2010

Does Growth Data Make a Difference?: Teacher Decision Making Processes Using Growth Data versus Status Data

Patricia Fox
Virginia Commonwealth University

Does Growth Data Make a Difference?: Teacher Decision Making Processes Using Growth Data versus Status Data

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

by

Patricia Fox
B.A. University of Virginia 1989
M.T. Virginia Commonwealth University 1995

Director: James McMillan, Ph.D.
School of Education

December 2010

Acknowledgment

An African Proverb that adorns my wall states "If you want to travel fast, travel alone. If you want to travel far, travel together." This seems a very apt saying, as I owe many thanks to a wide variety of people who have traveled this path with me.

First, I would like to acknowledge my committee members for their guidance and support. Dr. Jim McMillan, who served as my chair, was especially helpful in keeping me motivated throughout the process. Additionally, his expertise in educational research was invaluable. Drs. Nita Bryant, Kathleen Cauley, and JoLynne DeMary served faithfully on my committee, and their expertise, guidance, and support were much appreciated.

The division's first doctoral cohort also provided support and encouragement. Through coursework, group projects, study sessions, and finally the dissertation process, they were always there with a word of encouragement. I would especially like to thank Dr. Beth Teigen; without her persistence and insistence, this paper would undoubtedly still be "a work in progress".

Finally, my family has always played an essential role in any of my accomplishments, and this is no exception. To Mom and Dad, thank you for always believing in and supporting me.

Table of Contents

List of Tables

List of Figures

Abstract

DOES GROWTH DATA MAKE A DIFFERENCE?: TEACHER DECISION MAKING PROCESSES USING GROWTH DATA VERSUS STATUS DATA

By Patricia Fox, Ph.D.

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

Virginia Commonwealth University, 2010

Director: James McMillan, Ph.D.
School of Education

This experiment examined decisions made by teachers using only status data with those made by teachers using growth and status data. Middle school math teachers from five schools within a single school division located in Virginia participated in the study. Participants were randomly assigned to either the status only or growth and status group. They were then asked to analyze a sample set of class data and complete a survey in which they rated the success of four types of students, identified teacher strengths and weaknesses, and rated their confidence in and the usefulness of the data received. Teachers with access to growth and status data differed significantly in their ratings of three of the four types of students. Students with high growth/low achievement were rated more favorably by teachers with growth and status data ($p < .05$). Students with low growth/high achievement

and those with low growth/low achievement were rated less favorably by teachers with access to growth and status data ($p < .05$). Teachers with access to growth and status data also chose different strengths and weaknesses than those with access to only status data. Teachers did not differ significantly in their confidence in the data or the perceived usefulness of the data, although limitations may have influenced this finding.

This dissertation was created using Microsoft Word 2010.

Chapter 1
Introduction

Accountability in education is not a new phenomenon, but it does have a new face. Since 2001, the mantra of accountability, and legislation regulating accountability, has been No Child Left Behind (NCLB), the common name of the 2001 reauthorization of the Elementary and Secondary Education Act of 1965. It is this law that has guided public schools, school districts, and state boards of education for the last eight years. It is this law that requires schools, districts, and states to meet basic educational standards for all students, documenting the performance of various subgroups through disaggregated data – specifically by subgroups which have typically under-performed their peers in achievement. These subgroups include minorities, the economically disadvantaged, those with special needs, and English language learners.

NCLB includes a number of provisions for achievement testing. Specifically, states are required to test all students in grades three through eight annually in both reading/language arts and mathematics. Additionally, states are required to disaggregate the data in both reading/language arts and mathematics by subgroups. These subgroups include ethnic groups, the economically disadvantages, and students with disabilities. Each state is also required to set Annual Measurable Objectives

(AMO) for Adequate Yearly Progress (AYP) that will ensure that all students are proficient in reading/language arts and mathematics by the 2013-14 school year.

One result of this accountability movement is the amount of data that is available regarding student achievement. If all that was necessary for increased student achievement were the availability of data, or even a superficial analysis of data, the conditions of NCLB would have been met long ago. However, student mastery of basic reading and mathematics skills still lags below 100%, and an examination of subgroups specifically targeted by NCLB shows that these groups are still underachieving compared to their peers (Planty, et al., 2008). Data alone, then, are not the answer.

Statement of Problem

Data-driven decision making is a complex process that requires data to become first information and then knowledge. Statistics show that despite an abundance of data, for many educators the numbers remain simply that – data that is not placed into context to become either information or knowledge that they can use to influence learning. Research also identifies a variety of reasons that this is the case, including access, time, capacity, and trust. Under NCLB, a variety of methods for analyzing school effectiveness have now been approved (Carlson, 2001). These include both status models, successive cohort models, and longitudinal models. The question this research examines is whether change data, and specifically, longitudinal data, are a more effective tool than status data for classroom educators attempting to make data driven decisions. That is, does individual growth data have greater potential to inform and improve teaching, learning, and student achievement than status data?

Overview of the Literature

Framework for Data-Driven Decision Making in Schools

Drawing on research from the field of data-driven decision making in the private sector, several researchers have translated the framework for DDDM to the educational setting. These include Petrides and Guiney (2003), Light, Wexler, and Heinze (2005), and most recently Ikemoto and Marsh (2007). While Petrides and Guiney and Light et al merely adopt the framework, Ikemoto and Marsh, in a paper for the RAND Corporation, validate the framework using data gathered in two previous RAND studies.

The framework for DDDM addresses how to make data useful. Within the framework, data must first become information which can then become knowledge (Ikemoto & Marsh, 2007; Mandinach, Honey, & Light, 2006; Petrides & Guiney, 2003). Data, according to the framework, can exist in any state and may be usable or unusable. In order for data to become information, it must be placed into context by the individual. Information alone, however, does not have implications for future action. In order for decision-making which influences future action to occur, information must become knowledge. Within the framework, only information which is deemed useful can become knowledge that is used to guide action. The issue for educators is that, while there seems to be an abundance of data, there is very little evidence that the data available from standardized testing becomes either information or usable knowledge.

In a recent Occasional Paper for the RAND Corporation, Marsh, Pane, and Hamilton (2006) looked at data-driven decision making in education. Their paper draws on information obtained in four studies conducted by the RAND Corporation over a five year period. Because of this, they were able to examine data use at a variety of levels, including the district, school, and classroom. They identified several areas for future research at that time. Included in these recommendations were "to examine the relative utility of various types of data at all levels of the system…" and to identify "ways to present data and help staff translate different types of data into information that can be readily used for planning and instruction" (Marsh et al, 2006, p. 12).

Data Use in the School Setting

An examination of the use of data in three individual states paints a somewhat rosy picture of data use in the educational setting. Stecher and Hamilton (2006), conducting research on Standards Based Accountability (SBA) as a result of NCLB in California, Georgia, and Pennsylvania, found that a majority of teachers and administrators in those states used state assessment results for improvement purposes including changing instructional practices and identifying areas for professional development. These results do not, however, create a complete picture.

In 2007, the U.S. Department of Education published a report, Teachers' Use of Student Data Systems to Improve Instruction. It was the first of its kind, and a "baseline against which outcomes associated with new federal, state, and district efforts to promote the use of data systems to improve instruction and student achievement can be compared" (Means, et al., 2007, p. 17). Their findings indicate that teachers with access to data varied in their reports of how the data were used, with fewer than 25% of teachers indicating that data were used in identifying skill gaps or promising instructional strategies.

A follow-up study was completed in 2006-07. While more teachers had access to data, the percentage of teachers who used data for specific functions tended to remain constant. While the increase in the availability of data means a net increase in the number of teachers using the data, still less than 50% of all teachers are using data (Gallagher, Means, & Padilla, 2008).

In terms of the framework, while more teachers have access to data, progress still needs to be made in translating those data into information and knowledge on a consistent basis. Given the current availability of testing data and its potential uses, it is important to understand the factors which impact the use of data in the school setting and the relevance of state-level tests – those mandated by NCLB – as tools for teacher use.

Factors Influencing the Use of Data in the School Setting

Beyond access to data, there are a number of other factors which influence data use at the school level. One factor found in the research on data use in schools is time; time to analyze the data in order to use it effectively is a necessary element in its use (Ikemoto & Marsh, 2007; Ingram, Louis, & Schroeder, 2004; Lachat & Smith, 2005; Symonds, 2004). A second factor is the capacity to use data as addressed through supports and professional development (Feldman & Tung, 2001; Ikemoto & Marsh, 2007; Kerr et al, 2006; Symonds, 2004). These may include either district or school level initiatives, and the lack of training is often cited by teachers as a barrier to the use of data (Feldman & Tung, 2001; Kerr et al, 2006; Symonds, 2004).

Another barrier to data use involves trust, both in the process of using data and with regard to the data itself. Over time, teachers in education have come to associate data with negative outcomes or punitive intentions (Bernhardt, 2004; Ingram et al, 2004; Jones, 2007). Additionally, teachers have concerns about the validity and accuracy of student achievement data, limitations of achievement data, challenges in measuring achievement, and the usefulness of the data once collected (Ikemoto & Marsh, 2007; Ingram et al, 2004; Jones, 2007; Kerr et al, 2006; Pedulla et al, 2003).

Models of Student Achievement

In the current era of high-stakes accountability and frequent student testing, a new type of data has emerged: change data. These data exist in many forms but the basic concept is the same: rather than simply indicating where a student is on the achievement continuum, change data represent the growth of a student from a specified time in the past to the current time. Methods for calculating change include pretest-posttest over a single year, vertical scales, and vertically articulated standards.

The pretest-posttest method measures student ability at two distinct times within a given academic year. "The primary object of teaching is to produce learning (that is, change, and the amount and kind of learning that occur can be ascertained only by comparing an individual's or a group's status before the learning period with what it is after the learning period" (Davis, 1964, p. 234). Critical elements in measuring change include tests that measure the material that is taught, that are highly reliable, and that have few floor or ceiling effects (Davis, 1964).

Vertical scales differ from pretest-posttest design in that they measure achievement over time but not within the same academic year. A vertical scale is "a single (unidimensional) scale that summarizes the achievement of students" (Lissitz & Huynh, 2003, p. 3). It is derived from linking assessments from one year to the next based on overlapping curricula. Items from the preceding year's curriculum and the successive year's curriculum are embedded within a given year's test (Schafer, 2006). Performance on these items is used to develop a scaled score which shows not only the proficiency level of a given student, but also how much that student has grown over the course of a year (DePascale, 2006).

Vertically articulated standards, also called vertically moderated standards, are quite different from vertical scales. Rather than comparing scores, they compare proficiency levels as defined by the state (Schaffer, 2006) Huynh and Schneider (2005) identify the two basic elements for vertically moderated standards as common policy definitions and a consistent trend line for performance categories. When constructed correctly, vertically moderated standards enable schools "to predict whether each student is likely to attain the minimum, or proficient, standard…" (Lissitz & Huynh, 2005, p. 5) Vertically moderated standards are particularly useful in subject areas where the content is grade-level specific such as science and social studies (Huynh & Schneider, 2005).

