11-2010

Educational assessment: Exploring the relationship between computer adaptive testing, data analysis, and student achievement

Michael Donhost

DePaul University

School of Education

EDUCATIONAL ASSESSMENT: EXPLORING THE RELATIONSHIP BETWEEN COMPUTER ADAPTIVE TESTING, DATA ANALYSIS, AND STUDENT ACHIEVEMENT

A Dissertation in

Educational Leadership

by

Michael J. Donhost

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

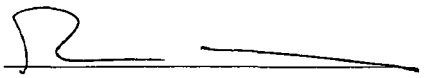
Doctor of Education

July 2009

SIGNATORY PAGE

We approve the dissertation of Michael J. Donhost.

Date of Signature



7/15/09

Gayle Mindes, Ed.D.

Professor of Education

Dissertation Advisor

Chair of Committee

7/15/09

Chris McCullough, Ed.D.

Assistant Dean for Assessment

7/15/09

Steve Hamlin, Ed.D.

Principal, Waukegan High School

ABSTRACT

The research on the potential relationship of the Northwest Evaluation Association's Measures of Academic Progress (NWEA MAP) and the Illinois Standards Achievement Test (ISAT) was conducted in two phases and focused on middle schools with a grade configuration of 6-8 in the state of Illinois. The first phase of the research examined whether or not a significant relationship existed between NWEA MAP test participation by middle schools in Illinois, and ISAT growth over time. A total of 86 schools were involved in the research associated with Phase 1. The second phase of the study aimed to find out if the NWEA members' self-reported level of data-driven decision making practices was associated with ISAT growth over time. An original data-driven decision making (DDDM) survey was employed to measure the use of data-driven decision making practices at each middle school selected for the study. The DDDM survey questions were created based on the major recommendations that are found in the book *Data Wise* (Boudett, 2005a). For each of the 8 major *Data Wise* recommendations, one question was formed to measure the level of implementation for each school. A total of 31 of the 43 identified principals participated in the study, for a return rate of 72%. Ultimately, this quantitative research failed to reject both of the null hypotheses. However, the DDDM survey did illuminate a discrepancy in the reported implementation level of *Assessment Literacy* compared to the other 7 survey questions. Recommendations for future research include conducting an in-depth study of a few schools that have a proven record of rapid growth, in an attempt to distill the specific factors that

allow certain NWEA member schools to experience success. A second approach would be to follow a cohort of schools from the implementation stage with NWEA through the first few years of use, attempting to identify the decisions and actions that lead to measurable growth. Finally, deciphering which terms and processes are crucial to an effective understanding of assessment literacy could also prove beneficial to the broader educational community.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express gratitude to the authors of *Data Wise* and the Harvard Education Press in granting copyright permission for the use of *Data Wise* concepts in the production of the data-driven decision making survey.

***Data Wise: A Step-by-Step Guide to Using Assessment Results to Improve Teaching and Learning*, edited by Kathryn Parker Boudett, Elizabeth A. City, and Richard J. Murnane, (Cambridge, MA: Harvard Education Press, 2005). For more information, please visit www.harvardeducationpress.org, or call 1-888-437-1437.**

I would also like to express gratitude to the Northwest Evaluation Association for providing requested data that proved instrumental to the success of this study.

Northwest Evaluation Association ("NWEA"), a nonprofit organization located at 5885 SW Meadows Road, Suite 200, Lake Oswego, Oregon 97035.

CHAPTER 1

INTRODUCTION TO THE STUDY

Background

In contemporary society, the term *assessment* is perceived as anything from the panacea for educational improvement, to a persistent threat to quality educational practices. The disparity in beliefs regarding assessment can be distilled, in part, to the presence of distinct foci to which assessments are geared, with two branches being assessments for accountability, and assessments for school improvement. Assessments for school improvement are those that have the capacity to inform instructional decision making. Assessments for accountability on the other hand, are not designed to directly improve education, rather, to summarize the achievement level of various educational entities.

Fueled in large part by the No Child Left Behind PL 107-110 legislation of 2002, public education today is consumed with the role that accountability-related standardized assessment regulations and practices play in day-to-day teaching and learning. Proponents, usually legislators and policy makers, see the standardized assessments as the primary method in which schools can prove accountable practice, usually through the use of state-developed instruments. Educators however, often view the standardized assessments as a threat to holistic instruction. The negatives associated with standardized assessments include a history of ill-informed test formation, including biased questioning techniques, a focus on low level knowledge and skills, and questions formed to sort and select, rather than to show mastery of material. More recently, the

negatives associated with standardized assessment have shifted towards how the test results are being utilized. The use of results from a single state test to make high-stakes, often life altering, decisions is perhaps the most polarizing current reality in education.

The state-level accountability standardized assessments have created an impetus for school improvement efforts, though ironically their structure and focus, make the state level assessment results of limited use in the very school improvement efforts they innately promote. The assessments, being summative in nature, lack the capacity to inform instruction, due to the length of time that elapses between test administration and the reporting of results, and the associated lack of content/standard specific detail.

In response to the growing public and political pressure for schools to improve, many districts are exploring alternative assessment techniques for assistance in the school improvement process. One technique, which is rapidly increasing in popularity, is the utilization of computer-adaptive assessments. Schools that utilize computer adaptive assessments seek to inform instruction/learning with detailed and timely information on the progress of each student.

The common denominator across the various forms and functions of assessment is that on some level, for some purpose, information is produced. The proliferation of information from the various assessments being utilized, has created a need for educators to understand how to make decisions based on the

information they accumulate. The information, in the form of data, is, to an ever-increasing degree, driving educational decision-making.

Problem

To prepare children for successful achievement on the state standardized achievement assessments, a growing number of schools in Illinois, and across the nation, are utilizing the computerized adaptive tests that were created by the Northwest Evaluation Association, or NWEA. The adaptive tests, called Measures of Academic Progress, or simply MAP, are aligned to each member's state standards, and aim to provide accurate information about student growth and achievement so that the schools can determine which children will need remedial or other instruction in order to show satisfactory progress on the high stakes state standardized tests. For Illinois the high stakes measure is ISAT, the Illinois Standards Achievement Test ("Student Assessment," 2009). Participation in MAP testing is voluntary for schools districts and requires that districts invest both time and money. Surprisingly, with more than 3000 member districts across the nation, including 3 million plus students taking the tests annually, there is currently no research on the effectiveness of these tests to improve educational achievement on middle school state-level accountability standardized assessments. A complicating factor when considering whether NWEA MAP test participation is effective as a means for increasing achievement is the fidelity of implementation in the various member districts. Does merely administering the computer adaptive assessments induce achievement gains, or, do increases in

student achievement rely on what is done with the information that is reported, the ability of schools to make data-driven decisions.

On the NWEA website, the following statement appears; "Assessment Should Make A Difference" ("Assessment System," 2008c). Ultimately, this study aimed to find out if NWEA MAP did make a difference. Did NWEA MAP test participation improve overall school district performance on the high stakes measure-ISAT?

Purpose

The purpose of this study was to ascertain whether or not a relationship existed between Northwest Evaluation Association's Measures of Academic Progress (NWEA MAP) test participation by middle schools in Illinois, and Illinois Standards Achievement Test (ISAT) growth over time. Growth was defined as the change in the percentage of students achieving the Meets or Exceeds rating on the Reading and Mathematics sections of the ISAT over the years. Secondarily, the study aimed to find out if the NWEA members' self-reported level of data-driven decision making practices was associated with ISAT growth over time.

Research Questions

1. Is there a statistically significant difference in growth on the Reading and Mathematics portions of the ISAT between schools that use NWEA MAP testing and schools that do not?

2. For schools that have participated in NWEA MAP testing, does growth on the ISAT Reading and Mathematics sections correlate with the reported implementation levels of data-driven decision making practices?

Null Hypotheses

$H_0$1: There is not a statistically significant difference in growth on the Reading and Mathematics sections of the ISAT between schools that use NWEA MAP testing and schools that do not.

$H_0$2: For schools that have participated in NWEA MAP testing, growth on the ISAT Reading and Mathematics sections does not correlate with the reported implementation levels of data-driven decision making practices.

Significance

A study on the impact of NWEA MAP testing and data-driven decision making practices at the middle school level was important for several reasons. For one, there was a dearth of research regarding the effectiveness of MAP testing in relationship to student achievement growth over time. The rationale for starting a base of research was energized by the sheer number of districts administering the computer adaptive assessments. Additionally, the study had the potential to illuminate which data-driven decision making practices, if any, correlated with increases in student achievement. The findings could have potentially led to improved school improvement practices at the middle school level.

Delimitations

1. The study was focused specifically on middle schools in the state of Illinois with a grade configuration of 6-8.
2. Reading and Mathematics were the only two subjects measured as potential indicators of increased student achievement.
3. Only those districts that administered NWEA MAP testing for each of the three school years encompassing 2005-2008 were included in the group of interest.
4. The data-driven decision making survey was only sent to school districts that administered NWEA MAP testing for each of the three school years encompassing 2005-2008.

Overview of Chapters

A review of the relevant and related literature on assessment, accountability, NCLB, ISAT, NWEA, and data-driven decision making will be presented in Chapter 2. While the order of the presentation is not necessarily linear in respect to time, the progression will allow the reader to understand how the topics are interconnected, and have led to the existing state of affairs. In Chapter 3, the research methodologies will be outlined. Chapter 4 will consist of a presentation and analysis of the data. Chapter 5 will be comprised of a summary, findings, and conclusions.

CHAPTER 2

REVIEW OF RELATED LITERATURE

Introduction

The purpose of this study was to ascertain whether or not a relationship existed between Northwest Evaluation Association's Measures of Academic Progress (NWEA MAP) test participation by middle schools in Illinois, and Illinois Standards Achievement Test (ISAT) growth over time. Growth was defined as the change in the percentage of students achieving the Meets or Exceeds rating on the Reading and Mathematics sections of the ISAT over the years. Secondarily, the study aimed to find out if the NWEA members' self-reported level of data-driven decision making practices was associated with ISAT growth over time. To provide adequate context to the study, the review of related literature addressed the two branches of assessment that were germane to this research, assessment for accountability, and assessment for school improvement.

Assessment for Accountability

In 1877, teachers and parents joined together to force Samuel King, the first superintendent of Portland Oregon, to resign after he published the district test results in the local newspaper. The published scores included each child's name and school attended, allowing public scrutiny of the teachers associated with each reported score (Tyack, 1974, p.48). The publication of test scores was received coldly by prominent national educational figures.

