		
	Diff	*SE*	*p*
English*Mathematics	-.163	.148	1.000
English*Science	-.196	.158	1.000
English*Social Studies	.119	.166	1.000
English*Elementary	.300	.140	.486
English*Other	.533	.370	1.000
Mathematics*English	.163	.148	1.000
Mathematics*Science	-.033	.158	1.000
Mathematics*Social Studies	.282	.166	1.000
Mathematics*Elementary	.436*	.140	.015
Mathematics*Other	.696	.370	.910
Science*English	.196	.158	1.000
Science*Mathematics	.033	.158	1.000
Science*Social Studies	.315	.175	1.000
Science*Elementary	-.496*	.151	.016
Science*Other	.729	.374	.781
Social Studies*English	-.119	.166	1.000
Social Studies*Mathematics	-.282	0166	1.000
Social Studies*Science	-.315	.175	1.000
Social Studies*Elementary	.181	.158	1.000
Social Studies*Other	.414	.377	1.000
Elementary*English	-.300	.140	.486
Elementary*Mathematics	-.436*	.140	.015
Elementary*Science	-.496*	.151	.016
Elementary*Social Studies	-.181	.158	1.000
Elementary*Other	.233	.367	1.000
Other*English	-.533	.370	1.000
Other*Mathematics	-.696	.370	.910
Other*Science	-.729	.374	.781
Other*Social Studies	-.414	.377	1.000
Other*Elementary	-.233	.367	1.000

** $p < 0.05$*

Table 4

Post Hoc Analysis for Frequency of Types of Summative Assessments Analyzed: Released SOL Items

Standard 4	*Diff*	*SE*	***p***
Analyze < 2 times/yr * Analyze 2-4 times/yr	.29*	.114	.039
Analyze < 2 times/yr * Analyze > 4 times/yr	-.05	.131	1.000
Analyze 2-4 times/yr * Analyze < 2 times/yr	-.29*	.114	.039
Analyze 2-4 times/yr * Analyze > 4 times/yr	-.33*	.135	.044
Analyze > 4 times/yr * Analyze < 2 times/yr	.05	.131	1.000
Analyze > 4 times/yr * Analyze 2-4 times/yr	.33*	.135	.044

p < .05

Table 5

Post Hoc Analysis for Frequency of Methods of Data Analysis: Examining Mode

Standard 3	*Diff*	*SE*	*p*
Analyze < 2 times/yr * Analyze 2-4 times/yr	.21	.154	.504
Analyze < 2 times/yr * Analyze > 4 times/yr	.34*	.139	.045
Analyze 2-4 times/yr * Analyze < 2 times/yr	-.21	.154	.504
Analyze 2-4 times/yr * Analyze > 4 times/yr	.13	.106	.695
Analyze > 4 times/yr * Analyze < 2 times/yr	-.34*	.139	.045
Analyze > 4 times/yr * Analyze 2-4 times/yr	-.13	.106	.695

Vita

Nancy R. Hoover was born in 1956 in Richmond, Virginia. The oldest of three children, she grew up in Chesterfield County and attended Virginia Tech upon graduation from Meadowbrook High School in 1974. After two years at Virginia Tech, Nancy returned to Richmond and completed the Clinical Laboratory Assistant program at Richmond Memorial Hospital. Upon completion of this program, she worked in the Chemistry Department for a local hospital for 3 years. She left the work force to raise her family and returned to Virginia Commonwealth University in 1992 in order to complete her degree so she could become a science teacher. She graduated in 1996 with a Bachelor's of Science in Interdisciplinary Science with a concentration in chemistry and a Masters in Teaching. She taught 8th grade physical science for four years before transferring to Lloyd C. Bird High School to teach physics. In 2005, Nancy became the director of the Governor's Academy for Engineering Studies at Lloyd C. Bird High School, a position she still holds.