Current models of student achievement can be divided into two categories: status and growth (CPE, 2007a). Status models use a single snapshot of student achievement to make decisions regarding the effectiveness of schools. Growth models use at least two measures of student achievement which are then compared. Growth models can be further divided into successive cohort models and longitudinal models. Successive cohort models include improvement and performance index models and compare the status results of one cohort to those of the successive cohort. Longitudinal cohort models include simple change, value-added, and growth to proficiency models. All three models calculate individual or group growth from one point in time to the next. Of these models, only those that meet the growth to proficiency requirements under NCLB are approved for determining Adequate Yearly Progress.

Research studies show that longitudinal growth models have the ability to present a clearer picture of student achievement and school effectiveness than standard status models or successive cohort models (Heck, 2006; Zvoch & Stevens, 2008). Specifically, longitudinal models can identify schools in which students experience high growth while still remaining below the benchmark and those that meet the benchmark while only exhibiting low or average growth.

Research Questions

1. Do teachers receiving both growth and status data differ in their perception of student success compared to teachers who receive only status data?
2. Do teachers receiving both growth and status data differ in their perceptions of their instructional effectiveness compared to teachers who receive only status data?
3. Do teachers receiving both growth and status data have greater confidence in the data as an accurate representation of student achievement than those who receive only status data?
4. Do teachers receiving both growth and status data have greater confidence in the data as an accurate representation of their instructional effectiveness than those who receive only status data?
5. Do teachers receiving both growth and status data perceive that data as more useful for making decisions regarding individualizing instruction in the classroom than those who receive only status data?
6. Do teachers receiving both growth and status data perceive that data as more useful for guiding their personal professional development than those who receive only status data?

Methodology

True experimental design was used. All middle school math teachers in five of the fourteen middle schools in the district were randomly assigned to either the status (control) group or the growth and status (intervention) group. Data reports and surveys were distributed to participants at department meetings within the schools. Two independent variables were used in the study. The first was the type of data report available to the teacher which had two levels, status only and status and growth. The second independent variable was type of student. This variable had four levels: high achievement-high growth, high achievement-low growth, low achievement,-high growth, and low achievement-low growth. The survey was the only data collection instrument. The survey was designed by the researcher and modified after being reviewed by members of a doctoral cohort and piloted with middle school math teachers in a neighboring school district. Data analysis included frequency distributions and *t*-tests; all data analysis was done by the researcher.

Key Terms

For the purposes of this study, key terms have been defined as follows:

<u>Status Data</u>: data that are the product of single point in time measurements of student achievement, and specifically data from state-mandated standardized testing.

<u>Growth Data</u>: data that are a comparison of student achievement on state-mandated standardized tests over time, requiring scores on at least two tests within a single content area.

<u>Adequate Yearly Progress (AYP)</u>: term used to describe the progress necessary for a school, district, or state to achieve 100% proficiency for all subgroups by the year 2013-14. AYP is calculated each year for each school, district, and state based on AMO’s (see below) set by the state.

<u>Annual Measurable Objective (AMO)</u>: minimum percentage of students, overall and within each subgroup, who must demonstrate proficiency in order for a school, district, or state to make Adequate Yearly Progress (AYP) under NCLB. Percentages are set by the state.

Chapter 2
Review of the Literature

With the advent of No Child Left Behind, data-driven decision making has moved from the private sector to the public schools. Data use in schools however, has not increased as would be expected; there is still far more data available than is used effectively to guide and inform instruction (U.S. Department of Education, 2006). The review of the literature will explore data use in schools as it pertains to student achievement. The four major sections of this chapter are as follows: Framework for Data-Driven Decision Making, Data Use in the School Setting, Factors Influencing Data Use in the School Setting, and Models of Student Achievement. These are followed by a definition of terms used.

Framework for Data Driven Decision Making

A number of researchers have adapted a business model for data-driven decision making for the school setting (Ikemoto & Marsh, 2007; Light, Wexler, & Heinze, 2004; Mandinach, Honey, & Light 2006; Marsh, Pane & Hamilton, 2006). This framework identifies data, information, and knowledge as separate entities, the latter of which should be derived from the first two through various processes which include, collecting, organizing, analyzing, summarizing, synthesizing, and prioritizing. Once these processes have been concluded, actions can be taken, and the cycle then begins again.

The first step of the framework focuses on the data. The processes used in this step are the collection and organization of raw data (Ikemoto & Marsh, 2007; Mandinach, Honey, & Light 2006). These data may take many forms, including input, process, outcome, and satisfaction (Ikemoto & Marsh, 2007; Marsh, Pane & Hamilton, 2006). Additionally, the type of data collected may vary depending on whether it is to be used at the classroom, school, district, state, or even federal level (Ikemoto & Marsh, 2007). "Data exist in a raw state [and] do not have meaning in and of themselves" (Light, Wexler & Heinze, 2004, p. 3). Whether the data are translated into information depends on how and by whom it is used (Light, Wexler, & Heinze, 2004).

The second step of the framework is where information is created from data. Information is defined as "data that is given meaning when connected to a context." (Light, Wexler & Heinze, 2004, p. 3). This happens through two processes: analysis and summarization (Ikemoto & Marsh, 2007; Light, Wexler, & Heinze, 2004; Mandinach, Honey, & Light 2006; Marsh, Pane & Hamilton, 2006). The translation of data into information helps the user to understand his or her environment, but it does not necessarily lead to action (Light, Wexler, & Heinze, 2004).

The third step in the framework is the creation of knowledge. "Knowledge is the collection of information deemed useful, and eventually used to guide action. (Light, Wexler & Heinze, 2004, p. 3). In this step, data users employ the processes of synthesizing and prioritizing the information, creating knowledge which inform the decisions that are made (Ikemoto & Marsh, 2007; Light, Wexler, & Heinze, 2004;

Mandinach, Honey, & Light 2006; Marsh, Pane & Hamilton, 2006). Once an action has been decided upon, the process repeats itself beginning with the collection and organization of data related to the decision that was implemented (Marsh, Pane, & Hamilton, 2006).

Critical in the framework for data-driven decision making is the data itself. As indicated, not all data are used, and not all data are considered valuable. One aspect of data use that will be addressed by this study is whether participants find change data more valuable than status data when making decisions.

Data Use in the School Setting

Although a preponderance of data on students and student achievement exists, as indicated by the U.S. Department of Education studies (2007, 2008), these data are not frequently used to make meaningful changes in instruction which may impact student learning. While the research shows that using data to make changes in instructional practices can increase student achievement (Symonds, 2004) and that some educators are using state assessment data (Brunner et al., 2005; Stecher & Hamilton, 2003); it also reveals that the data, when they are available, are not used by all educators, or even a majority of educators for such purposes (Gallagher, Means, & Padilla, 2008; Ingram, Louis, & Schroeder, 2004; Means, Gallagher, & Padilla, 2007).

Symonds (2004) investigated and compared schools that were successful in closing the achievement gap with those that were not as identified by California's Academic Performance Index (API) ranking system. For the study, four years of API data were examined and thirty-two schools were identified for further investigation.

A gap-closing school was defined as one in which “low-performing students make more rapid progress” than high-performing students while in a non-gap-closing school, the opposite was true (Symonds, 2004, p. 1). Each of the thirty two schools was then surveyed the use of data. Three gap-closing schools were selected for further in-depth case study analysis using interviews, observations, and document review. Additional information was collected from six other schools using teacher and student focus groups.

Findings from the study indicated that schools that were most successful in closing the achievement gap had timely and frequent access to reliable data. Data used in these schools went beyond state-mandated tests to include interim assessments that were used to inform instruction. In gap-closing schools, teachers reported more frequent analysis of data to determine skill gaps: over 67% of teachers in gap-closing schools regularly used the practice compared to less than 25% in non-gap-closing schools. Additionally, on-going assessments were administered more frequently in the gap-closing schools than the non-gap-closing schools with almost all teachers in the gap-closing schools using monthly assessments compared to less than half of the teachers in the non-gap-closing schools.

There is also evidence that state level assessment data can be an informative tool in guiding school improvement at the classroom level. Stecher and Hamilton (2003) studied three states – California, Georgia, and Pennsylvania – and their usage of test-score data at the school and classroom level. Using stratified random sampling, the researchers identified 100 schools located in 25 districts in each of the three states. For each school, surveys were collected from the principal as well as regular education math and science teachers. Case study visits were conducted at two schools in each state.

With regard to annual state assessments, teachers in all three states indicated that the tests and data from the tests were useful in a variety of ways (Stecher & Hamilton, 2003). Teachers reported paying careful attention to the results and using the results for improvement. In each state, a majority of teachers used the results to improve their instructional efforts. This included being more aware of the standards tested and searching for ways to teach more effectively. Using state assessments to focus on standards a moderate amount or great deal was reported by 66%, 72%, and 69% of the middle school teachers in California, Georgia, and Pennsylvania, respectively. Percentages were slightly higher for elementary school teachers. With regard to more effective teaching methods, 58%, 69%, and 59% of teachers in each state reported using annual assessments to guide their search.

State assessments were also reported as being used by teachers to guide professional development and tailor individual instruction, although teachers in Georgia were more likely than teachers in either California or Pennsylvania to use the assessments in this manner (Stecher & Hamilton, 2003). Specifically, 78% of Georgia middle school teachers agreed or strongly agreed that the state tests were useful for tailoring instruction compared to 50% of middle school teachers in Pennsylvania and only 35% of those in California. Similarly, 79% of Georgia middle school teachers found the data from state tests useful for guiding professional learning while only 60% of middle school teachers in Pennsylvania and 55% of those in California reported such usage.

The researchers indicate that while no clear explanation for the differences in data use was explicitly explored in this study, there are several potential reasons. These included that Georgia teachers were more like to report that data were clear and easy to understand (93% vs. 67% and 74%). Additionally, Georgia teachers were more likely to report that the test was a good measure of the content standards as identified by the state.

Brunner et al. (2005) studied the use of specific reports in New York City. The study was done in phases, with the first two phases using interviews and observations to develop a survey used in phase three. Fifteen schools from four districts participated in the study. Prior to the study, The Grow Network was contracted by the city to produce reports for mathematics and language arts teachers in grades four through eight. These reports, called Grow Reports, provided teachers with information on the incoming students. It is important to note, especially given the title of the reports, that these reports were status measures of student ability based on one year of testing. The study examined the use of these reports by teachers and administrators.

This study also found that standardized assessment data is being used by teachers to inform decisions. Of those teachers participating in the survey, 37% reported using the Grow Reports on a monthly basis while 32% reported using the reports between three and six times during the year (Brunner et al. 2005). Of the teachers using the Grow Reports, 91% indicated that the reports were useful for determining strengths and weaknesses of the class as a whole and that they altered their instruction based on this information (Brunner et al. 2005). This was done in a variety of ways. Eighty-nine percent of teachers used the information to set priorities (Brunner et al. 2005). The majority of teachers also used the reports in planning lessons, with 76% indicating the reports were used for general planning, 71% for mini-lessons, and 51% for year-long planning (Brunner et al. 2005). Differentiating instruction was another common practice, with 89% of teachers indicated using the reports in this manner (Brunner et al. 2005). Methods for differentiating instruction included modified lessons, classwork, and homework, and grouping students according to needs for small group or partner work (Brunner et al. 2005).