> Emerson E. White, a noted school superintendent and leader in the National Education Association, complained that test scores "should not be used to compare schools and teachers. A careful observation of this practice for years has convinced me that such comparisons are usually unjust and mischievous" (Tyack, 1974, p.48).

Starting in the mid 1900's, schools were held accountable for scores on standardized tests with ever increasing intensification.

> This began in the 1940s with college admission tests. Next came district-wide standardized tests in the 1950s and 1960s. The 1970s was the decade of the state assessment. In the 1980s and 1990s, we added national and international assessments (Stiggins, 2005a).

Coming nearly full circle in 2002, 125 years after Samuel King resigned, President George W. Bush signed into law legislation that included the requirement for school districts and states to produce annual report cards for public consumption, spelling out the performance of each educational entity on an annual basis. While the reports do not include student names, the public accountability is certainly reminiscent of King's vision.

The current link between assessment and accountability is unmistakable. In fact, "...90% of Congress and more than 85% of parents (Ravitch, 2001) want to know how students are doing in school" (Reeves, 2005, p.40). This desire to hold schools accountable, has catapulted standardized assessments into the forefront of the public's attention.

> The chief reason for what seems to be an explosion of educational testing is that U.S. educational policymakers, bent on making the nation's educators more accountable, want hard evidence regarding how well public schools are performing. These policymakers, and most of our citizens as well, believe that student test performance should be the ultimate yardstick by which we measure a school's effectiveness (Popham, 2003a, p.v).

There is a great deal of discomfort on the part of educational scholars in regards to how singular assessment results are being utilized to make high stakes decisions under the guise of accountability.

> So our investment of billions of dollars over six decades in district, state, national, and international testing for accountability has produced scant evidence that these tests have increased student achievement or provided the motivation to learn. At the same time, we have seen mounting evidence of great harm for some segments of our student population (Stiggins, 2004, p. 23).

Ultimately, the demand for accountability, with the focus on assessment, has likely been a major impediment to schools truly achieving improvement (Black, 2005, p.260).

> As someone who has spent his entire career doing research, writing, and thinking about educational testing and assessment issues, I would like to conclude by summarizing a compelling case showing that the major uses of tests for student and school accountability during the past 50 years have improved education and student learning in dramatic ways. Unfortunately, that is not my conclusion (Linn, 2000, p.14).

The quote, written by Robert Linn, an international expert on standardized testing and accountability, encapsulates the frustration that many educators have experienced with standardized testing and the accountability movement in the United States.

No Child Left Behind

Following 15 years of standards-based reform, No Child Left Behind (NCLB) emerged as the most recent catalyst to test driven accountability (Jennings, 2006, p. 110).

> The already thriving national obsession with educational testing intensified in early 2002, when President George W. Bush signed the No Child Left Behind Act, an enormously significant piece of federal legislation laced with loads of assessment-and-accountability provisions (Popham, 2003a, p.v).

NCLB is the merely the latest iteration of the U.S. Elementary and Secondary Education Act (ESEA). According to Popham, standardized tests have been used to evaluate America's schools since ESEA became law in 1965. "That statute provided for the first major infusion of federal funds into local schools and required educators to produce test-based evidence that ESEA dollars were well spent" (Popham, 2005, p.39). Thus, the tradition of using tests to prove achievement results began. The original ESEA legislation was designed to provide dollars to assist school districts with the education of children living in poverty (Thomas, 2005, p.52). In comparison to the original ESEA legislation, NCLB broadened the focus to include all public schools, and all learners- those with special education needs and those who are English Language Learners (ELL), also, NCLB infused penalties into the equation.

The main thrust of the NCLB Act PL 107-110 is the requirement for all students to be proficient in the areas of Reading and Mathematics by the year 2014 (Cronin, 2007, p.5). As of 2006, every state had instituted the required

yearly testing in grades 3-8, as well as one grade at the high school level, in the areas of Reading and Mathematics (Jennings, 2006, p.111). Schools were also required to administer annual Science assessments once in grades 3-5, 6-9, and 10-12, respectively, starting during the 2007-2008 school year (NCLB, Sec. 1111). Under the legislation, students have experienced a dramatic increase in the number of standardized tests they encounter, now in excess of 45 million each year (Tucker, 2009, p.3). The tests are administered to prove that students in a variety of defined subgroups are making Adequate Yearly Progress (AYP). If a school is unable to meet the defined AYP requirement for two consecutive years, a school improvement process is mandated that includes initiating a two-year turn-around plan, and permitting students to choose a new school to attend within the district. Failing to reach AYP in subsequent years induces intensified consequences (NCLB, Sec. 1116).

Furthermore, the act empowers each state with the ability to customize their requirements. Thus, states each act to establish their own definitions of proficiency, their own tests, and their own cut scores (McCall, 2004, p.3). The effects of state level customization have been dramatic.

> The findings of this inquiry are sobering, indeed alarming. We see, with more precision than previous studies, that "proficiency" varies widely from state to state, with "passing scores" ranging from the 6th percentile to the 77th. We show that, over the past few years, twice as many states have seen their tests become easier in at least two grades as have seen their tests become more difficult (Cronin, 2007, p.3).

The credibility of NCLB's impact on practice is severely diminished by the fact that each state is in charge of defining its own levels of proficiency, based largely

on arbitrary decisions, and states have already shown the propensity to decrease the difficulty level of their tests over time.

Illinois Standards Achievement Test

The trend toward reducing the difficulty of the state imposed standardized tests came to fruition in Illinois, when between the years 2003-2006, the ISAT Reading test became dramatically easier in grades 3 and 8, along with the Math test in grade 8. These declines in difficulty came in lieu of the fact that Illinois had already expected a less demanding level of proficiency than most states (Cronin, 2007, p.79). While "State accountability systems that are based on test data and the No Child Left Behind Act have put educators under great pressure to improve their students' scores on standardized tests" (Boudett, 2005b, p. 700), the same tests fail to provide meaningful information back to schools to promote growth.

> ...Many teachers and principals in these three districts felt that state assessment data were not ideal for analyzing student performance and driving instructional decisions. School staff reported that state assessment data are not timely or adequately aligned with daily instruction to be particularly useful, are limited in subject and content coverage and often in the grade levels tested, and have a significant time lag before results are released (Karr, 2006, p. 515).

While educators want to be accountable for student achievement and learning, there are too many detractors engulfing the current method for demonstrating school effectiveness.

There are a number of arguments that have surfaced against using state level standardized test results as a measure of school effectiveness.

1. The high stakes testing environment has led to a narrowing of the curriculum and resulted in pushing instruction toward lower level cognitive skills (Darling-Hammond, 2004, p.1049).
2. The high stakes environment has increased the number of dropouts and decreased the graduation rate, particularly in regards to minority students (Stiggens, 2005, p.13).
3. The pressure to raise scores has resulted in cases of improper test preparation and test administration (Popham, 2006b, p.124). "An analysis of the entire Chicago data reveals evidence of teacher cheating in more than two hundred classrooms per year, roughly 5 percent of the total" (Dubner, 2005, p.34).
4. Unfortunately, standardized test data has been mishandled and misused for decades. "A physician, John Cannell (1987), forcefully brought to public attention what came to be known as the Lake Wobegon effect (Koretz, 1988), that is, the incredible finding that essentially all states and most districts were reporting that their students were scoring above the national norm" (Linn, 2000, p.7).
5. "...a meaningful amount of what's measured by today's accountability tests is directly attributable not to what the students have been taught in school, but to what those children brought to school because of their families' socioeconomic status or the academic aptitudes they happened to inherit" (Popham, 2006a, p. 330).

6. The tests, which are customarily given only once per year, are incapable of providing teachers with immediate feedback concerning the achievement level of their students. The information delay, with states often taking 6 months or more to report results, does not allow teachers to alter instruction in a responsive manner (Stiggins, 2002, p. 759).

The root of the current assessment dilemma seems to stem from a desire to have standardized assessments serve a dual role. First, to act as a watch guard, effectively protecting the taxpayers' investments in the schools. Second, is the, often latent, expectation that assessments should actually lead to improvements in education (Popham, 2006a, p. 1). The confluence of assessment and accountability has led to misguided test formation and inappropriate comparisons. While the prospect of improving assessment for the purpose of accountability seems adrift, the prospect for improving assessment as a vehicle to improve instruction seems promising. Therefore, the next section will depart from focusing on accountability, acknowledging its role in the evolution of our standardized assessment systems, and focus henceforth on assessment that has the potential to positively impact instruction.

Assessment for School Improvement

Educational assessment experts have shared a number of recommendations to increase the likelihood that assessment can act as a vehicle to improve instruction.

1. Measure progress more frequently, in order to inform instruction (Petersen, 2007, p. 42).

2. Clarify the highest priority content standards, providing a clear assessment description for each, and accurately assess only those identified (Popham, 2003a, p. 145).

3. Provide teachers with a detailed report regarding a student's standard-by-standard mastery (Popham, 2003a, p. 134).

4. Provide assessment results in a timely manner (Stiggins, 2002, p. 759).

5. "Place more emphasis on comparisons of performance from year to year than from school to school. This allows for differences in starting points while maintaining an expectation of improvement for all" (Linn, 2001a, p. 5).

6. While it's potentially impossible to remove all bias, work toward reducing bias particularly in relation to SES (Popham, 2003a, p. 58).

"In other words, most educational assessments really ought to help teachers do a better job of teaching (Popham, 2006a, p.54)."

Northwest Evaluation Association

Over the last few years, a number of school districts across the United States have started to voluntarily utilize, with a financial cost and time commitment, a growth-based test that meets many of six the previously outlined

recommendations.

The Measures of Academic Progress, or MAP, is the state-aligned computer adaptive NWEA test that is administered to over 3 million students annually (NWEA, 2008a).

> Each test contains a balanced sample of questions testing the four to eight primary standards in that state's curriculum. The assessment is designed to be adaptive, meaning that high- and low-performing students will commonly respond to items that are aligned to the state's content standards, but are offered at a level of difficulty that reflects the student's current performance rather than the student's current grade (Cronin, 2007, p.9).