While these reports indicate that data can and are being used to influence both achievement and instruction, two U.S. Department of Education reports examine the use of data by teachers on a broader scale. These reports use survey data gathered from the National Educational Technology Trends Study (NETTS), which surveyed district technology coordinators and teachers. The original survey was conducted in 2005; the follow up was conducted in 2007. The initial report, *Teachers' Use of Student Data Systems to Improve Instruction*, was the first of its kind, and a "baseline against which outcomes associated with new federal, state, and district efforts to promote the use of data systems to improve instruction and student achievement can be compared" (Means, Gallagher, & Padilla, 2007, p. 17). The findings indicate a discrepancy in the reported availability of data by districts and teachers; while 60% of districts reported giving teachers access to data, only 48% of teachers reported having that access. Among those with access to data, the type of data available varied; while 74% of teachers reported having access to attendance data (the most commonly available), only 39% had access to standardized test scores for their current students. How the data were used also varied according to teacher reports. The most common use of data was to inform parents of student progress with 70% of teachers with access indicating using data this way. With regard to instruction, only 55% of teachers with access to data, or just over 25% of all teachers, reported using data to identify skill gaps while just 41% of teachers with access to data, or less than 20% of all teachers, used the data to identify promising practices for classroom instruction.

The follow-up study was completed in 2006-07. In terms of access to data, more teachers reported having access to an electronic student data system in 2007 (74%) than in 2005 (48%). Attendance data remained the most commonly available, and less than 50% of teachers reported having access to standardized test scores. The percentage of teachers with access who used data for specific functions tended to remain constant. In 2005 and again in 2007, 68% of teachers with access reported using data to inform parents, the most common practice. Sixty-three percent of teachers with access in 2005 and 65% of those with access in 2007 reported monitoring student progress through data. Similarly, 40% of teachers with access in 2005 and 39% in 2007 reported using data to identify promising practices. While the increase in the availability of data means an increase in the percentage of all teachers using data, still less than 50% of all teachers are using data for any given practice other than informing parents, and that stands at just 51% of all teachers (Gallagher, Means, & Padilla, 2008).

In addition to the lack of use of data to inform practice, there is additional evidence that the use of assessment data to make meaningful decisions is scarce. Ingram, Louis, and Schroeder (2004) conducted a qualitative study of data use in urban high schools. Nine high-schools that had been identified as using Continuous Improvement (CI) as a model for change were selected for study. Between 1996 and 1998, researchers conducted interviews to examine individual practices and interviews and focus groups to examine the organizational culture as it relates to data use and data-driven decision making.

In the study, Ingram et al. (2004) examined the implications of teacher decision making in terms of standards and accountability policies. One theme that emerged from their research was “a strong tendency to rely on data that are gathered anecdotally rather than systematically” (Ingram et al, 2004, p. 1270). The use of standardized testing data to evaluate teaching effectiveness was rare, and even when the description of test data was expanded to included teacher assessments, its use was low (Ingram et al, 2004). Thus, although external accountability measures rely primarily, if not solely on standardized tests scores to measure effectiveness, less than 50% of the respondents mentioned any measure of student achievement as a factor in determining teacher effectiveness.

In terms of the framework, while more teachers have access to data, progress still needs to be made in translating those data into information and knowledge.

Given the current availability of testing data and its potential uses, it is important to understand the factors which impact the use of data in the school setting and the relevance of state-level tests – those mandated by NCLB – as tools for teacher use.

Factors Influencing Data Use in the School Setting

Given the availability of data and its potential role in raising student achievement, several researchers have examined the use of data in the school setting. The research in this area identifies a number of factors which either promote or inhibit data use for decision making at the school level.

Access to Useful Data

Timely access to useful data is one factor which impacts data use, as found in a number of studies. Lachat and Smith (2005) and Kerr et al (2006) found that having timely access to data could promote data use. Additionally, both studies highlight the ability to disaggregate the data as an important factor relating to its perceived usefulness.

Lachat and Smith (2005) investigated data use at five high-poverty urban high schools. These schools were selected in part because data use was a core component of their improvement effort. A four-year case study of the five schools was conducted to look for factors which influenced data use. Data were collected through documents, field notes, data archives, and interviews with various school personnel.

Over the course of the study, timely access to data was determined to be a critical element in their use (Lachat & Smith, 2005). Initially, many of the district data-systems contained incomplete or inaccurate data, often due to student mobility and drop-out rates. Once accuracy issues were resolved, school personnel were able to more effectively use the data to examine effectiveness of certain programs and to target instruction.

Lachat and Smith (2005) found that, in addition to timely access to accurate data, the ability to disaggregate data also emerged as a theme of their analysis. Disaggregated data "became more meaningful to school staff and were used more meaningfully in making instructional decisions" (Lachat and Smith, 2005, p. 342). This ability was dependent on district level systems which needed to provide reliable data in a manipulatable format that enabled teachers to analyze the data for patterns and trends. Access to data in a format that could be manipulated increased the chances of staff ownership of data as well as their ability to find meaning in the data and use the data to target individual student and teacher concerns.

Kerr et al (2006), studying data use in three districts, also found the format of the data provided to schools to be important. They examined three urban districts using a mixed methods design. The initial research was a comparative case study using site visits, interviews, and focus groups for data collection. This was followed by a survey instrument distributed to principals and teachers in the three districts. In addition to having multiple data sources, the ability to disaggregate the data was found as an advantage in the two districts that were more successful in using data. The emphasis in the successful districts varied; one district focused on school improvement and staff development while the other focused on the use of interim assessments. In the third district, timely access to data was seen as the primary barrier to data use, as reports had to be requested from the district or outside agencies.

The framework for data-driven decision making suggests that data are used when placed in context. Growth data may have greater potential to become information for the educator because its context is established as change in a given student from one time point to another.

Time

Lack of time to devote to data analysis is another barrier to the use of data in the educational setting. This is acknowledged by many of the leading authors in the field of data use for school improvement, including Bernhardt (2004) and Holcomb (1999). This assertion is supported by more recent field research.

Ingram, Louis, and Schroeder (2004) found time to be an important element related to the use of data in schools. Using data from a longitudinal study of nine high schools with a commitment to continuous improvement, they found that collecting and analyzing data was viewed as competing with other tasks for time in the teachers' schedule. Schools, and the school day, do not provide adequate time for teachers to collect and analyze data for decision making.

Lachat and Smith (2005) went a step further with regard to time. In their study of urban high schools, they found that the need to provide structured time for collaboration was an essential factor for successful use of data. Specifically, they state that "adequate, uninterrupted meeting time" was seen as essential (p. 346).

Incorporating data use in already structured settings such as team meetings, department meetings, and faculty meetings is one way to begin the transition.

Symonds (2004) identified a number of ways in which time could impact the use of data to impact instruction. Studying gap-closing and non-gap-closing schools in the San Francisco Bay area, it was found that gap-closing schools were different from their non-gap-closing schools in several ways related to the use of time. First, in gap-closing schools, time was set aside during the school day to analyze data and plan based on what the data revealed. In addition, time was provided for teachers to collaborate with each other in discussing the data and reflecting on what it revealed about their practices. Finally, the teachers identified time as an important factor in classroom implementation of the practices that had been identified through the data. A school schedule that has been modified to provide this time at both the small group and whole faculty level was seen as an important factor in producing a data driven culture in the school.

Again, growth data have the potential to affect the decision-making framework because it presents the data in a format that is easily understood. This format has the potential to reduce the amount of time needed for data to become actionable knowledge.

Capacity for Data Use

The ability to use available data is another factor which influences the level to which data are used effectively within a school. Feldman and Tung (2001) found that many school level personnel, including teachers and administrators, lacked the expertise necessary to use data effectively. Additionally, they found that while outside support can alleviate some of the problem, there was a perceived need to increase the internal capacity for meaningful use of data.

Kerr et al (2006) also discuss the capacity of a school to use data. Their findings indicate that, while less than 50% of teachers in all three districts felt prepared to use data, those in districts with centralized supports were more likely to feel prepared than those with less centralized supports (36% and 43% compared to 23%). Centralized supports included district-level personnel that provided support by preparing reports and meeting with schools for planning purposes. They highlight district level capacity to support school-level staff as a critical need in promoting data use.

Finally, Symonds (2004) reports on the need for professional development in the area of understanding and applying data. In studying gap-closing versus non-gap closing schools, she found that, in addition to using data more effectively, the teachers at the gap-closing schools related greater frequency and variety of training with regard to data use than the teachers at the non-gap-closing schools. This included professional development in understanding, analyzing and using data as well as linking data and instruction and tailoring instruction based on what the data reveal.

Standards-based tests that are used to report student status use complex processes to ensure that student evaluation from year to year is consistent and fair. These same processes create data that is less than transparent for teachers, complicating the process of using data to make decisions. Growth data derived from these same results may have the potential to be consistent with teachers' previous experience and therefore enhance their capacity for data analysis and decision making.

Confidence in the Data

A final theme that emerges from the literature on data use is the concept of trust. Trust for the purposes of this research is defined as confidence in the data. This is separate from the process of using data, although trust in the process has also been identified as a potential barrier.

In their research on urban high schools, Ingram et al (2004) found two themes that emerged regarding teachers confidence in student achievement data. First, even teachers who mentioned the use of achievement data were likely to mention the limitations of those data in terms of the information that it provided. This was tied to the second finding, that there were measurement challenges associated with achievement tests. These two factors were mentioned by teachers when discussing what data they used and how they used data to determine their own effectiveness. The teachers in the study were more likely to rely on more common assessments, such as teacher-made tests, or anecdotal information to evaluate their effectiveness.

Kerr et al (2006) studied data use at the district level with similar results. In each of the three districts studied, teachers expressed greater confidence in classroom level testing data than either state or local measures. Issues identified by the teachers in the study included that the data received from such assessments were limited in scope and not as useful as other data sources. In their conclusions, they identify the usefulness of data type at different organizational levels as an area of further inquiry that would be beneficial to the field.

Research by Pedulla et al (2003) focused not on how teachers use data, but rather on teacher perceptions of state-mandated tests. This study used an 80 question survey that was taken by teachers in 47 of the 50 states (three states were excluded based on characteristics of their testing program at the time of the survey). While the authors were interested in learning if teacher perceptions varied depending on the stakes of the test (high, medium, or low for schools and students) and the level of the school (elementary, middle, or high), a great deal of similarity was found across all levels for teachers perceptions of the tests' value.

In general, teachers' perceptions of the tests' values were low (Pedulla et al, 2003). For instance, regardless of the stakes attached to the test at either the school or student level or the level of the teacher, less than 20% of teachers believed that the scores on the tests accurately reflected the quality of education that students received. Additionally, less than 20% of teachers in each category reported believing that the test was as accurate a measure of student achievement as their own judgment.

Teachers also questioned the relationship between student characteristics and test scores. When examined by the stakes attached to the test, 80% of teachers in each category believed that the score differences from year to year were reflective of student characteristics rather than school effectiveness (Pedulla et al, 2003). When comparing the same question based on the level of the school, over 75% of teachers in each category believed the scores differed based on school characteristics (Pedulla et al, 2003). When comparing different schools, teachers believed that the differences in scores were related to student characteristics rather than school effectiveness. This was true regardless of the stakes attached to the test, as greater than 75% of each group reported believing this, or the level of the school, as greater than 80% of each group reported believing this (Pedulla et al, 2003).

Another pattern that emerged from the national survey of teacher perceptions was a lack of confidence in the test to accurately represent the abilities of certain subgroups (Pedulla et al, 2003). Teachers were asked specifically about English as a second language (ESL) students and minority students. Regardless of stakes or school level, greater than 90% of the teachers reported that the tests were not able to accurately measure the abilities of ESL students. Similarly, greater than 70% of teachers in each category did not believe the state-mandated tests to be an accurate reflection of the abilities of minority students.

While teachers may be reporting only perceptions when describing their confidence in standardized test data, there is data to support their opinions. Numerous studies of based test scores indicate that student achievement status is highly correlated with non-school factors (Heck, 2006; Zvoch & Stevens, 2006). Some of the reluctance on the part of the teachers to use standardized test data may result from this perception that it does not provide an accurate account of what occurs in the classroom. Growth data, however, are more likely to reflect what has taken place in the classroom over the course of the year, and thus may have greater potential to be seen as valid by teachers because it reflects the influence of the school, and specifically the time frame for which growth is calculated, in addition to the non-school factors that are so prevalent in status scores.

Models of Student Achievement

Achievement is the primary measure of school accountability under NCLB. Measures of student achievement are used for both rewards and sanctions, but many of these measures have come under scrutiny in recent years. According to Goldschmidt and Choi (2007), "NCLB presumes that monitoring the percentage of students who are proficient in reading and mathematics is sufficient to identify schools that are doing a good job and schools that need improvement" (p. 3). Different states, however, use different concepts of quality and progress, and recently different models for ascertaining school effectiveness have been approved under NCLB (Carlson, 2001; CPE, 2007b.) These models can be divided into two categories, status and growth.

Status Models

One model for measuring student achievement is the status model. The Center for Public Education (2007a) defines a status model as a "method for measuring how students perform at one point in time..." (¶ 3). In terms of NCLB, a status model is "a snapshot of a subgroup's or school's level of student proficiency at one point in time" which is then compared with an established target (Goldschmidt et al., 2005).

Status models are designed to answer the question, "on average, how are students performing this year?" (Goldschmidt & Choi, 2007). In order to make what is defined as Adequate Yearly Progress (towards 100% proficiency in 2014), a school must meet the annual measurable objective (AMO) defined by the state. The term adequate yearly progress may be misleading, however, as most models for ascertaining adequate yearly progress use status data and require that students meet proficiency levels in the given year (Barone, 2009; Goldschmidt & Choi, 2007). A status model is a simple model that relies on data from one point in time from which decisions regarding a schools' status in terms of NCLB is determined (Carlson, 2001; Goldschmidt & Choi, 2007).

There are two primary advantages of status models. The first is that they are easy to use (Carlson, 2001; Heck, 2006; Zvoch & Stevens, 2006). The second is that they are easy to understand (Barone, 2009, Carlson, 2007; Heck; Zvoch & Stevens, 2006). Additionally, when states first began reporting student performance, the capacity for tracking longitudinal changes in student performance was different than it is today (Barone, 2009).

Status measures of student achievement are not without their limitations, however. A primary limitation of any status model is that it "does not take into account where each student started at the beginning of the year in assessing performance" (Barone, 2009, p. 2). This is also problematic because assignment to a given school is not a random one, but rather, one influenced by both economic and political processes (Zvoch & Stevens, 2006). The challenge in meeting proficiency is different, and greater, for schools serving students who are disadvantaged and those who start behind their peers (Heck, 2006; Zvoch & Stevens, 2006). Indeed, according to Carlson (2001) the results from a high performing school may be more closely related to the student population than to any processes that are occurring within the school itself.

Another limitation of status models is the possible misclassification of schools. Because status models rely on proficiency as measured at a single point in time, it is possible for a school's population to make substantial progress but still fall short of the benchmark (Barone, 2009; Center for Public Education, 2007b; Heck, 2006). And as indicated before, a high-performing school may get credit for students demonstrating proficiency when there is no indication of the contribution of the school to that measure (Carlson, 2001).

In terms of student achievement, Goldschmidt and Choi (2007) identify four questions that status models fail to answer. Those are:

1. To what extent is previous student performance influencing current performance?
2. What student background factors are influencing achievement?
3. How does current performance relate to achieving the 100% proficiency target?
4. How accurate is this model in identifying school in need of improvement? (p. 4)

Because of the lack of information inherent in status measures, these models are of limited use for making inferences about schools and making decisions about school policy (Carlson, 2001; Goldschmidt & Choi, 2007).

Growth Models

The final type of model is growth models. These models attempt to answer the question, "Is this an effective school?" (Carlson, 2001). A growth model is defined as "a method for measuring the amount of academic progress each student makes between two points in time." (CPE, 2007a, ¶ 4). The Center for Public Education recognizes five types of growth models: improvement, performance index, simple growth, growth to proficiency, and value-added. These five models can be classified into two basic types; models that measure growth by comparing successive cohorts of students, and those that measure individual growth over time, or longitudinal models.

Successive Cohort Models

There are two basic types of successive cohort models; the improvement model and the performance index model. The primary question addressed by these models is "Is [the] achievement level of [the] school improving?" (Carlson, 2001). In order to answer this question, the results of successive cohorts, which are composed of different students, are compared (Barone, 2009; Carlson, 2001; Goldschmidt & Choi, 2006; Goldschmidt et al., 2005).

Both the improvement model and the performance index model are currently used as measures of AYP at the federal level. The improvement model is the "safe harbor" provision (Barone, 2009; Goldschmidt et al., 2007). In order to make safe harbor, a school must decrease its failure rate from one year to the next; however, the determination is still based on reaching proficiency (Barone, 2009).

The performance index model differs from the improvement model in that schools are given credit for students who move upward in proficiency levels even if they do not reach that target score (CPE, 2007b). For example, states would receive partial credit for the percentage of students who moved from the "below basic" into the “basic” category, even though neither of these categories is considered proficient (CPE, 2007b). Performance index models are currently approved for use in at least twelve states (CPE, 2007b).

Improvement models and performance index models, while labeled growth models, are still considered by many to be status models. They are categorized as growth models because they measure proficiency at two points in time, meeting the basic definition of a growth model (Carlson, 2001; Goldschmidt et al., 2007). However, because they rely on proficiency categories in order to determine the effectiveness of a given school rather than measuring student growth they are also described as status models (Barone 2009; Goldschmidt et al. 2007).

As with status models, successive growth models have several advantages in evaluating schools. As indicated, these models include both status and growth, which is seen as an advantage over status models (Carlson, 2001). Additionally, they recognize schools making progress toward proficiency even if they have not reached the established objective (AMO) for a given year (Carlson, 2001). Finally, neither improvement nor performance index models require scaling, either of tests or standards, in order to measure progress (Carlson, 2001).

Successive cohort models are not without their limitations. The primary limitation is the lack of ability to ascribe any meaningful differences to the school context. Carlson (2001) and Goldschmidt et al. (2007) indicate that the observed differences may be due to initial differences in the groups that are being compared rather than being indicative of learning based on instructional programs.

Additionally, student mobility has been shown to have effects even in schools with large, generally stable populations (Carlson, 2001).

Longitudinal Models

The remaining growth models can be described as longitudinal models and attempt to answer the question, "Is this an effective school?" (Carlson, 2001). Each of these models compares the growth of a single cohort of students, either individually or as a group, to their performance at a previous time (Carlson, 2001; Goldschmidt et al., 2009). Students in these models can be either matched or unmatched, depending on the specific model; matched samples are true longitudinal models while unmatched samples are considered quasi-longitudinal (Carlson, 2001). In terms of measurement, while true longitudinal is less noisy, it is also more difficult to compute, and some studies suggest that the results for both types are very similar (Carlson, 2001).

Simple growth models are exactly what the name implies, and they are the simplest of the longitudinal models. In simple growth models, changes in individual student scores are calculated from year to year, and growth scores are averaged to determine the growth within the school (CPE, 2007b). This model is not approved for determining AYP under NCLB as there is no means of determining whether the students will eventually reach proficiency.

Value-added models are the most complex of the growth models. Value-added models are defined as "a method of measuring the degree in which teachers, schools, or educational programs improve student performance." (CPE, 2007a, ¶5). Value-added models examine a student's current performance in terms of both his previous performance and in terms of the expected growth as determined by previous students with similar characteristics (CPE, 2007b). Again, because there is no explicit expectation that proficiency will be reached with value-added models, these models are not approved under NCLB as a model for making AYP

The longitudinal models that are approved under NCLB are growth to proficiency models. Growth to proficiency models use existing data to predict whether a student with a given growth pattern will be proficient at some point in the future, and states are given credit for these students even though they have not reached proficiency at the time tested (Barone, 2009; CPE, 2007b). Beginning with the 2005-06 school year, growth models were approved for calculating AYP, and as of January, 2008, 15 states had growth models that had been approved by the US Department of Education (US DOE Press Release, 2009).

Growth models offer several advantages over the status models (including successive cohort models) currently used by many states to determine AYP. One advantage is that growth models more closely reflect the business of schools, learning, which takes place over time (Heck, 2006; Zvoch & Stevens, 2006). Perhaps because of this, another advantage is that growth models produce a better picture of the effects that school are having. These models are more likely to reveal schools which are successful with challenged populations, providing more accurate, valid information about effective schools and better identification of those in need of improvement (Goldschmidt & Choi, 2006; Goldschmidt et al., 2005; Zvoch & Stevens, 2006). Because these models measure growth, the influence of student background and initial status, both of which have been found to influence current academic achievement, is reduced (CPE, 2007a; Goldschmidt et al., 2005; Zvoch & Stevens, 2006).

The advantage of growth models that is of the greatest interest to this research, however, is the use of growth models to influence student learning. Carlson (2001) states that one use of growth models is understanding student progress, while the Center for Public Education (2007a) states:

> Quite possibly the most effective use of information from growth models is not for high stakes accountability but for such low stakes applications as informing instructional improvement, evaluating the effectiveness of academic programs, and targeting professional development for teachers and administrators (p. 3).

Heck (2006), also cites this advantage, stating that "current pass-fail information provided…is not sufficiently detailed to assess the school's instructional processes in ways that can be used to formulate a comprehensive improvement strategy" and that "if we are to raise the effectiveness of the nation's schools…high-quality information about school processes and outcomes is essential" (p. 671).

Growth models are not without limitations, however. Value-added models and growth to proficiency models require complex statistical calculations that make them less transparent to stakeholders (Barone, 2009). Additionally, measuring growth is noisier than measuring status and requires either equated forms of the same test, vertically scaled tests, or vertically articulated standards, all of which contribute to the decrease in simplicity and transparency (DePascale, 2006).

Studies Comparing Achievement Models

Achievement models serve a variety of purposes. These include school accountability, teacher evaluation, improving practice, and evaluating teacher preparation programs (CPE, 2007a). Different models, however, are better suited to different purposes. Several studies comparing achievement models have been published recently, and the findings are reported here.