For example, a high performing sixth grade student might receive questions at the ninth-grade level, while a low performing peer might receive questions geared at the fourth-grade level. The test adapts to the ability level of the student based on their ability to answer questions correctly, pulling questions from a pool of 2000 individualized for each state (Cronin, 2007, p.9). After a few incorrect answers, the grade level difficulty of the questions decreases, and conversely, after a few correct answers, the difficulty level ramps up. Throughout the 40-55 question test, the program is attempting to pinpoint the exact grade level equivalency the student is working from. This adaptability deceases the number of questions required to pinpoint a student's ability level, by decreasing the questions that are out of the student's range, on both the high and the low end of the scale. The adaptability also acts to provide a consistently challenging test for each individual student. Many students that normally excel on standardized tests experience questions that they find appropriately challenging throughout the

computerized testing experience. Conversely, students that have little success on grade level standardized tests feel appropriately challenged as the computer decreases the difficulty level to match their level of understanding.

NWEA currently offers MAP testing in the areas of Reading, Mathematics, Language Usage, and Science. The tests are administered 1, 2, 3, or 4 times per year by the 3,000 plus partner school districts. The tests are downloaded to school computers and the numeric results are displayed immediately following completion of each subject area test (Huen, 2006). Within one week of testing, all individual and group reports are available to teachers, principals, and school district administrators.

> MAP reports achievement on a RIT (Rasch Unit) scale, an equal-interval vertical measurement scale that enables educators to measure growth independent of grade level and to evaluate and compare performance data across years. The RIT is infinite, although most students' scores fall between the values of 140 and 300. The scale is equal-interval, meaning the distance between 170 and 182 is the same as the distance between 240 and 252 (Olson, 2007).

Each subject area test consists of several categories. For example, Mathematics test results are reported in the sub categories of Algebraic Functions, Computation, Data Analysis/Statistics/Probability, Geometry, Measurement, Number Sense, and Problem Solving (NWEA, 2006, p.6). The specificity of the results provides diagnostic information for individual students, classes, grade level teams, as well as school-wide data. The immediacy of the results, the ease of the reporting, and the connection to state standards and objectives make the NWEA MAP test an appealing assessment instrument to

school districts. Summarized by Allan Olson, former president of NWEA,

> In short, an assessment built on growth measurement not only assures that educators can appropriately challenge all children and raise student learning, it also can provide the foundation for better decision making at a district level, thereby improving how schools are organized and programs delivered (Olson, 2007).

The potential uses and benefits associated with utilizing NWEA MAP testing include,

> "Identifying the skills and concepts individual students have learned. Diagnose instructional needs. Monitor academic growth over time. Make data driven decisions at the classroom, school and district levels. Place new students into appropriate instructional programs" (NWEA, 2008c).

Though NWEA MAP testing has been administered since the year 2000, there is still a lack of research to substantiate many of the claims that the NWEA website purports. The only identifiable previous study that attempted to measure the effectiveness of NWEA MAP testing on improving student performance on state level standardized tests was a dissertation by Susan E. H. DeLong at Indiana State University. In the 2007 dissertation, DeLong attempted to find a relationship between NWEA MAP test participation and increases on the Indiana State accountability test, the ISTEP+, for elementary students throughout the state. The author was unable to detect a positive correlation between NWEA MAP participation and student achievement at the elementary level in Indiana.

There have also been a few critiques of computer adaptive testing's ability to inform instruction. The first critique concerns the design undergirding

computer adaptive assessments, item-response theory, or IRT.

> 'Item-response theory, or the mechanism used to determine which items are easier or harder, ... assumes there's a universal definition of hard and easy,' he said. In some subjects--reading, for instance--that assumption may hold, he said, but for other subjects--such as high school math, which may combine algebra and geometry questions--that assumption isn't always correct (Ash, 2008, p.3).

The preceding quote is from a 2008 Education Week article, in which Katie Ash quoted associate professor Neal Kingston of the University of Kansas. Kingston also added, "'The adaptive-testing model assumes that everyone has taken [courses] or learned [subjects] in the same way,' which is not always the case, Mr. Kingston said (Ash, 2008, p.3). "

In addition to the concerns expressed over this application of IRT, others express concern over the limited number of testing cycles per year, stating that 2 or 3 tests prove too infrequent to provide meaningful/actionable information to teachers. Overall, even the critics seem to acknowledge some value in the computer adaptive testing model. "It's a pretty nice framework for making certain types of tests for certain purposes," said Mr. Marion, "but the promises--from what I've seen--far exceed the practice" (Ash, 2008, p.3).

Considering the lack of research to date and the critiques outlined, the number of districts that are participating in NWEA MAP testing is somewhat startling. "In states like South Carolina, Indiana and Minnesota, anywhere between 60 and 90 percent of the school districts statewide are using a formative assessment based on growth measures to make informed decisions about each

student's education" (Olson, 2007). Furthermore, the NWEA website reports that over 3100 "partners", usually school districts, utilize MAP testing. A relatively small number of "partners", 70, are from international locations (NWEA, 2008b). Since there are more than 15,000 school districts in the United States, controlling for private schools that are included in the member district number, approximately 15-20 percent of our nation's school districts are currently utilizing NWEA MAP testing.

Data-Driven Decision Making

While the literature concerning data-driven decision making is certainly not all positive, as the 2006 *Education Week* article entitled "Data-Driven to Distraction" illuminates, most of the negativity, when distilled, is centered on standardized summative assessments and NCLB, rather than utilizing data sources to inform instruction. To date, the widespread use of data to inform instruction has been limited in large part due to a lack of understanding and skills in assessment and data analysis. The following critiques shed light on just how deficient most educators are in their understanding of assessment.

> Low human capacity to support data-driven inquiry has frequently been noted as a barrier to effective data use in schools. Supovitz and Klein (2003) were "shocked" by the limited technical capacity of faculty even in schools that had been identified as innovative data users. Just 19 percent of teachers and administrators in those schools felt that they had the skills to manipulate data to answer the questions they were interested in (Karr, 2006, p.500).

> This situation is analogous to asking doctors and nurses to do their jobs without knowing how to interpret their patient's charts. Because health professionals are evaluated according to the longevity and physical well

> being of their patients, you can be certain that those professionals thoroughly understand how to ascertain a patient's vital signs. They're called vital signs because they're vital (Popham, 2006b, p.84)!

While schools have lacked sufficient student data in the past, that is no longer the case. "...many are now snowed under with data. They are data rich but analysis poor" (Thomas, 2006, p.37). The lack of skills concerning assessment and data analysis is prevalent among administrators and teachers.

> School leaders in all three sites were very forthright in their anxieties about using data. Even when they were positively disposed to looking at data as part of their decision making, they expressed insecurity about their skill in gathering, interpreting and making sense of the information about their school. Many of them indicated that they had not had training or experience in research, data collection, data management, or data interpretation (Earl, 2003, p.388).

> Few teachers are prepared to face the challenges of classroom assessment because they have not been given the opportunity to learn to do so. It is currently the case that only about a dozen states explicitly require competence in assessment as a condition to be licensed to teach (Stiggens, 2002, p.762).

The political conditions of a district can also lead to mixed feelings about data use, with teachers at some high performing schools seeing data as empowering, while teachers in low achieving, high poverty, diverse communities can feel devalued and disenfranchised by a focus on data (Karr, 2006, p.499). In addition to the lack of training and political realities of a district,

> Common barriers to transforming data into knowledge in education settings often include poorly designed or nonexistent data systems, disorganized record management; moody gatekeepers-data mavericks-who hold back data to preserve power; or personnel who simply fail to ask the right questions of the available data (Mills, 2006, p.44).

Though obstacles exist, data-driven decision making has been found to be highly effective in the promotion of increased student achievement.

> A study of 32 San Francisco Bay Area K-8 schools released in 2003 by the Bay Area School Reform Collaborative (now Springboard Schools) found that "what matters most [in closing the achievement gap] is how schools use data." In fact, those schools that had accelerated the progress of their low-performing students- who were catching up with high-performing students-were those that regularly captured data for the purpose of improving results (Petersen, 2007, p.37).

"Recent Research suggests that data-based decision making can have a positive impact on student achievement and on other aspects of schooling" (Karr, 2006, p.500). Similarly, In the 2002 Spring issue of the *Journal of Staff Development*, Mike Schmoker states that there is ample evidence to suggest that teachers working in teams will experience results when they:

> Focus substantially-though not exclusively- on assessed standards. Review simple, readily available achievement data to set a limited number of measurable achievement goals in the lowest-scoring subjects or courses and target specific standards where achievement is low within that courses or subject area. Work regularly and collectively to design, adapt, and assess instructional strategies targeted directly at specific standards of low student performance revealed by assessment data... (Schmoker, 2002, p.11).

While many educators are still wary of data, some districts have embraced the use of data in guiding decision making in the areas of curriculum, instruction, and professional development (Zavadsky, 2006, p.32). Ultimately, data should precipitate a conversation about what is working, what is not, and what will change as a result (Petersen, 2007, p.42).

To harness the positive effects of data-driven decision making, "...faculty and doctoral students at the Harvard Graduate School of Education (HGSE) and school leaders from three Boston public schools worked together for two years" to develop a guide to school level data-driven decision making practices (Boudett, 2005a). In the end, the group published *Data Wise, A Step by Step Guide to Using Assessment Results to Improve Teaching and Learning. Data Wise* encapsulates many of the data-driven decision making recommendations put forth by leading scholars. Understanding that "small amounts of time and one-shot workshops will not suffice" (Boudet, 2005a, p.134), the authors have offered a series of steps that are designed to cultivate a team of data competent individuals within a school community for the ultimate purpose of utilizing data to improve teaching and learning. The book, released in 2005, outlines an 8-step data process for schools to undertake. The steps are:

1. Organize for Collaborative Work
2. Build Assessment Literacy
3. Create Data Overview
4. Dig into Student Data
5. Examine Instruction
6. Develop an Action Plan
7. Plan to Assess Progress
8. Act and Assess

Conclusion

The brief history of assessment, as presented in this piece, shows struggle. The current age of accountability, propelled by NCLB, has shifted a great deal of attention to high stakes standardized test scores. There have been many positive effects of the NCLB legislation, such as "schools are paying much more attention to the alignment of curriculum and instruction and are analyzing test score data much more closely (Jennings, 2006, p.111), and "schools are paying much more attention to achievement gaps and the learning needs of particular groups of students" (Jennings, 2006, p.111). However, the prior reports of increased student achievement following the launch of NCLB have been tempered as researchers have noticed a decline in the difficulty level of state tests (Cronin, 2007, p. 4).