The Heck Study

In 2006, Heck compared three models of student achievement to determine their accuracy, equity, and utility. In terms of accuracy, the overlap in identification of schools that were perceived as meeting the benchmarks for NCLB, as defined by the state, and described as having made AYP in the current year was examined. The equity issue concerns one of the main limitations of status models, accounting for factors, such as student background, that are beyond the control of the school. Finally, the concept of usefulness was related to how the information could be employed to examine school effectiveness and potential improvement in effectiveness.

In order to complete the study, it was necessary to have a data set that could be compared as both a successive cohort and a longitudinal cohort. Heck (2006) used data collected over a four year period, 1994-1997, from all 123 comprehensive K-6 schools in the state of Hawaii. Sample sizes ranged from a low of 6,394 students in year one to a high of 6,970 students in year three. In constructing the longitudinal model, only those students who were in the same school all four years were used, resulting in 75% of the population being represented; a possible limitation of the study.

A number of variables were used in the analyses (Heck, 2006). Major variables included were student background, school quality, and school socioeconomic composition. Also examined were school size, teacher experience level, attendance rates, and the percentage of students identified as needing special education services. Student proficiency, defined differently for the three models, was the dependent variable.

Student background was comprised of a number of different variables (Heck, 2006). Specifically, gender, socioeconomic status, and minority status were used. The composition of the population was consistently 49% female. For socioeconomic status, participation in the federal lunch program was used as a proxy, a common practice under NCLB. The percentage of students designated as low SES ranged from 36 to 39 over the four years. For minority status, students of historically underachieving subgroups within the category of Asian or Pacific Island ancestry, which makes up 72% of the overall sample, were used. These were Filipinos, Hawaiians, and Samoans; accounting for between 45 and 49% of the total student population over the course of the study.

School quality was examined using the *Effective Schools Survey*, which is administered in the state on regular cycles. The survey collects information from all certified staff and fifth graders and a randomized sample of parents equal to approximately 20% of the total parent population. The survey measures six items related to school quality: principal leadership, teacher practices in monitoring student progress, school expectations regarding student achievement, emphasis on academics, school climate, and home–school relations. Internal consistency coefficients were reported for each category and ranged from a low of 0.80 for principal leadership to a high of 0.90 for home–school relations.

School socioeconomic composition was a weighted composite score based on census data and participation in the lunch program (Heck, 2006). Information for calculating the scores was obtained from the 1990 census and included percentage living in poverty, percentage receiving public assistance, median income, percentage of high school graduates, and per capita income.

The other independent variables were operationalized as follows. School size was dichotomous, with schools having fewer than 600 students coded as small schools. Teacher experience also had two levels, less than five years experience and five or more years experience. No information was included on how attendance or percentage requiring special education services were operationalized.

Two measures of student proficiency were used. The first was student cut scores. The cut score was established at the 40th percentile of the Stanford Achievement Test Edition 8 using scaled scores for reading, math, and language. The second measure of proficiency was school AYP standards, which varied according to the model of student achievement used. For Model 1, unadjusted proficiency level, and Model 2, adjusted proficiency level, percentage of students scoring at or above the 40th percentile was used. A trajectory was established at 48% for year one, 56% for years two and three, and 62% for year four.

Three models of student achievement were examined in the study (Heck, 2006). Model 1 was an unadjusted proficiency level, or status model. Model 2 was an adjusted proficiency model, also a status model. Scores for this model were adjusted based on within school clustering and within and between school measurement error. Model 3 considered both adjusted proficiency level, as defined in Model 2, and growth. Growth was measured using both initial achievement and rate of change over the course of the four year period.

The study compared the three models in terms of consistency, equity, and utility (Heck, 2006). In terms of consistency, there were no two models that converged on a similar set of schools with the standard set at 85-90% overlap. Values ranged from a low of 66% overlap of Model 2 schools that were identified in Model 3 to a high of 81% overlap for Model 1 schools in Model 2. An alternative comparison, using only year 4 data for model 2 schools, found 83% overlap of Model 3 schools that had met Model 2 standards, still below the minimum of 85% set by the researcher.

Equity was examined in terms of how each model accounted for various factors known to impact student achievement (Heck, 2006). In all three models, between-school comparisons revealed that the majority of differences were explained by school SES. Additionally, school quality has a small but positive effect in all three models. In models 2 and 3, adjusted proficiency level, small, positive effects were associated with being female while small, negative effects were found for minority and low SES students. When measuring growth, minority status and low SES were associated with higher than average growth and steeper growth trajectories. Additionally, there was an interaction between the two, with low SES minority students experiencing the most growth. Part of this may be, as Heck points out, due to regression toward the mean, with those students with the lowest performance measures in the beginning having the most opportunity to show growth.

School context and school processes were found to differentially impact achievement and growth using Model 3 (Heck, 2008). As indicated, there was a dominant relationship between school SES and achievement in all three models; however, there was a much weaker relationship between school SES and growth. School processes, however, had a greater effect on growth than on achievement. These results indicate that "Model 3 increased the equity and validity of comparisons between schools when the focus was on growth instead of on proficiency" (Heck, 2008, p. 687).

The usefulness of each model was also examined. For models 1 and 2, the classification of schools is dichotomous; those that met the benchmark and those that did not. Conversely, Model 3 can be used to identify four categories of schools once a growth standard has been set: those that meet both proficiency and growth standards, those that meet neither standard, those that meet proficiency but not growth standards, and those that meet growth but not proficiency standards. Using the standards of 62% proficiency and top 20% in growth for high growth, Heck (2008) found that 15 schools, or 12%, met both standards while 84 schools, or 68%, met neither standard. These schools would fall into the same identification category in either Model 1 or Model 2. An additional 24 schools, or 19% of the sample, met only one of the standards. Specifically, 14 schools, 11% of the sample, met the proficiency but not the growth standard, indicating that there may be issues with the school processes that need to be addressed if the school is to continue meeting ever increasing proficiency standards. Similarly, 10 schools, 8% of the sample, had high growth while still not meeting the proficiency targets, indicating that while there is still growth needed, their needs may be different than those schools with whom they would have been classified under models 1 and 2.

The Zvoch and Stevens Study.

Zvoch and Stevens (2008) conducted similar research using middle schools in a single district in the southwestern United States. Achievement data for middle school students (grades 6-8) from three cohorts were used in the analysis. For the first cohort, starting sixth grade in 1997-98, only sixth and eighth grade test data were used as the state did not require testing of seventh graders during the 1998-99 school year. For the remaining cohorts, entering sixth grade in 1998-99 and 1999-2000, three years of data were available.

The selection process was similar to, but also differed from, that used by Heck (Zvoch & Stevens, 2008). Both researchers eliminated students who transferred over the course of the data collection period, but while Heck excluded any student who did not remain in the same school for all four years, Zvoch and Stevens only excluded those students who transferred out of the district. Additionally, only one year of data was required for a student in order to be retained in the Zvoch and Stevens study. Subject retention was high for all three cohorts in the study, with cohort 1 having a retention rate of 92% and cohorts 2 and 3 having a retention rate of 89%.

Zvoch and Stevens (2008) used many of the same, but far fewer, variables in their analysis. Both researchers used eligibility for free or reduced lunch as a proxy for socio-economic status (Heck, 2006; Zvoch & Stevens, 2008). Additionally, both considered school size in their analyses; Zvoch and Stevens used three levels of school size, small (<200 students, medium (200-300), and large (>350) (Heck, 2006; Zvoch & Stevens, 2008). Finally, both researchers used outcome data from norm-referenced tests to examine proficiency, with Zvoch and Stevens analyzing data from the TerraNova/CTBS5 Survey Plus. However, while Heck examined reading, mathematics, and language, Zvoch and Stevens used only mathematics achievement data (Heck, 2006; Zvoch & Stevens, 2008). Additionally, Zvoch and Stevens did not use either gender or school quality indices to examine differences in achievement.

Zvoch and Stevens (2008) used three types of models in their study. The first model of student achievement was a standard status model with the benchmark set at the 40^{th} percentile using 6^{th} grade scores (Zvoch & Stevens, 2008). The second model was an improvement model. For this model, change in proficiency was calculated using successive cohorts of 6^{th} graders (Zvoch & Stevens, 2008). The final model used was a longitudinal model. In this case, change scores for each student were calculated using the difference between 8^{th} and 6^{th} grade scores for individual students which were then averaged to determine the school change score (Zvoch & Stevens, 2008). Additionally, between-cohort change scores were calculated by comparing the within cohort difference scores (Zvoch & Stevens, 2008). Data were then used to produce three-level unadjusted longitudinal models: individual student growth trajectories, and within- and between-school variations in status and growth (Zvoch & Stevens, 2008). This three-level model was then repeated in an adjusted model, which incorporated the increasing performance expectation of NCLB.

Several relationships between achievement and growth were reported. In examining differences in student achievement using successive cohort models, Zvoch and Stevens (2008) found negative relationships between initial status of the cohort and cohort gains in achievement. That is, groups with high initial achievement showed low growth and groups with low initial achievement showed high growth (Zvoch & Stevens, 2008). This implies that the variation in achievement and growth, when comparing different cohorts, is due to non-systemic factors such as the composition of the cohort rather than school processes (Zvoch & Stevens, 2008).

Correlations between SES status and scores, proficiency, and growth were also reported. Both mean achievement scores and percentage passing were strongly correlated with SES (-.96 and -.97, respectively), as seen in previous studies (Zvoch & Stevens, 2008). SES was positively, though less strongly, correlated with growth in achievement (.77) and increases in percentage proficient (.48) (Zvoch & Stevens, 2008). This is not surprising given the previously reported negative relationship between achievement and growth.

Much smaller relationships were discovered when examining the relationship between growth and SES. Similar to the relationships Heck (2006) found, Zvoch and Stevens (2008)reported that high levels growth were found in schools with both high and low percentages of students characterized as disadvantaged; this was also true for schools with low growth. They conclude that “Knowing the poverty status of schools thus provided little insight into the rate at which students learned mathematics across cohorts or the change in mathematics growth between cohorts” (Zvoch & Stevens, 2008, p. 587).

Relevance

Heck concludes by stating that, while further research is necessary, “growth models provide a more comprehensive framework for school assessment and a direct means for superintendents and principals to identify student and school needs and engage in planned efforts to strengthen instructional processes” (2006, p. 695). Zvoch and Stevens (2008) conclude that “These results suggest that conclusions regarding the performance and instructional practices of schools could vary widely and/or be misguided depending on the indicators used in a school accountability system” (p. 588). These findings are directly related to the proposed research in that the hypothesis is that teachers with access to growth data will be able to improve instructional processes through better identification of students needs and their own strengths and weaknesses in influencing student learning.

Summary

Recent changes in accountability have resulted in an abundance of data related to student achievement. A framework for data-driven decision making indicates that raw data, or facts, must first become information by being placed into context and then become knowledge when that information is embodied within an individual. While the accountability movement has provided data, research shows that data use within the school setting is limited, especially in terms of guiding student achievement. While data use is evident in a number of individual settings, there is also evidence that this practice is limited. A number of factors which influence data use have also been identified in the research. In general, these are considered barriers to data use and include access to meaningful data, and the time and capacity to turn data into actionable knowledge. Additionally, there is a general distrust of standardized achievement data as an indicator of the quality of education. Rather, teachers view the results of these tests as reflective of student characteristics. This is not unfounded, as a number of studies have demonstrated a clear correlation between socio-economic status and student achievement.

A recent development in accountability has been growth measures. Because of the annual testing that is required under NCLB, many states have begun to track changes in achievement, or growth, in addition to static achievement levels, or status. Beginning in 2005, the United States Department of Education incorporated approved growth models into measures for calculating Adequate Yearly Progress. Additionally, a number of states have begun to evaluate teachers based on the growth of students.

What has not been examined, and what this study begins to examine, is the potential value of growth data to teachers.

Chapter 3
Methodology

The purpose of this study was to examine the potential of growth data for overcoming some of the primary barriers to data use in schools. This chapter is divided into three sections: design, procedures, and data analysis. The design section includes the type of design, a description of the participants and the assignment method, and information on the variables as well as the instrument used to measure the dependent variable. The procedure section discusses how treatment was implemented and the collection of data. The analysis section identifies the methods to be used in answering each of the research questions.

Research Questions

1. Do teachers receiving both growth and status data differ in their perception of student success compared to teachers who receive only status data?
2. Do teachers receiving both growth and status data differ in their perceptions of their instructional effectiveness compared to teachers who receive only status data?

3. Do teachers receiving both growth and status data have greater confidence in the data as an accurate representation of student achievement than those who receive only status data?
4. Do teachers receiving both growth and status data have greater confidence in the data as an accurate representation of their instructional effectiveness than those who receive only status data?
5. Do teachers receiving both growth and status data perceive that data as more useful for making decisions regarding individualizing instruction in the classroom than those who receive only status data?
6. Do teachers receiving both growth and status data perceive that data as more useful for guiding their personal professional development than those who receive only status data?

Research Design

The study design was experimental. True experimental design is rare in education, but this study allowed for its use for a number of reasons. First, the nature of the study and the location in which it was conducted required new information to be provided to participants. Because these participants had not been clustered in any manner, it was possible to completely randomize the assignment of participants to either the intervention or the control group. The random assignment of subjects to either the intervention or control group reduced the potential that selection was a threat to internal validity (McMillan, 2004).

Population and Sampling Procedures

Participants in the study were the middle school math teachers from a large, diversified school district in the southeast region of the country. The school system serves over 60,000 students in sixty-three schools, fourteen of which are middle schools. One of these is an alternative middle school.

A convenience sampling of math teachers in six of the fourteen middle schools in the county was used. The purpose in using only six of the schools was to reduce the burden on the school system. The selected schools were representative of the overall population of the county and chosen by the county. Assignment to either the status only or growth and status data group was random. Demographic information from participants was collected in order to compare the sample to the population.

Measures/Data Sources

Independent Variable

The study had two independent variables. The first independent variable was the type of report received. This variable had two levels: status reports only and growth and status reports. Because the state in which the study was conducted does not currently generate growth reports, these reports were developed by the researcher. The design of both reports was based on those currently available to teachers in the state. For each group of students tested, identified as a single class for the district in which the study was conducted, teachers can receive a summary report. This report includes three pieces of information for each student tested: numeric scaled score for the overall test, proficiency level of the student based on the scaled score, and reporting category scaled scores. Overall scaled scores range from 0 to 600. Students scoring between 0 and 399 are rated as either below basic or basic, both of which are failing scores. Scores between 400 and 600 are passing scores. Students who receive scores between 400 and 499 are rated as proficient while students scoring between 500 and 600 are rated as advanced proficient. Reporting category scores range from 0 to 50. The cut score for being proficient within each reporting category is approximately 30.

The growth reports differed from the status reports only in the addition of growth information. The researcher used information provided by several states in order to guide the development of the growth report. States that responded to the request for information included Alaska, Arizona, Arkansas, Florida, Iowa, and Michigan. Of these states, only Florida and Michigan indicated that data on growth was provided to teachers. Michigan provided a sample report that teachers receive as well as information on interpreting the report (P. Bielawski, personal communication, February 19, 2009). The class roster report used in Michigan was used to create a similar class roster report for this study. The Michigan class roster report contains information on individual students which includes change in achievement level which is categorized as Significant Decline

(SD), Decline (D), No Change (N), Improvement (I), or Significant Improvement (SI) (P. Bielawski, personal communication, February 19, 2009). This Michigan report also includes scores for each student by strand (P. Bielawski, personal communication, February 19, 2009). These aspects of the Michigan report were used to create the sample growth report with modifications to reflect the current reports available in the state in which the research was conducted (See Appendix B).

The second independent variable was type of student. For the growth report, a simple growth model was used and a matrix of possible combinations of status and growth was created. Status was given two levels: high and low. Growth was given three levels: high, expected, and low. Participants were asked about students in four of the six categories. The final matrix shows the four types of students for whom data was collected (see figure 1).

		Growth		
		High	Expected	Low
Status	High	Type 1	-	Type 3
	Low	Type 2	-	Type 4

Figure 1. Matrix of student type for growth data.

Current information available in the state was the basis for the general format of the report. With regard to mathematics, students receive an overall scaled score ranging from 0 to 600 and a scaled score for each of five reporting categories ranging from 0-50. The five reporting categories are consistent between third and eighth grades but differ once students enter Algebra I. For this study, teacher reports included both scaled score and reporting category information. Teachers receiving growth reports received growth data based on overall scaled score and reporting category performance. Each teacher received data pertaining to one class that included class averages and individual student information.

The final report given to teachers included 27 students. Twelve of these students represented the subjects for use in examining the first research question: do teachers receiving student growth and status data differ in their perception of student success compared to teachers who receive only status data? Three students of each type were included in the report. Figure 2 indicates which students were of interest for each type.

Type 1 High Status/ High Growth	Type 2 Low Status/ High Growth	Type 3 High Status/ Low Growth	Type 4 Low Status/ Low Growth
Student 11	Student 02	Student 09	Student 07
Student 14	Student 18	Student 10	Student 12
Student 15	Student 24	Student 25	Student 26

Figure 2. Types of students used as subjects in teacher reports.

An additional fifteen students appeared in each report; these fifteen students were used to create average reporting category scores and frequency distributions for growth to address the second research question: do teachers receiving student growth and status data differ in their perceptions of their own effectiveness compared to teachers who receive only status data?

Included with the reports was information on the data contained in the reports. The information provided varied only in the addition of the description of the growth data for teachers in the intervention group. The informational pages are included in Appendix A and the sample report is included in Appendix B.

Dependent Variable

The dependent variables are the decisions teachers make using the data. The survey was the only data source, and it included demographic information that enabled the researcher to describe the sample and compare the sample to the population. Teachers were asked to respond to a series of questions based on the reports provided to them. For each target student, teachers were asked to rate the success of the student on a six point likert-type scale. Additionally, each teacher was asked to identify which reporting category represents an area of strength for them and which represents an area of potential growth. The final section of the survey asked about the teacher's confidence in the data in terms of student achievement and teacher effectiveness and offered an opportunity for open-ended response regarding the data.

The first draft of the survey was reviewed by members of a doctoral cohort. Following modifications, a pilot of the survey was run using middle school math teachers from outside the district. Appendix C contains the survey instrument.

Data Collection Procedures

Data collection was done through math department meetings at the county and school level. At a county department chair meeting, the researcher first administered the survey to department chairs at participating schools. Once department chairs had completed the questionnaire, the researcher explained the procedure to be used at the individual schools and answered any questions the participants had. These chairs were then provided with the reports and survey instruments they would need to complete the administration at their school. The reports were stacked in an alternating manner, and chairs were asked to distribute them in the order in which they were received. This process was modeled in the division level department chair meeting. Every middle school math teacher in the selected schools who attended the meeting received one of the two data reports and an opportunity to participate in the study. Each department chair was provided with an envelope to collect the surveys which were then be returned through the interoffice mail system. Administration of the survey took place in June, 2010.

Data Analysis

All data analyses were performed by the researcher. After receipt of the surveys, the researcher entered information from each respondent into an SPSS database. In order to examine reliability of the instrument, Cronbach alpha scores were calculated. An alpha score was calculated for each of the four types of students. Alpha scores were also calculated for the questions about confidence in and usefulness of the data in terms of examining or informing practice related to both teachers and students. Two reliability scores were calculated for each question. The first set of reliability scores compared the questions about confidence in the data with each other and the questions about usefulness of the data with each other. The second set compared the questions based on whether they provided information about teacher effectiveness and needs or student success and needs.

Additional information was calculated based on the research questions. Following an outlier analysis, descriptive statistics were run. These included frequency distributions, means, and standard deviations for all items. Inferential statistical analyses were run according to the research questions. For research question 1, perceptions of student success, one score per respondent was calculated for each type of student by averaging their scores for the individual students within each category. The means for each type of student were compared for teachers receiving only status data and those receiving both status and growth data. An alpha level of .05 was used. For research question 2, perceptions of instructional effectiveness, cross-tabs were run for predicted responses for both strengths and weaknesses. Responses from those teachers who did not choose a predicted strength (reporting category 1 and 5) or weakness (reporting category 4 or 3) were not considered in the cross tabs. For perceived strengths, a chi-squared test with Yates correction was run. For perceived weaknesses, no analyses beyond frequency distribution were performed as the value of one cell was zero. For the remaining questions, those regarding confidence in and usefulness of the data, mean scores for each group of teachers were compared using a *t*-test with an alpha level of .05.

Research Hypotheses

Based on the literature review, the following research hypotheses have been formulated. They are presented in order of the research questions proposed.

1. Teachers receiving both growth and status data will differ in their perceptions of student success compared to teachers who receive only status data. Specifically,
    a. Teachers will rate Type 1 students (high status, high growth) similarly.
    b. Teachers receiving both growth and status reports will rate Type 2 students (low status, high growth) more favorably than teachers receiving only status reports.
    c. Teachers receiving both growth and status reports will rate Type 3 students (high status, low growth) less favorably than teachers receiving only status reports.

d. Teachers will rate Type 4 students (low status, low growth) similarly.

2. Teachers receiving both growth and status reports will indicate different strengths and weaknesses than teachers receiving only status data.
3. Teachers receiving both growth and status data will have greater confidence in the data as an accurate representation of student achievement than those who receive only status data.
4. Teachers receiving both growth and status data will have greater confidence in the data as an accurate representation of their instructional effectiveness than those who receive only status data.
5. Teachers receiving both growth and status data will perceive those data as more useful for making decisions regarding individualizing instruction in the classroom than those who receive only status data.
6. Teachers receiving both growth and status data will perceive those data as more useful for guiding their personal professional development than those who receive only status data.

Chapter 4
Results

The focus of this study was teachers' decision making processes using growth and status data. The purpose was to discover if teachers receiving both growth and status data made different decisions than those receiving only status data. Research questions were developed to examine teacher perceptions of student success and instructional effectiveness. Additional questions were developed to examine teachers' confidence in the data and the perceived usefulness of the data. The six research questions were:

1. Do teachers receiving both growth and status data differ in their perception of student success compared to teachers who receive only status data?
2. Do teachers receiving both growth and status data differ in their perceptions of their instructional effectiveness compared to teachers who receive only status data?
3. Do teachers receiving both growth and status data have greater confidence in the data as an accurate representation of student achievement than those who receive only status data?