Educators only recently, during the last decade or two, started to broadly think about assessment as a catalyst or mechanism for improving instruction and achievement (Popham, 2008, p.4). NWEA has enjoyed an increased footprint throughout the United States, and even abroad, with over 3 million students annually taking the computer adaptive tests. Though the district partnerships have surpassed 3,000, there is currently no research evidence that these types of tests are educationally beneficial (Popham, 2008, p.10). While the information that is provided by NWEA is much more detailed and timely than traditional standardized assessments, ultimately, "A process of human interpretation and creating meaning has to happen to change data into information and ultimately into workable knowledge" (Earl, 2003, p.389). The act of testing students and

simply expecting achievement gains is analogous to putting an obese person on the scale and expecting weight loss (Quindlin, 2005). NWEA MAP testing provides rich and timely data, and data-driven decision making has been linked to producing gains in achievement, but does the former necessarily lead to the latter? The lack of assessment literacy and data analysis skills within the broader educational community has potentially compromised the promise of using this assessment, NWEA MAP, as an assessment for school improvement.

CHAPTER 3

RESEARCH DESIGN AND METHODOLOGY

Introduction

The purpose of this study was to ascertain whether or not a relationship existed between Northwest Evaluation Association's Measures of Academic Progress (NWEA MAP) test participation by middle schools in Illinois, and Illinois Standards Achievement Test (ISAT) growth over time. Growth was defined as the change in the percentage of students achieving the Meets or Exceeds rating on the Reading and Mathematics sections of the ISAT over the years. Secondarily, the study aimed to find out if the NWEA members' self-reported level of data-driven decision making practices was associated with ISAT growth over time.

1. Is there a statistically significant difference in growth on the Reading and Mathematics portions of the ISAT between schools that use NWEA MAP testing and schools that do not?

The second phase of the study involved survey research aimed at answering the second research question.

2. For schools that have participated in NWEA MAP testing, does growth on the ISAT Reading and Mathematics sections correlate with the reported implementation levels of data-driven decision making practices?

The sections that follow will outline the research design, provide descriptions of the population and samples, define the instrumentation and data collection procedures, overview the methods of data analysis, and report on the limitations of this study.

Research Design

Correlational research methodology was employed to identify if an association existed between NWEA MAP test participation and growth over time on the ISAT Reading and Mathematics sections, and subsequently between reported data-driven decision making practices and ISAT growth for the two academic subjects. This methodology was selected so that an understanding of the efficacy of current practices in middle school assessment could be determined. The advantage of using correlational research methodology was that the existence or absence of a relationship among the various variables should have been detectable based on easily accessible archival data sets. However, by using this method, it was understood that direct cause and effect would not be proven, only that an association may be present (Kachigan, 1991, p.118).

Population/Sample

The population for this study consisted of public middle schools in the state of Illinois. Middle schools were defined as schools serving grades 6-8, though 6-8 grade schools that housed Preschool and/or Kindergarten programs were also included in the population. Based on information obtained from the Illinois State Board of Education website, 398 schools met the criteria for membership in the population. For the first phase of the research, schools were

selected from the population and placed into one of two categorical groups. To facilitate in the group selection process, NWEA was contacted for a list of school districts that utilized MAP testing during the 2005-2006, 2006-2007, and/or 2007-2008 school years. The number of districts utilizing Map testing in Illinois grew from 84 during the 2005-2006 school year, to 117 in 2006-2007, to a total of 154 in 2007-2008. The three lists were compared to distill which school districts had participated in MAP testing for all three school years. The list of 76 school districts identified as 3-year participants in MAP testing was then cross-referenced with the list of 398 middle schools that met the criteria for inclusion in the population, to arrive at a total of 43 schools that would ultimately comprise Group A.

Group A was comprised of middle schools that have administered NWEA MAP testing for at least three consecutive years, including 2007-2008. This group was considered an in-tact group, meaning that the sample was not chosen randomly, but instead for the purpose of their participation on the dimension of interest (Kachigan, 1991, p. 213).

Group B was defined as middle schools that did not administer NWEA MAP testing during any of the three school years spanning from 2005 to 2008. Following the removal of Group A schools from the study population, along with schools that administered MAP testing for only a portion of the previous three year period, the formation of Group B began. The remaining schools were resorted, and numbered sequentially starting with 001. The table of random digits from Kachigan's Second Edition of Multivariate Statistical Analysis was

employed to form Group B. Group B was comprised of an equal number of schools as Group A, 43 schools respectively, chosen by use of random sampling procedures.

Materials/Instrumentation

The first phase of the study utilized information reported on each school's Illinois School Report Card in 2006 and 2008. In addition to the percentages of students that Meet and Exceed Standards in Reading and Mathematics at each individual grade level 6-8, Low-Income Rate, School and District Total Enrollment numbers, and the Equalized Assessed Valuation per Pupil were captured from each school's annual Illinois School Report Card.

Phase 2 of the research also incorporated the information from each school's state report card. Additionally, Phase 2 involved the use of an original survey. The purpose of the survey was to measure the prevalence of data-driven decision making practices within the middle schools being studied, and more importantly, to produce data that would help ascertain if the reported practices related to growth on the ISAT Reading and Mathematics sections over the last three years. The data-driven decision making survey, hereafter referred to as the DDDM survey, located in the appendix, utilized a Likert Scale and was administered via a combination of email and paper mailings. Snap Survey Software was deployed to create and administer the electronic version of the survey, and a paper copy of the survey was mailed to principals that did not respond to either of two email prompts. The DDDM survey questions were created based on the major recommendations that are found in the book *Data*

Wise (Boudett, 2005a). For each of the eight major *Data Wise* recommendations, one question was formed to measure the level of implementation for each school. To increase the likelihood of garnering participation, the survey was designed to take less than five minutes, with a total of eight questions. The DDDM survey was field tested by sending it for review by three current principals. The principals reviewed it for clarity of instructions and questions, difficulty, and length, i.e. time to complete. The reviewers each sent input electronically, indicating that the survey took less than 5 minutes to complete, and that the format and content proved easy to understand. Two of the reviewers commented on the benefit of having the key indicators following each question.

Data Collection Procedures

Phase 1- The Illinois State Board of Education website, www.isbe.net, was accessed to download each selected school's Illinois School Report Card. Data was gathered from each of the Illinois School Report Cards for the years 2006 and 2008. Report card information on the percentage of students meeting or exceeding Standards in Reading and Mathematics at each individual grade level 6-8, Low-Income Rate, School and District Total Enrollment numbers, and the Equalized Assessed Valuation per Pupil was harvested in March of 2009 and compiled into a Statistical Package for the Social Sciences (SPSS) file for analysis.

Phase 2- The survey was administered through a three-step process. Step one included sending an email letter (Copy in the appendix) containing the

survey link to all principals in Group A. After one week, an email reminder and duplicate letter was sent to all Group A members that failed to complete the survey following the initial prompt. After one additional week, a paper version of the letter and survey was sent to all members of Group A that had not responded to either email prompt. The administration of the survey took place in March 2009, a historically slower month for school administrators, to maximize the survey return rate (Roberts, 2004, p. 142).

Data Analysis Plan

The Statistical Package for the Social Sciences (SPSS), a statistical software application, was utilized to conduct an analysis of the data for both phases of the research.

Phase 1- ISAT data was gathered from Group A and B's Illinois School Report Cards from 2006 and 2008. The Reading and Mathematics Meets or Exceeds Standards percentage for each grade level, 6-8, was documented for each of the two reporting years. Additionally, four potential covariates: Low-Income Rate, School Total Enrollment, District Total Enrollment, and Equalized Assessed Valuation per Pupil were recorded from each of the school report cards.

The ISAT scores were reported as the percentage of students in each school that were able to Meet or Exceed state standards for each of the two subject areas. Growth was defined as the difference in ISAT Reading and Mathematics scores from 2006 to 2008. The 2006 scores acted as the pretest measure, and the 2008 scores were installed as the posttest measure. The

growth was reported to the hundredths position, as either positive or negative.

The years 2006 and 2008 were selected because of the availability of relevant data. The ISAT is comprised of a variety of subject area tests at various grade levels. The subject areas tested at each grade level have changed numerous times over the last few years, with constant data at each grade level in Reading and Mathematics only dating back to 2006. Additionally, schools do receive an overall school percentage that encompasses all of the given tests for that year. However, since the subject areas and number of tests administered at each grade level have shifted, comparing the overall school scores across the years would be of limited benefit. Reading and Mathematics scores have been analyzed separately, with each set of scores acting as dependent variables in Phase 1 of the research.

The independent variable in Phase 1 was NWEA MAP membership status. Membership status, in this study, was a dichotomous categorical variable. Schools were coded with a 1 if they had participated in MAP testing for the previous three years, including the 2007-2008 school year, and a 0 if the school had not administered MAP testing during the three year period. There were two dependent variables in Phase 1 of the study. The dependent variables were growth over time on the ISAT Reading and Mathematics sections, individually.

The first stage in the analysis included reviewing the descriptive statistics for the continuous variables pertinent to the study. Subsequently, a series of independent samples *t*-tests were run to identify the existence of relationships

between NWEA group membership and the demographic variables.

Next, independent samples *t*-tests were employed to test whether NWEA group membership was associated with growth on the Reading and Mathematics sections of the ISAT. Based on the group differences that existed between NWEA and non-NWEA schools, an evaluation of potential covariates was conducted using Pearson's correlation coefficient. The analysis of covariance, ANCOVA, statistical procedure was employed to further analyze the data from Phase 1. ANCOVA was selected due the procedure's ability to utilize pretest scores as covariates to adjust posttest means (Elliot, 2007, p.188). Adjusting the post-test means in this manner had the effect of showing growth from 2006 to 2008. The decision to utilize the ANCOVA procedure was also based in part on hoping to build upon the aforementioned dissertation by Susan E. H. DeLong. Delong (2007) utilized an ANCOVA in her research on MAP testing but did not add a pretest covariate.

Phase 2- Phase 2 sought to find the relationship between data-driven decision-making indicators on the DDDM survey, with growth over time on the ISAT Reading and Mathematics sections. In addition to utilizing the ISAT data harvested for Phase 1, Phase 2 employed the use of the DDDM survey. The DDDM survey was completed by 19 middle school principals in the electronic format, which allowed for a simple data transfer into SPSS. Twelve additional middle school principals submitted their completed surveys by mail.

The first stage in the analysis of Phase 2 included reviewing the descriptive statistics for the continuous variables pertinent to the study.

Subsequently, a series of independent samples *t*-tests were run to compare the mean scores of survey participants and non-survey participants to make sure that significant differences did not exist between the schools that elected to complete the survey and those that did not. The next step was to present the survey statistics for each individual question. The final stage of the analysis for Phase 2 included an evaluation of the linear relationship between each of the eight survey questions and the ISAT growth scores for Reading and Mathematics, measured using Pearson's correlation. While a multivariate multiple linear regression procedure seemed fitting based on the number and types of variables involved, the potential for multicollinearity was simply too great considering the highly correlated nature of the survey items.