4. Do teachers receiving both growth and status data have greater confidence in the data as an accurate representation of their instructional effectiveness than those who receive only status data?
5. Do teachers receiving both growth and status data perceive that data as more useful for making decisions regarding individualizing instruction in the classroom than those who receive only status data?
6. Do teachers receiving both growth and status data perceive that data as more useful for guiding their personal professional development than those who receive only status data?

Sample

The sample consisted of math teachers from five of the fourteen middle schools from a large, suburban school division in Virginia. Six schools were purposively identified by the county to participate in the survey. These schools were considered by the county to be representative of their overall population and the programs available. Of the six schools selected to participate; one school did not return any surveys. Demographic information for the population and the sample are presented in Table 1. This table shows that of the 62 middle school math teachers in the five schools, 45 completed at least part of the survey. Within the population, 84% were female and 16% male; within the sample, 78% were female, 16% male, and 7% did not indicate a gender. With regard to ethnicity, the population was 87% Caucasian, 10% African-American, and

3% other; the sample was 73% Caucasian, 9% African-American, and 9% other or unspecified while an additional 9% did not provide an answer. Finally, with regard to level of education attained, 69% of the population held a bachelor's degree and 31% held a master's degree or higher as compared to the sample, for which 47% held a bachelor's degree, 50% held a master's degree or higher, and 4% chose not to respond to the question. Table 1 shows the percentage of teachers in the overall population and the corresponding percentages for the sample.

Table 1

Characteristics of Population and Sample

Characteristic	Population	Sample
N	62	45
Gender		
Female	84%	78%
Male	16%	16%
did not answer		7%
Ethnicity		
White	87%	73%
Black	10%	9%
Other/Unspecified	3%	9%
did not answer		9%
Education		
Bachelor's	69%	47%
Post-graduate	31%	50%
did not answer		4%

Instrument

Cronbach alpha scores were calculated for a number of questions within the survey. Scores for each type of student (1-4) were calculated and are shown in Table 2.

Table 2

Reliability Scores for Student Types

Type	N	Items	alpha
1	43	3	0.79
2	43	3	0.76
3	45	3	0.86
4	44	3	0.70

All of these scores fall within the acceptable range for reliability.

Reliability scores were also calculated for the questions about teacher confidence in and perceived usefulness of the data. Questions were grouped in two ways for these calculations. These results are reported in Table 3.

Table 3

Reliability Scores for Confidence in and Usefulness of the Data

Question	N	Items	alpha
Confidence in data	41	4	0.65
Usefulness of data	41	4	0.58
Student success/individualizing instruction	41	4	0.79
Teacher effectiveness/professional development	41	4	0.62

Of these scores, only the coefficient for student success/individualizing instruction falls within the acceptable range.

Research Question 1

The first research question concerned teacher perceptions of student success. For this question, teachers were asked to examine data, either status data or growth and status data, and rate the success of twelve different individuals representing four types of students. Three students of each type were included in the matrix (see figure 2).

Type of Student

Type 1 High Status/ High Growth	Type 2 Low Status/ High Growth	Type 3 High Status/ Low Growth	Type 4 Low Status/ Low Growth
Student 11	Student 02	Student 09	Student 07
Student 14	Student 18	Student 10	Student 12
Student 15	Student 24	Student 25	Student 26

Figure 2: Types of students used as subjects in teacher reports

Teachers rated each student on a six-point scale ranging from very successful (1) to very unsuccessful (6). For each type of student, the three scores were averaged for each teacher, resulting in a single mean score per respondent. Prior to comparing the means, an outlier analysis was conducted using boxplots. Outliers were defined as those scores which fell between 1.5 and 3 IQR's from the upper and lower limits of the interquartile range. One respondent was an outlier for two of the four types of students and was excluded for the purposes of research question 1. Three other respondents were an outlier in only one category; they were included in the analyses for research question 1.

These means were compared using independent samples t-tests, the results of which are shown in Table 4. Type 1 students are those who showed high levels of both growth and achievement; while the mean rating of teachers receiving both types of data were higher than those receiving only status data (both = 1.737, status = 2.027), they were not significantly different ($t=1.582$, $p>.05$). Type 2 students were those who showed high growth while still having low achievement. Mean ratings for growth data teachers were significantly different than those teachers with only status data ($t = 2.687$, $p<.05$). Teachers with access to growth data rated these students as more successful ($\bar{x} =2.632$) than those with only status data ($\bar{x} = 3.147$). The opposite was true for Type 3 students, those with low growth but high achievement. Teachers with access to both types of data rated Type 3 students as less successful than teachers with only status data (growth $\bar{x} = 3.175$, status $\bar{x} =2.093$). This difference was also significant ($t = -4.427$, $p<.05$). For the final type of student, those with low achievement and low growth, there was also a significant difference in the perceptions of student success ($t = -3.345$, $p<.05$). As with Type 3 students, these students were perceived as less successful by the growth teachers ($\bar{x} =3.912$) than the status only teachers ($\bar{x} =3.300$).

Table 4

Comparison of Mean Scores for Type of Student by Type of Data Received

	Status Data *n=25*		Growth and Status Data *n=19*					Effect Size
Type	*M*	*SD*	*M*	*SD*	*t*		*p*	*d*
1	2.03	0.58	1.74	0.62	1.58		0.12	0.48
2	3.15	0.49	2.63	0.78	2.69	*	0.01	0.81
3	2.09	0.71	3.18	0.91	-4.43	*	0.00	-1.33
4	3.30	0.53	3.91	0.70	-3.35	*	0.00	-0.99

$p < .05$

Research Question 2

The second research question asked if teachers receiving both growth and status data differed in their perceptions of their instructional effectiveness compared to teachers who receive only status data. Data to analyze this question was collected in the form of each teacher indicating one area of strength and one area of weakness of out five possible reporting categories.

For the area of strength, the target categories were reporting categories one and five. In the status data, mean student scores for reporting category 1 were higher than those for the remaining four categories. In the growth data, students showed growth in reporting category five more frequently than in the other four categories. In order to perform the analyses, cases were limited to those respondents who chose either of the two categories, eliminating any respondent who chose category two, three, or four. Crosstabs were then performed for all remaining cases. The results, seen in Table 5, showed that while 86% of the teachers with only status data chose category 1 as a strength, only 50% of those receiving both growth and status data selected reporting category 1 as a strength. A chi-squared test with Yates correction was performed and significance was found [χ^2 (Yates) = 3.972, p<.05].

Table 5

Areas of Strength by Type of Data Received

	Reporting Category		
Type of Data	1	5	Total
Status only			
Count	19	3	22
% within type	86%	14%	100%
Growth and status			
Count	7	7	14
% within type	50%	50%	100%

Similar analyses were performed for areas of weakness. In this case, reporting categories 4 and 3 were the targeted categories for status and growth data respectively. For status data, student mean scores were the lowest in reporting category 4. For the growth data, reporting category 3 had the highest frequency of students showing low growth. Analyses were performed for those cases where either of the two categories was chosen as the weakness by the respondent. As can be seen in Table 6, while 100% of the teachers receiving status data selected reporting category 4 as a weakness, only 43% of those receiving both types of data chose it as their weakness. Because the value of one cell was zero, the chi-squared test was not performed.

Table 6

Areas of Weakness by Type of Data Received

	Reporting Category		
Type of Data	1	5	Total
Status only			
Count	0	21	21
% within type	0%	100%	100%
Growth and status			
Count	8	6	14
% within type	57%	43%	100%

Research Questions 3 and 4

Research questions 3 and 4 were designed to measure the teachers' confidence in the data. Question 3 focused on their confidence in the data as an accurate portrayal of student achievement while question 4 examined their confidence in the data as an accurate portrayal of teacher effectiveness. For each of these questions, teachers were asked to rate their level of confidence in comparison to annual assessment data they had received in the past. The scale was a five-point, likert-type scale with the choices ranging from much more confident (5) to much less confident (1). Because teachers receiving only status data received the same data that is currently available in the state, a neutral response of same level of confidence was included. A boxplot was used to examine the data for outliers with those scores falling between 1.5 and 3.0 IQR's from the upper and lower limits of the IQR defined as outliers. Following the outlier analysis, an independent samples *t*-test was run to compare the means.

Research question 3 examined teachers' confidence in the data as an accurate portrayal of student achievement. The boxplot revealed one outlier; this case was excluded from the *t*-test. As shown in Table 7, the means were not significantly different.

Table 7

Teachers' Confidence in the Data as an Accurate Portrayal of Student Achievement

Status Data n=24		Growth and Status Data n=16				Effect Size
M	*SD*	*M*	*SD*	*t*	*p*	*d*
3.29	0.75	3.38	0.89	-0.32	0.75	-0.11

Teachers' confidence in the data as an accurate portrayal of teacher effectiveness was examined for research question 4. The boxplot revealed ten outlier values, representing greater than 20% of the cases; no cases were excluded for the *t*-test. The results of the independent samples *t*-test failed to show significant differences (see Table 8).

Table 8

Teachers' Confidence in the Data as an Accurate Portrayal of Teacher Effectiveness

Status Data n=24		Growth and Status Data n=16				Effect Size
M	*SD*	*M*	*SD*	*t*	*p*	*d*
2.88	*.61*	*3.06*	*1.03*	*-.72*	*.48*	*-.22*

Research Questions 5 and 6

The final research questions examined the potential usefulness of the data for decision making. Teachers were again asked to compare the data used to complete the survey to annual assessment data they had received in the past. The same five point likert-type scale was used with five representing "much more useful" and one being "much less useful". The neutral choice of same level of usefulness was included as some teachers received the same data that was currently available in the state. Boxplots were used to examine the data for outliers prior to running an independent samples *t*-test to compare means.

Research question 5 asked the participants to consider the usefulness of the data for making decisions regarding individualizing instruction in the classroom. The boxplot revealed one outlier case, which was excluded from further analysis. The *t*-test failed to show significant differences (see Table 9).

Table 9

Perceived Usefulness of the Data for Individualizing Instruction

Status Data		Growth and Status Data				Effect Size
n=24		n=16				
M	*SD*	*M*	*SD*	*t*	*p*	*d*
3.50	*.83*	*3.63*	*.96*	-*.44*	*.66*	-*.15*

The final research question asked the participants to consider the usefulness of the data for guiding professional development. The boxplot revealed one outlier case, which was excluded from further analysis. The *t*-test failed to show significant differences (see Table 10).

Table 10

Perceived Usefulness of the Data for Guiding Professional Development

Status Data		Growth and Status Data				Effect Size
n=24		n=16				
M	*SD*	*M*	*SD*	*t*	*p*	*d*
3.17	*.72*	*3.53*	*.72*	*-1.55*	*.13*	-*.50*

Open Ended Responses

Because of the exploratory nature of the research, participants were given the opportunity to provide any additional comments they may have had regarding the data and/or questionnaire. Though a limited number of participants chose to comment, there were similarities in the responses. One teacher with status data commented that "testing is only a small piece of the picture" while one with growth data reported that "pretty much all testing done… using a multiple choice format is not any indication of teacher effectiveness or student achievement." A number of teachers with status data indicated that a lack of knowledge about the individual students hindered their ability to evaluate the students. These included that "what may be success for one may not be for another" and "it is difficult to state a confidence level on data that is not representative of my actual students". A teacher with growth data indicated that "without student names, there is no way to individualize" instruction.