The independent variable in each correlation was the reported level of implementation for each survey topic. The dependent variable in each correlation was either the Reading or Mathematics growth over time for each school.

Limitations

The limitations of this study included:

1. School improvement processes that fell outside the scope of NWEA MAP testing and the data-driven decision making (DDDM) survey may have gone undetected and could have inadvertently corrupted the final analysis.

2. The number of testing cycles utilized by NWEA member districts presumably ranged from 1 to 4 per year, though the variation was not considered in the analysis.
3. Focusing on middle schools specifically may have decreased the validity of the findings, considering the wide range of variance in their K-5 feeder schools.
4. The survey on data-driven decision making may have been somewhat subjective in nature, being reported from a single source within each district.
5. This study assumed that the ISAT scores were instructionally sensitive enough to be an accurate gauge of achievement growth over time.

CHAPTER 4:
ANALYSIS OF DATA

Introduction

As outlined in previous chapters, the purpose of this study was to ascertain whether or not a relationship existed between Northwest Evaluation Association's Measures of Academic Progress (NWEA MAP) test participation by middle schools in Illinois, and Illinois Standards Achievement Test (ISAT) growth over time. Growth was defined as the change in the percentage of students achieving the Meets or Exceeds rating on the Reading and Mathematics sections of the ISAT over the years. Secondarily, the study aimed to find out if the NWEA members' self reported level of data-driven decision making practices was associated with ISAT growth over time.

This chapter was organized into two main sections based on the two previously defined research questions.

Phase 1

Question 1

Is there a statistically significant difference in growth on the Reading and Mathematics sections of the ISAT between schools that use NWEA MAP testing and schools that do not?

The first phase of the study utilized information reported on each school's Illinois School Report Card in 2006 and 2008. In addition to the percentages of

students that Meet and Exceed Standards in Reading and Mathematics at each individual grade level 6-8, School Low-Income Rate, School Total Enrollment, District Total Enrollment, and the Equalized Assessed Valuation per Pupil figures were captured from each school's annual Illinois School Report Card.

Demographics

A total of 86 schools were involved in the research associated with question 1. One-half of the middle schools were chosen purposefully based on their involvement with NWEA MAP testing over the previous three school years, including 2007-2008, while the other one-half were chosen randomly from a list of Illinois middle schools that were not associated with NWEA MAP testing during the previous three years. Based on 2008 data, the school enrollments ranged in size from 54 to 2,522 students, while the associated school districts ranged in size from 296 to 18,532. Two key economic indicators were also collected, including the School Low-Income Rate and the Equalized Assessed Valuation per Pupil. The Low-Income Rate ranged from 0.4 to 94.1. Meanwhile, the Equalized Assessed Valuation per Pupil ranged from $22,102.00 to $1,830,941.00.

Hypothesis Testing

$H_0 1$ There is not a statistically significant difference in growth on the Reading and Mathematics sections of the ISAT between schools that use NWEA MAP testing and schools that do not.

Table 4.1

Descriptive Statistics for Phase 1 Categorical Variables- Group Size

Group	N	%
NWEA	43	50
Non-NWEA	43	50

Table 4.2

Descriptive Statistics for Phase 1 Continuous Variables-

Variable	n	*M*	SD
School Enrollment	86	610.94	354.29
District Enrollment	86	4449.92	4552.26
School Low-Income Rate	84	25.94	24.26
Equalized Assessed Value	86	287,248.63	257,467.83
Reading 2006	86	80.45	11.60
Reading 2008	86	83.94	10.32
Mathematics 2006	86	84.91	10.68
Mathematics 2008	86	86.88	9.67

Looking deeper into the eight variables described in Table 4.2, Table 4.3 contains a comparison of the group means, Non-NWEA and NWEA, on each of the eight variables.

Table 4.3

Comparison of Means for Non-NWEA and NWEA Schools-

Variable and membership	n	*M*	SD	Std. Error of Mean
School Enrollment				
Non-NWEA	43	545.05	298.14	45.47
NWEA	43	676.84	395.29	60.28
District Enrollment				
Non-NWEA	43	4830.30	5349.83	815.84
NWEA	43	4069.53	3608.69	550.32
School Low-Income Rate				
Non-NWEA	42	38.61	26.75	4.13
NWEA	42	13.27	12.10	1.87
Equalized Assessed Value				
Non-NWEA	43	185,629.79	276,556.06	42,174.40
NWEA	43	388,867.47	191,128.06	29,146.76
Reading 2006				
Non-NWEA	43	74.39	12.43	1.90
NWEA	43	86.50	6.56	1.00
Reading 2008				
Non-NWEA	43	78.46	10.73	1.64
NWEA	43	89.42	6.24	0.95

Mathematics 2006				
Non-NWEA	43	79.50	12.11	1.85
NWEA	43	90.32	4.90	0.75
Mathematics 2008				
Non-NWEA	43	82.48	11.12	1.70
NWEA	43	91.27	5.10	0.78

The two economic indicators and the ISAT results appear to vary substantially between the two groups. Four independent samples t-tests were employed to further examine the demographic variables in relationship to group membership.

Table 4.4

Independent Samples t-Test Comparing Non-NWEA and NWEA Groups on Demographic Data-

Variable	Non-NWEA M	NWEA M	t	df	Sig. (2-Tailed)
School Low-Income Rate	38.61	13.27	5.594	57.093	.000
School Enrollment	545.05	678.84	-1.745	78.102	.085
District Enrollment	4830.30	4069.53	.773	73.665	.442
Equalized Assessed Valuation per Pupil	185,629.79	388,867.47	-3.964	74.668	.000

There is not evidence to substantiate the existence of a difference between Non-NWEA and NWEA group means on the measures of School Enrollment, t (78.10) = -1.75, p = 0.09, and District Enrollment, t (73.67) = 0.73, p = 0.44. There is however, evidence that a difference may exist between Non-NWEA and NWEA group means on the measures of school Low-Income Rate, t (57.09) = 5.59, p = 0.00, and Equalized Assessed Valuation per Pupil, t (74.67)= -3.96, p = 0.00.

To start testing the null hypothesis directly, H_0, an independent samples t-test was employed to compare Non-NWEA and NWEA schools on Reading and Mathematics growth between 2006 and 2008.

Table 4.5

Independent Samples t-Test Comparing Non-NWEA and NWEA Group Growth-

Variable	Non-NWEA M	NWEA M	t	df	Sig. (2-Tailed)
Reading Growth	4.07	2.92	-1.20	65.47	.234
Mathematics Growth	2.98	0.95	-2.35	68.25	.022

The growth of Mathematics scores from 2006 to 2008 for Non-NWEA schools (M = 2.98 %, SD = 4.87, N = 43) was significantly higher than that of NWEA schools (M = 0.95 %, SD = 2.89, N = 43), t (68) = -2.35, p = 0.022. The growth of Reading scores from 2006 to 2008 did not illuminate a significant difference between Non-NWEA and NWEA schools, t (65) = -1.20, p = 0.234.

Considering the overall group differences present in the ISAT scores between Non-NWEA and NWEA schools, along with the likelihood that demographic differences may be affecting the rate of growth, correlations were run to identify potential covariates, and several analysis of covariance tests were run to further examine question 1.

Table 4.6

Correlation Matrix Between Potential Covariates and ISAT Post-Tests-

Variable		Post-Test Reading	Post-Test Mathematics
School Low-Income Rate			
	n	84	84
	r	-.897**	-.849**
	p-value	.000	.000
School Enrollment			
	n	86	86
	r	.055	-.037
	p-value	.615	.738
District Enrollment			
	n	86	86
	r	-.061	-.111
	p-value	.575	.307
Equalized Assessed Valuation per Pupil			
	n	86	86
	r	.514**	.456**
	p-value	.000	.000

**Correlation is significant at the 0.01 level (2-tailed)

An analysis using Pearson's correlation coefficient indicated that there is a significant correlation between Post-Test Reading scores and School Low-Income Rate $r(82) = -0.90$, $p < 0.001$, as well as between Post-Test Mathematics scores and School Low-Income Rate $r(84) = -0.85$, $p < 0.001$. The analysis also indicated a significant relationship between Post-Test Reading scores and Equalized Assessed Valuation per Pupil, $r(84) = 0.51$, $p < 0.001$, and Post-Test Mathematics scores and Equalized Assessed Valuation per Pupil, $r(84) = 0.46$, $p < 0.001$. No further significant correlations were identified with Reading and Mathematics Post-Tests. Since School Low-Income Rate and Equalized Assessed Valuation per Pupil were also significantly correlated with each other, only School Low-Income Rate will be utilized in subsequent calculations.

A total of four ANCOVA's were run, utilizing a pre-test/post-test format. The 2006 ISAT scores acted as a covariate in each analysis, effectively adjusting the 2008 ISAT group means on the Reading and Mathematics sections as if all schools had started from an identical position. The adjusted 2008 post-test figures represent growth since 2006. The first table contains ANCOVA results which included only one covariate in the analysis, the Pre-Test RE06.

Table 4.7

Tests of Between-Subject Effects for Dependent Variable: Post-Test Reading with Pre-Test RE06 as a Covariate

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7773.980a	2	3886.990	251.842	.000
Intercept	520.229	1	520.229	33.706	.000
NWEA	29.811	1	29.811	1.931	.168
RE06	5191.584	1	5191.584	336.368	.000
Error	1281.041	83	15.434		
Total	615004.451	86			
Corrected Total	9055.021	85			

a. R Squared = .859 (Adjusted R Squared= .855)

The ANCOVA test for equality of means reported in Table 4.7 was not significant (F =1.931 with 1 and 83 *df*, p = .168). The following table contains ANCOVA results which include the School Low-Income Rate as an additional covariate.

Table 4.8

Tests of Between-Subject Effects for Dependent Variable: Post-Test Reading with Pre-Test RE06 and School Low-Income Rate as Covariates

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7468.891a	3	2489.630	187.208	.000
Intercept	529.422	1	529.422	39.810	.000
NWEA	8.348	1	8.348	.628	.431
RE06	571.581	1	571.581	42.980	.000
SLIR	201.206	1	201.206	15.130	.000
Error	1063.901	80	13.299		
Total	603833.551	84			
Corrected Total	8532.793	83			

a. R Squared = .875 (Adjusted R Squared= .871)

The ANCOVA test for equality of means reported in Table 4.8 was not significant (F =0.628 with 1 and 80 *df*, p = .431). The analysis then shifted to ISAT Mathematics growth. Table 4.9 contains only the Pre-Test MA06 as a covariate.

Table 4.9

Tests of Between-Subject Effects for Dependent Variable: Post-Test Mathematics with Pre-Test MA06 as a Covariate

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6771.062a	2	3385.531	239.719	.000
Intercept	226.847	1	226.847	16.062	.000
NWEA	1.899	1	1.899	.134	.715
MA06	5110.235	1	5110.235	361.840	.000
Error	1172.203	83	14.123		
Total	657035.090	86			
Corrected Total	7943.265	85			

a. R Squared = .852 (Adjusted R Squared= .849)

The ANCOVA test for equality of means reported in Table 4.9 was not significant (F =0.134 with 1 and 83 *df*, p = .715). The final ANCOVA looked at ISAT Mathematics growth, with the added covariate of School Low-Income Rate.

Table 4.10

Tests of Between-Subject Effects for Dependent Variable: Post-Test Mathematics with Pre-Test MA06 and School Low-Income Rate as Covariates

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6597.353a	3	2199.118	181.524	.000
Intercept	107.377	1	107.377	8.863	.004
NWEA	16.031	1	16.031	1.323	.253
MA06	1147.207	1	1147.207	94.695	.000
SLIR	44.774	1	44.774	3.696	.058
Error	969.179	80	12.115		
Total	644438.306	84			
Corrected Total	7566.532	83			

a. R Squared = .852 (Adjusted R Squared= .849)

The ANCOVA test for equality of means reported in Table 4.10 was not significant (F =1.323 with 1 and 80 *df*, p = .253). Overall, the research fails to reject the null hypothesis $H_0$1: There is not a statistically significant difference in growth on the Reading and Mathematics portions of the ISAT between schools that use NWEA MAP testing and schools that do not.

Phase 2

Question 2

For schools that have participated in NWEA MAP testing, does growth on the ISAT Reading and Mathematics sections correlate with the reported implementation levels of data-driven decision making practices?

The second phase of the study was focused solely on Group A, the schools that were utilizing NWEA MAP testing. To delve deeper into understanding the relationship between MAP testing and ISAT scores, a data-driven decision making survey (DDDM) was employed to measure the use of data-driven decision making practices at each school in Group A. A total of 19 electronic surveys were completed, along with 12 paper versions of the survey, for a combined response rate of 31/43, or 72%. The second phase of the study also utilized information reported on each school's report card in 2006 and 2008, as previously reported in Phase 1.

Demographics

The variability of schools in Phase 2 of the research is markedly less than that of Phase 1. A total of 43 schools were involved in the research associated with question two. All 43 of the purposively sampled Illinois middle schools had been utilizing NWEA MAP testing over the previous three school years. Based on 2008 data, the school enrollments ranged in size from 160 to 2,522 students, while the associated school districts ranged in size from 439 to 14,347. Two key economic indicators were also collected, including the School Low-Income Rate

and the Equalized Assessed Valuation per Pupil. The School Low-Income Rate ranged from 0.7 to 61.7. Meanwhile, the Equalized Assessed Valuation per Pupil ranged from $83,121.00 to $851,857.00.

Hypothesis Testing

> $H_0$2 For schools that have participated in NWEA MAP testing, growth on the ISAT Reading and Mathematics sections does not correlate with the reported implementation levels of data-driven decision making practices.

The following table contains descriptive statistics for NWEA schools that participated in the data-driven decision making survey.

Table 4.11

Descriptive Statistics for Data-Driven Decision Making Survey Participants-

Variable	n	*M*	SD	Std. Error of Mean
School Enrollment	31	724.74	423.09	75.99
District Enrollment	31	4284.74	3410.29	612.51
School Low-Income Rate	31	13.33	13.40	2.41
Equalized Assessed Valuation per Pupil	31	393,350.55	182,130.11	32,711.53
Reading Growth	31	2.58	2.14	0.38
Mathematics Growth	31	0.96	2.02	0.36

There were also 12 schools that did not participate in the survey. The following table contains the descriptive statistics for the NWEA schools that chose not to complete a survey.

Table 4.12

Descriptive Statistics for Data-Driven Decision Making Survey Non-Participants-

Variable	n	*M*	SD	Std. Error of Mean
School Enrollment	12	553.08	291.95	84.28
District Enrollment	12	3513.58	4187.67	1208.88
School Low-Income Rate	11	13.08	7.82	2.36
Equalized Assessed Valuation per Pupil	12	377,286.17	220,924.38	63,775.38
Reading Growth	12	3.79	4.62	1.33
Mathematics Growth	12	0.93	4.55	1.31

To ensure that the survey respondents did not differ significantly from non-survey respondents, a series of t-tests were run to compare mean scores for the two groups.

Table 4.13

Comparison of Mean Scores- Data-Driven Decision Making Survey Participants and Survey Non-Participants Independent Samples t-Tests-

Variable	*t*	Survey Participant	Survey Non-Participant	*df*	Sig. (2-tailed)
School Enrollment	1.513	724.74	553.08	29.10	.141
District Enrollment	.569	4284.74	3513.58	16.96	.577
School Low-Income Rate	.074	13.33	13.08	30.61	.941
Equalized Assessed Valuation per Pupil	.224	393,350.55	377,286.17	17.12	.825
Reading Growth	-.868	2.59	3.79	12.87	.402
Mathematics Growth	.021	0.96	0.93	12.71	.983

The t-Tests indicated that there were no significant differences between the descriptive statistics of the survey participants and the non-survey participants. The analysis then shifted toward understanding the results from the data-driven decision making survey. The survey prompted principals to rate their school's

degree of implementation for the 8 data-driven decision making recommendations found in the book *Data Wise* (Boudett, 2005a).

Table 4.14

Data-Driven Decision Making Survey Questions-

Question	Question Content
Number 1	To what degree have you and your staff organized for collaborative data analysis?
Number 2	To what degree have you and your staff built assessment literacy?
Number 3	To what degree have you and your staff created a data overview?
Number 4	To what degree have you and your staff "dug into" the data?
Number 5	To what degree have you and your staff examined instruction?
Number 6	To what degree have you and your staff developed an action plan?
Number 7	To what degree have you and your staff planned to assess progress?
Number 8	To what degree have you and your staff taken action and assessed results?

Appendix C contains the DDDM Survey questions in their entirety, including key indicators for each question.

The survey employed a 7-point Likert scale, ranging from 1 for *Not at all*, to 4 for *Moderate*, to 7 for *Highest*.

Table 4.15

Data-Driven Decision Making Survey Statistics-

Question	n	*M*	SD
Number 1	31	4.48	1.71
Number 2	31	3.52	1.21
Number 3	31	4.65	1.50
Number 4	31	4.68	1.38
Number 5	31	4.61	1.48
Number 6	31	4.52	1.71
Number 7	31	4.42	1.29
Number 8	31	4.35	1.38

M Based on a 7-point Likert scale

To test the null hypothesis, a series of correlations were run to measure each question in relation to the growth in ISAT scores in Reading and Mathematics from 2006 to 2008, for each of the 31 participant schools.

Table 4.16

Correlations Between Data-Driven Decision Making Survey Questions and ISAT Reading and Mathematics Growth Scores-

Question		ISAT Reading Growth 2006-2008	ISAT Mathematics Growth 2006-2008
Number 1			
	r	.049	.019
	p-value	.795	.919
	n	31	31
Number 2			
	r	.274	.310
	p-value	.135	.089
	n	31	31
Number 3			
	r	.063	.017
	p-value	.735	.930

	n	31	31
Number 4			
	r	.233	.171
	p-value	.207	.357
	n	31	31
Number 5			
	r	.253	.232
	p-value	.169	.208
	n	31	31
Number 6			
	r	.022	-.235
	p-value	.905	.203
	n	31	31
Number 7			
	r	.069	-.177
	p-value	.711	.342
	n	31	31
Number 8			
	r	.190	.152
	p-value	.306	.414
	n	31	31

The analysis using Pearson's correlation coefficient indicated that there does not appear to be a statistically significant linear relationship between the individual survey responses and growth from 2006 to 2008 on the Reading and Mathematics sections of the ISAT. Of the 8 survey questions, question 2 pertaining to assessment literacy had the strongest correlations with ISAT growth, though neither was statistically significant. Mathematics Growth, $r(29) = 0.310$, $p = 0.089$. Reading Growth, $r(29) = 0.274$, $p = 0.135$. Due to the highly correlated nature of the 8 survey questions, the individual survey responses could not be combined to conduct further analysis.

The research failed to reject the null hypothesis $H_0 2$: For schools that have participated in NWEA MAP testing, growth on the ISAT Reading and Mathematics sections does not correlate with the reported implementation levels of data-driven decision making practices.

CHAPTER 5:
SUMMARY, FINDINGS, AND CONCLUSIONS

Summary

This chapter contains a brief summary of the study and highlights the conclusions drawn from the data presented in Chapter 4.

Purpose Statement

The purpose of this study was to ascertain whether or not a relationship existed between Northwest Evaluation Association's Measures of Academic Progress (NWEA MAP) test participation by middle schools in Illinois, and Illinois Standards Achievement Test (ISAT) growth over time. Growth was defined as the change in the percentage of students achieving the Meets or Exceeds Standards rating on the Reading and Mathematics sections of the ISAT from 2006-2008. Secondarily, the study aimed to find out if the NWEA members' self reported level of data-driven decision making practices was associated with ISAT growth over time.

Research Questions

1. Is there a statistically significant difference in growth on the Reading and Mathematics sections of the ISAT between schools that use NWEA MAP testing and schools that do not?
2. For schools that have participated in NWEA MAP testing, does growth on the ISAT Reading and Mathematics sections correlate with the

reported implementation levels of data-driven decision making practices?

Methodology Review

Phase 1 of the research involved the comparison of group means between schools identified as NWEA MAP test participants and schools that were not associated with NWEA MAP testing during the three year period spanning from 2005 to 2008. There were two dependent variables studied in Phase 1, growth over time on the ISAT Reading, and growth over time on the ISAT Mathematics sections. The analysis included reviewing the descriptive statistics for the dichotomous groups, running a series of *t* -tests, and ultimately employing an analysis of covariance, ANCOVA, statistical procedure. The ANCOVA procedure utilized pretest scores as covariates to adjust posttest means, effectively showing growth for the 86 schools from 2006 to 2008 on the ISAT Reading and Mathematics tests.

The focus of Phase 2 centered on the 43 middle schools that were identified as NWEA MAP test participants from 2005-2008. Utilizing the ISAT data harvested for Phase 1, Phase 2 added the use a data-driven decision making survey. The research sought to find the relationship between data-driven decision making indicators on the DDDM survey, with growth over time on the ISAT Reading and Mathematics sections. The DDDM survey was completed by 31 of 43 identified middle school principals. The analysis for Phase 2 included reviewing the descriptive statistics for the continuous variables, running a series of independent samples *t*-tests, and concluded with an evaluation of the linear

relationship between each of the eight survey questions and the ISAT growth scores for Reading and Mathematics, measured using Pearson's correlation.

Findings

Ultimately, the quantitative research failed to reject both of the null hypotheses.

$H_0$1: There is not a statistically significant difference in growth on the Reading and Mathematics sections of the ISAT between schools that use NWEA MAP testing and schools that do not.

$H_0$2: For schools that have participated in NWEA MAP testing, growth on the ISAT Reading and Mathematics sections does not correlate with the reported implementation levels of data-driven decision making practices.

While the research did not produce statistically significant findings on the dimensions of interest, there were a few findings of potential importance. An analysis using Pearson's correlation coefficient indicated that there was a significant correlation between Post-Test Reading scores and School Low-Income Rate $r\ (82) = -0.90$, $p < 0.001$, as well as between Post-Test Math scores and School Low-Income Rate $r\ (84) = -0.85$, $p < 0.001$ for the 86 schools in Phase 1. The statistically significant inverse relationship between School Low-Income Rate and ISAT scores appears to support previous research findings and the widely held understanding that a student's economic situation tends to have a linear relationship with his or her academic achievement.

The survey responses provided data of potential importance as well. The survey was completed by 31 of 43 principals included in Phase 2 and employed a 7-point Likert scale, ranging from 1 for *Not at all*, to 4 for *Moderate*, to 7 for *Highest*. Therefore, a score of 4 was the midpoint score, or what may have been perceived as average. Collectively, the principals rated their schools implementation of data-driven decision-making practices above moderate "4" on 7 of the 8 questions. The only prompt to which the group's mean response fell below the moderate level, was question 2 "To what degree have you and your staff built assessment literacy?" The lower rating for question 2 was readily apparent during visual reviews of the completed surveys. In fact, 17 of 31 principals rated question 2 as their school's lowest degree of implementation. This result is of particular importance considering that the focus of question 2 could be seen as a foundational understanding that should undergird the other 7 data-driven decision making implementation recommendations. For example, a school's ability to analyze data as represented by question 4, "To what degree have you and your staff "dug into" the data?" would presumably be hampered by a lack of assessment literacy. Regardless, question 4 had an average response of 4.68 while question 2 had an average score of 3.52.

Conclusions

The failure to reject both of the null hypotheses should not be viewed as an indictment on the practice of schools utilizing NWEA MAP testing. The findings should be merely viewed as adding to the body of research that fails to draw

connections between the use of interim assessments and increased student achievement (Popham, 2008, p.10).

Implications

The amount of money being spent and the time being allocated by schools to partake in NWEA MAP testing is substantial. As such investments are substantial, so too should be the call for further research on the effectiveness of NWEA MAP testing. The failure to draw a connection between NWEA MAP testing and growth on the ISAT is not necessarily a direct weakness of the instrument, but instead could illuminate a weakness in how the MAP testing is jelling with the overall educational program of a school or school district. Additionally, the DDDM survey results appear to warrant an increased focus on assessment literacy in teacher education, administrative certification, and professional development programs for middle school educators in Illinois.

Recommendations for Further Research

Considering the apparent lack of a linear relationship between NWEA MAP testing and school level growth on large-scale accountability tests, focusing future research on small cohorts of schools implementing NWEA MAP testing is recommended. One direction that researchers could take is to conduct an in-depth study of a few schools that have a proven record of rapid growth, in an attempt to distill the specific factors that allow certain NWEA member schools to experience success. A second approach would be to follow a cohort of schools

from the implementation stage with NWEA through the first few years of use, attempting to identify the decisions and actions that lead to measurable growth.

The results from Phase 2 of the research also illuminate potential paths for future researchers. Research could be conducted using the DDDM survey with a variety of populations, not simply those utilizing NWEA MAP testing. Additionally, the group mean response to survey question 2 warrants further examination, particularly due to the question's approximation of a significant correlation with ISAT growth. Beyond further exploration of the relationship between assessment literacy and growth on large scale standardized assessments, deciphering which terms and processes are crucial to an effective understanding of assessment literacy could also prove beneficial to the broader educational community.

Concluding Remarks

The act of implementing NWEA MAP testing gives the appearance of a school or district being dedicated to and focused on school improvement, though unfortunately, the act of testing does not guarantee subsequent data analysis and purposeful decision making. While the research conducted in both phases of this study failed to reject the null hypotheses, the findings should not lead one to the conclusion that the utilization of NWEA MAP testing is decidedly ineffective as an ingredient in the school improvement process. One potential reason for the lack of a linear relationship between MAP testing and growth on state sponsored standardized tests is the current lack of assessment literacy amongst

educators. The following quote from James Popham may seem unduly harsh at first glance, "What most of today's educators know about education assessment would fit comfortably inside a kindergartener's half-filled milk carton" (Popham, 2006b, p. 84). However, when coupled with the self-reported results from question 2 on the DDDM survey, pertaining to assessment literacy, the statement seems appropriately critical of the skill set of contemporary educators. The deficiency in the skill set of educators is not necessarily the fault of individuals, but rather may be emblematic of the situation in which they were formally educated, the era that preceded the current high stakes data driven environment.

REFERENCES

Ash, K. (2008). Adjusting to Test Takers. *Education Week, 28*(13), 19-21.

Black, P., & Wiliam, D. (2005). Lessons from around the World: How Policies, Politics and Cultures Constrain and Afford Assessment Practices. *Curriculum Journal, 16*(2), 249.

Boudett, K. P., City, E. A., Murname, R. J. (2005). *Data Wise A Step-by-Step Guide to Using Assessment Results to Improve Teaching and Learning.* Cambridge, MA: Harvard Education Press.

Boudett, K. P., Murnane, R. J., City, E., & Moody, L. (2005). Teaching Educators: How to Use Student Assessment.

Cronin, J., Dahlin, M., Adkins, D., Kingsbury, G. G., & Northwest Evaluation Association, O. R. (2007). *The Proficiency Illusion*: Thomas B. Fordham Institute.

Darling-Hammond, L. (2004). Standards, Accountability, and School Reform. *Teachers College Record, 106*(6), 1047.

DeLong, S. (2007). A study of the relationship between the utilization of NWEA MAP testing and student achievement. Ph.D. dissertation, Indiana State University, United States, Indiana. Retrieved June 28, 2009, from Dissertations & Theses: A&I. (Publication No. AAT 3286443).

Dubner, S. D. L. a. S. J. (2005). *Freakonomics: A Rogue Economist Explores The Hidden Side Of Everything.* New York, NY: William Morrow.

Earl, L., & Fullan, M. (2003). Using Data in Leadership for Learning. *Cambridge Journal of Education, 33*(3), 383.

Elliot, A. C., & Woodward, W. A. (2007). *Statistical Analysis Quick Reference Guidebook.* Thousand Oaks, CA: Sage Publications.

Huen, C. (2006). Making Online Assessment Work [Electronic Version], *2008*, from http://www.techlearning.com/showArticle.php?articleID=193700236

Illinois State Board of Education. (2009). Student Assessment. Retrieved 4/28/09, 2009, from http://www.isbe.state.il.us/assessment/ISAT.htm

Jennings, J., & Rentner, D. S. (2006). Ten Big Effects of the No Child Left Behind Act on Public Schools. *Phi Delta Kappan, 88*(2), 110.

Kachigan, S. K. (1991). *Multivariate Statistical Analysis* (Second ed.). New York, NY: Radius Press.

Karr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to Promote Data Use for Instructional Improvement: Actions, Outcomes, and Lessons from Three Urban Districts. *American Journal of Education, 112*(4), 496.

Linn, R. (2000). Assessments and Accountability. *Educational Researcher, 29*, 4-16.

Linn, R. L. (2001). Assessments and Accountability. (Condensed Version). *Practical Assessment, Research & Evaluation, 7*(11).

Linn, R. L. (2001). A Century of Standardized Testing: Controversies and Pendulum Swings. *Educational Assessment, 7*(1), 29.

McCall, M. S., Kingsbury, G. G., Olson, A., & Northwest Evaluation Association, O. R. (2004). *Individual Growth and School Success. A Technical Report from the NWEA Growth Research Database*: Northwest Evaluation Association.

Mills, L. (2006). Transforming Data into Knowledge. *Principal Leadership, 7*(2), 44.

No Child Left Behind Act of 2001, 20 U.S.C. §§ 6301 et.seq (2002).

Northwest Evaluation Association. (2006). *Step 3- Climbing the Data Latter*: Northwest Evaluation Association.

Northwest Evaluation Association. (2008). About Us. Retrieved 7/23/08, 2008, from http://nwea.org/about/index.asp

Northwest Evaluation Association. (2008). Members. Retrieved 7/23/08, 2008, from http://nwea.org/about/members.asp

Northwest Evaluation Association. (2008). Assessment System. Retrieved 4/28/09, 2008, from http://www.nwea.org/system.asp

Northwest Evaluation Association. (2008). Home Page. Retrieved 7/27, 2008, from http://nwea.org/index.asp

Olson, A. (2007). Growth Measures for Systemic Change. *School Administrator, 64*(1).

Petersen, J. L. (2007). Learning Facts: The Brave New World of Data-Informed Instruction. *Education Next, 2007*(1), 36.

Picciano, A. G. (2006). *Data-Driven Decision Making For Effective School Leadership*. Upper Saddle River, New Jersey Pearson.

Pink, D. (2006). *A Whole New Mind*. New York, NY: The Berkley Publishing Group.

Popham, J. W. (2003). *Test Better, Teach Better The Instructional Role of Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Popham, W. J. (2003). The Seductive Allure of Data. *Educational Leadership, 60*(5), 48.

Popham, W. J., & George Lucas Educational Foundation, S. R. C. A. (2005). *For Assessment. Edutopia, April/May 2005*: George Lucas Educational Foundation.

Popham, J. W. (2006). *Assessment for Educational Leaders*. Boston, MA: Pearson Education.

Popham, J. W. (2006). Needed: A Dose of Assessment Literacy. *Educational Leadership, 63*(6), 84.

Popham, J. W. (2008). *Transformative Assessment*. Alexandria, VA: ASCD.

Quindlen, A. (2005). Testing: One, Two, Three [Electronic Version]. *Newsweek*. Retrieved 7/28/08, from http://www.newsweek.com/id/50105/page/1

Reeves, D. B. (2005). A Constructive Alternative in a Destructive Debate. *Principal Leadership, 5*(7), 38.

Roberts, C. M. (2004). *The Dissertation Journey*. Thousand Oaks, California: Corwin Press.

Schmoker, M. (2002). Up and Away. *Journal of Staff Development, 23*(2), 10.

Stiggins, R. J. (2002). Assessment Crisis: The Absence of Assessment for Learning. *Phi Delta Kappan, 83*(10), 758.

Stiggins, R. (2004). New Assessment Beliefs for a New School Mission. *Phi Delta Kappan, 86*(1).

Stiggins, R., & Chappuis, J. (2005). Using Student-Involved Classroom Assessment to Close Achievement Gaps. *Theory Into Practice, 44*(1), 11.

Thomas, J. Y., & Brady, K. P. (2005). The Elementary and Secondary Education Act at 40: Equity, Accountability, and the Evolving Federal Role in Public Education. *Review of Research in Education, 29*, 51-68.

Thomas, R. S. (2006). How to Survive Data Overload. *Principal Leadership, 7*(2), 37.

Tucker, B. (2009). *Beyond the Bubble: Technology and the Future of Student Assessment*: Education Sector, (ERIC Document Reproduction Service No. ED504266) Retrieved June 28, 2009, from ERIC database.

Tyack, D. B. (1974). *The One Best System*. Cambridge, MA: Harvard University Press.

Zavadsky, H., & Dolejs, A. (2006). Data: Not Just Another Four-Letter Word. *Principal Leadership, 7*(2), 32.

APPENDIX A

Definition of Terms

Content Standards- "Content standards refer to the knowledge and skills that educators want students to learn" (Popham, 2006a, p.25).

Curriculum- "The ends- that is, the learning objectives sought for students" (Popham 2006a, p. 43).

Data-Driven Decision Making- "The simplest definition of data-driven decision making is the use of data analysis to inform, when determining courses of action involving policies and procedures" (Picciano, 2006, p.6). For the purposes of this study, data-driven decision making is specifically focused on improving instruction and achievement.

Data-Driven Decision Making Survey- Based on the chapter titles and key indicators presented in the book *Data Wise,* the DDDM survey was sent to schools that utilized MAP testing over the school years 2005-2008, to measure if a relationship existed between ISAT growth and self reported data-driven decision making practices.

Data Wise- A data analysis guide developed by faculty and doctoral students at the Harvard Graduate School of Education and administrators in the Boston Public Schools (Boudett, 2005a).

Educational Assessment- "A process by which educators use student's responses to specially created or naturally occurring stimuli in order to make inferences about students' knowledge, skills, or affective status" (Popham, 2006a, p.3).

Formative Assessment- Assessments that occur before or during a dose of education, with the intention of informing instruction.

High Stakes Accountability- "...the use of scores on achievement tests to make decisions that have important consequences for examinees and others, as a primary strategy to promote accountability" (Darling, Hammond,
2004, p. 1048).

ISAT- Illinois Standards Achievement Test, administered yearly to students in grades 3 through 8.

MAP- "NWEA Measures of Academic Progress (MAP) are state-aligned computerized adaptive assessments that provide accurate, useful information about student achievement and growth" (NWEA, 2008c).

Middle School- For the purpose of this study, middle school is defined as a public school having a grade configuration of 6^{th}, 7^{th}, and 8^{th}, regardless of the philosophical underpinnings and structures of the school.

NWEA- "The Northwest Evaluation Association (NWEA) is a national non-profit organization dedicated to helping all children learn. NWEA provides research-based assessments, professional training, and consulting services to improve teaching and learning (NWEA, 2008d)."

Standardized Test- "A standardized test is any assessment device that's administered and scored in a standard, predetermined manner" (Popham, 2003a, p.125).

Summative Assessment- Assessment that is administered following a dose of education.

APPENDIX B

Survey Letter

Dear Principal,

Based on your school's experience with The Northwest Evaluation Association's Measures of Academic Progress, NWEA MAP, assessment and your status as a principal of a grade 6-8 school within the state of Illinois, you are being invited to participate in a research study conducted by Michael J. Donhost, a doctoral candidate at DePaul University.

The broad aim of this study is to explore the relationship between NWEA MAP test participation and growth on the ISAT Math and Reading tests in grades 6-8, over time. The first phase of research is now complete, and consisted of a statistical analysis of the aforementioned variables to determine if a correlation exists.

To examine the relationship further, phase two of the study involves conducting survey research on data-driven decision making practices for each of the NWEA MAP test participant schools. The goal being to ascertain whether reported implementation of data-driven decision making practices coupled with the usage of MAP testing, correlates with ISAT Math and Reading growth over time.

Participation in this survey is entirely voluntary. However, it is the goal that the results obtained on this portion of the research study will generate useful information on NWEA MAP testing and data-driven decision making practices for middle schools in Illinois.

Due to the limited number of middle schools that meet the criteria for inclusion in this study, your participation is highly important to the success of this research. You/your school will not be identified individually in any of the findings.

If you have any questions or concerns, please don't hesitate to contact me. If you have questions about your rights as a research subject, you may contact Susan Loess-Perez, DePaul University's Director of Research Protections at 312-362-7593 or by email at sloesspe@depaul.edu.

The eight question survey was designed to take five minutes or less to complete. Thank you in advance for your time and participation in this research study.

Sincerely,

Michael J. Donhost

michaeldonhost@mac.com

Version 3/23/09

APPENDIX C

Survey

Data-Driven Decision Making

The following questions and indicators were created using ideas presented in the book *Data Wise*

1. **To what degree have you and your staff organized for collaborative data analysis?** (Indicators include – forming a data team, creating a data inventory, deciding on systems for collecting and storing data, creating an inventory of instructional initiatives, meeting regularly, using protocols and lesson planning for meetings, adopting an improvement process.)

1	2	3	4	5	6	7
Not at all			**Moderate**			**Highest**

2. **To what degree have you and your staff built assessment literacy?** (Indicators include knowledge of the following terms and practices- sampling, discrimination, measurement error, reliability, score inflation, norm/criterion/standards referenced tests, developmental scales, cohorts, measuring improvement, and interpreting scores.)

1	2	3	4	5	6	7
Not at all			**Moderate**			**Highest**

3. **To what degree have you and your staff created a data overview?** (Indicators include – deciding on the educational questions to focus on, producing graphic displays, drawing attention to critical comparisons, comparing the performance of groups, displaying performance trends, providing opportunities for teachers to work with the data and ask questions, and allowing teachers to experience and discuss the actual tests.)

1 2 3 4 5 6 7

Not at all **Moderate** **Highest**

4. **To what degree have you and your staff "dug into" the data?** (Indicators include – looking carefully at a single data source, reviewing test item responses to understand student thinking, challenging assumptions, triangulating data sources, developing a shared understanding of the knowledge and skills students need, developing a common language, and identifying the learner-centered problem.)

1 2 3 4 5 6 7

Not at all **Moderate** **Highest**

5. **To what degree have you and your staff examined instruction?** (Indicators include – linking learning and teaching, developing the skill of examining practice, developing a shared understand of effective practices, drawing on internal and external resources, analyzing current practices, identifying resources, articulating the problem of practice.)

1 2 3 4 5 6 7

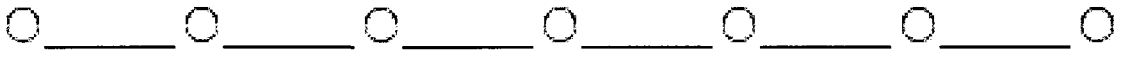
Not at all **Moderate** **Highest**

6. **To what degree have you and your staff developed an action plan?** (Indicators include – deciding on an instructional strategy to address, brainstorming solutions, selecting a solution to implement, developing a common vision for implementation, putting the plan down on paper, assigning responsibilities and time frames, planning for how to support teachers in their new work.)

1 2 3 4 5 6 7

Not at all **Moderate** **Highest**

7. **To what degree have you and your staff planned to assess progress?** (Indicators include – choosing assessments to measure progress, plan to use short-term, medium-term, and long-term data, setting appropriate improvement and proficiency goals.)



1 2 3 4 5 6 7

Not at all **Moderate** **Highest**

8. **To what degree have you and your staff taken action and assessed results?** (Indicators include- communicating the action plan early, integrating the action plan into ongoing school work, using teams for support and internal accountability, visiting classrooms frequently, promoting consistency rather than conformity, adapting professional development plans to meet ongoing needs that emerge from the work, checking in with teachers regularly about learning outcomes, helping teachers see the big picture, honestly evaluating what is working and what is not, celebrating success, revisiting your criteria and raising the bar, keeping the work fresh and ongoing.)

1 2 3 4 5 6 7

Not at all **Moderate** **Highest**

VITA

Michael J. Donhost

Education

Ed.D. **DePaul University**
Doctor of Education in Educational Leadership
July 2009, Chicago, Illinois

M.A. **Michigan State University**
Master of Arts in Educational Administration
August 2000, East Lansing, Michigan

B.S. **Western Michigan University**
Bachelor of Science in Secondary Education
December 1997, Kalamazoo, Michigan

Leadership Experience

Principal- Lake Bluff Middle School
Lake Bluff, Illinois — 7/05-6/08

Associate Principal- Elm Place Middle School
Highland Park, Illinois — 8/02-6/05

Publications

Donhost, M., & Hoover, R. (2007). Creating Change Through Staff Development. *Leadership* (Association of California School Administrators), 37(1). — 9/07

Presentations

ACSA- *Constructing Change*
Presented at the 2007 Association of California School Administrators' Leadership Summit in Santa Clara, California — 11/07

NASSP- *Breaking Ranks and Constructing Change*
Presented at the 2007 National Association of Secondary School Principals' National Convention in Las Vegas, Nevada — 2/07

NSDC- *A Constructivist Approach to Introducing Research and Building Consensus for Change*
Presented at the 2005 National Staff Development Council's Annual Conference in Philadelphia, Pennsylvania — 12/05

NMSA- *Constructivist Staff Development: Moving to the Middle*
Presented at the 2004 National Middle School Association's National Conference in Minneapolis, Minnesota — 11/04