Summary of the Results

This research examined the potential of growth data on teachers' perceptions of student success and teachers' perceptions of their own effectiveness. It also examined teachers' confidence in and the perceived usefulness of the data. Several statistically significant differences were found. When rating student success, teachers receiving growth and status data differed significantly from those receiving only status data for three of the four types of students. Specifically, students with high growth were rated more favorably by teachers with growth and status data than by those with only status data. Similarly, students with low growth were rated less favorable by teachers with growth and status data than by those with only status data. Additionally, teachers receiving growth and status data were different in their reporting of their strengths and weaknesses. No significant differences were found with regard to teachers' confidence in the data as a measure of student achievement or teacher effectiveness. Additionally, no significant differences were found for teachers' perceived usefulness of the data for individualizing instruction in the classroom or guiding professional development, however, the differences for guiding professional development did approach significance.

Chapter 5
Discussion, Conclusions, and Recommendations

This chapter is divided into three sections: discussion, conclusions, and recommendations. The discussion portion of the chapter examines the findings in light of the existing literature. This is followed by conclusions which can be drawn from this study and a section on recommendations for practice and further research.

Discussion

In this section, the findings of this study are examined in relationship to the existing literature and the research hypothesis. Major areas that are discussed include perceptions of student success, teacher strengths and weaknesses, and confidence in and usefulness of the data. The section concludes by looking at the potential limitations of the study.

Teacher Perceptions of Student Success

Several significant differences were found when considering teacher perceptions of student success. These included significant differences for three of the four student types. Additionally, each of these differences had large effect sizes.

Teachers with access to growth and status data clearly distinguished between students who were making progress despite not meeting the benchmark by ranking Type

2 students (low achievement, high growth) as more successful than those teachers with only status data ($t = 2.687$, $p < .05$). Additionally, even though the sample size was small, the effect size was large ($d = .81$). This finding is consistent with the research hypothesis - that teachers receiving both growth and status reports would rate Type 2 students (low status, high growth) more favorably than teachers receiving only status reports.

For Type 1 students (high achievement, high growth), the study failed to find a significant difference between the two groups. This is consistent with the research hypothesis, which was that teachers from both groups would rate Type 1 students similarly. It is noteworthy, however, that while not statistically significant, teachers with access to growth data rated these students more favorably than those with only status data and the effect size (Cohen's $d = .48$) was medium, indicating that there may practical significance. This finding confirms that teachers value growth as was seen with Type 2 students.

Similarly, students with low growth were viewed differently regardless of their achievement level as both Type 3 (high achievement) and Type 4 (low achievement) students received lower success ratings from the teachers with access to growth and status data than from teachers with access to status only data. For Type 3 students (high status, low growth), the hypothesis was that these students would be rated less favorably by teachers receiving both growth and status reports when compared with the ratings from teachers receiving only status reports. The findings confirmed the research hypothesis (t = -4.427, p < .05). Additionally, the effect size was large (d = 1.33) even with a small sample size.

Type 4 students were those students with both low achievement and low growth. It was hypothesized that these students would be rated similarly by teachers with access to both types of data and those with access to only status data. In this case, the hypothesis is rejected as significant differences were found. Specifically, teachers with access to both types of data rated Type 4 students less favorably than teachers with access to only status data (t =-3.345, p < .05). Additionally the effect size was again large (d = .99).

Another important difference is the value that teachers place on growth. This is clearly indicated when comparing the means of Type 2 and Type 3 students. Type 2 achievement students were defined as low achievement/high growth while Type 3 students were defined as high achievement/low growth. Despite the high achievement level of the Type 3 students, teachers with growth data rated them less favorably than the Type 2 students. Clearly the teachers with access to growth data viewed growth in a student as more important in defining success than the achievement level of the student.

Several research studies (Heck, 2006; Zvoch and Stevens, 2006; Zvoch and Stevens, 2008) indicate that a very different picture of student achievement is revealed when using growth data in combination with status data rather than using status data alone. Models of student achievement that take into account growth and status data reveal four types of schools rather than just two. This study indicates that the same is true for teachers with access to growth and status data rather than just status data. Furthermore, the ability to identify four types of schools and students has implications in terms of the needs of those schools and, in this case, those students. Status models, which identify only two types of schools or students, misidentify Type 2 and Type 3 schools and students.

The anticipated needs of these schools and students as identified by a growth-achievement matrix are very different from those identified by a status only system (Heck, 2006; Zvoch & Stevens, 2008). The growth experienced by Type 3 schools and students indicates that the instructional practices and programs that are currently in place are effective, while a status only model would indicate that the school is not effective or the student is not learning. The needs of this school or student would be very different from a Type 4 school, which has both low growth and low achievement, yet they would be identified as similar in a status only model. Similarly, as the AMO for making AYP increases, Type 2 schools, which have previously made AYP based on their high achievement in a status model, may eventually fall short of the AMO if they continue to show low growth. The needs of a Type 2 school or student, where achievement is high but growth low, may more closely align with the needs of a Type 4 school or student, but this would not be evident from a status only model.

Teacher Strengths and Weaknesses

In addition to a changing picture of student success, the understanding of teacher strengths and weaknesses was impacted by the availability of growth data. Teachers with access to growth data relied on the information in the growth report to identify their strengths and weaknesses at least as often as they relied on the status data. Data to analyze this question was collected in the form of each teacher indicating one area of strength and one area of weakness of out five possible reporting categories. Teachers having access to both growth and status data differed significantly from those teachers having access to only status data in their perceptions of their strengths [χ^2 (Yates) = 3.972, $p < .05$]. Additionally, although a chi-square test was not run for teacher-identified weaknesses (zero in one cell), the frequencies reported indicated that greater than half of the teachers with access to growth data used the information in the growth report to make the determination rather than the information in the status report. Given the reservations that teachers have regarding data, their willingness to use a data report that they have never seen before rather than one that has been available for many years speaks to the value of growth to teachers.

Both Heck (2006) and Zvoch and Stevens (2008) indicate that the growth data are an important component in planning school improvement strategies. Heck concluded by stating that, while further research is necessary, “growth models provide a more comprehensive framework for school assessment and a direct means for superintendents and principals to identify student and school needs and engage in planned efforts to strengthen instructional processes" (p. 695). Zvoch and Stevens (2008) concluded that "These results suggest that conclusions regarding the performance and instructional practices of schools could vary widely and/or be misguided depending on the indicators used in a school accountability system" (p. 588). This study suggests that teachers can also extract different information about their professional needs based on growth data, which can used to identify needs and inform practice.

Confidence In and Usefulness of the Data

This study also examined the potential of growth data to overcome barriers to data use in the educational setting based on a model of data driven decision making. This was asked in two ways: whether teachers had greater confidence in the data as a representation of student achievement or instructional effectiveness and whether they perceived the data to be more useful for individualizing instruction or guiding professional development. In each case, teachers receiving growth and status reports did not differ significantly from those receiving only status reports. With one question, however, the results did approach significance. When asked whether the data was useful for guiding professional development, teachers with access to growth and status data rated the data more useful, although not significantly more useful, for guiding their own professional development.

Several factors could have contributed to the non-significant findings in this area. First, it is important to examine the reliability data for these questions. When comparing the two questions that asked about students (portrayal of students success, individualizing instruction), the reliability was adequate (alpha = .79). When comparing the two questions that asked about teachers confidence in the data, the reliability was lower (alpha = .65). Reliability was also lower when comparing the two questions that asked about the usefulness of the data (alpha = .58). Finally, reliability was also low when comparing the question which asked about teacher effectiveness with that asked about using the question to guide professional development (alpha = .61). Although none of the four questions reached statistical significance, the low reliability may be a contributing factor. With more reliable scores, significance may have been observed.

Another explanation for the lack of significance may have been in the structure of the question. With regard to wording of the question, because some teachers were receiving the same data that are currently available in the state while others were receiving the additional growth data, the questions were phrased as “compared to annual assessment data that you have received in the past.” This resulted in a narrow scale that may have masked any differences. The narrowness of the scale is complicated by an already identified barrier to data use: access. For this study, teachers in the status only group received reports that mirrored the data that are currently available to them. Given this, the predicted score for their confidence in and perceived usefulness of the data "*compared"* to what they have seen before should have been "the same as". However, for three of the four questions, the mean score for teachers from the status only group was above "the same as". This indicates that while the data have been available, teachers may not have had access to it.

Finally, it may be that the lack of confidence in and usefulness of the data expressed by the teachers in the study is a reflection of their overall experience with standardized testing data, rather than specifically related to the growth data. Pedulla et al (2003) found that the perceptions of the value of standardized tests among teachers was low and that teachers did not believe the test was able to measure the abilities of students or the effectiveness of schools and school processes. Despite this general distrust of standardized test data, two of the teachers in the growth and status report group made positive statements about the growth data in the open comment portion of the survey. One wrote, "This would eliminate so many issues when the scores are looked at!" A second wrote, "If this type of info was supplied, I would want it in addition to the 'status report' data." Both of these statements indicate the information contained in the growth report not only supplied additional information, but that the teachers valued the information it contained.

The framework for data-driven decision making (DDDM) indicates that data (raw figures) must first become information by being placed into context which can then become actionable knowledge when embodied in an individual (Ikemoto & Marsh, 2007;

Mandinach, Honey, & Light, 2006; Petrides & Guiney, 2003) Such knowledge, once generated, has the potential to be used for planning and instruction (Marsh, Pane, & Hamilton, 2006). Unfortunately, recent reports published by the U. S. Department of Education on data use by teachers indicate that the data available are not being used in this manner (Gallagher, Means, & Padilla, 2008; Means, Gallagher, & Padilla, 2007). Research into data use has revealed several factors which may either promote or inhibit the use of data; these include access to useful data, time to devote to data analysis, capacity for data use, and trust in the data. While the findings related to models of student achievement and teacher strengths and weaknesses indicate that growth data provide educators with a different perspective, it is unclear at this time how useful growth data are in overcoming the barriers to data use.

Limitations

For this study, true experimental design was used. An advantage of this methodology is that it limits, and in some cases eliminates, a number of concerns that are often found in educational research regarding limitations of the study. However, several threats to both internal validity and generalizability of this study still exist. These include treatment fidelity, instrumentation, differential attrition, sample size, and subject matter.

Treatment fidelity is an issue because of the number of survey administrations and administrators. The researcher administered the survey at the initial department chair meeting and at the math department meeting at one of the four schools. The department chairs at the remaining four schools administered the survey to their respective math departments. During both administrations conducted by the researcher, despite explicit instructions that were read to participants prior to receiving the documents to complete the survey individually, there was some discussion on the part of the participants. While the researcher was able to discourage this discussion at the administrations conducted personally, there is no way of knowing if the same discussions took place at the remaining administrations. There is also no way of knowing how these discussions were handled by the survey administrator if they did take place.

Another potential threat is instrumentation. One concern with the instrumentation is the actual reports that were used by the participants. The state in which the survey was conducted does not currently report growth data. Because of this, the growth reports had to be created based on models from states that do report growth and the way data are reported in the state. For all participants, the administration of the survey was the first time that growth reports were seen. Indeed, for many, this was most likely the first exposure to reporting growth using the state standardized test data. Given this, there was a need to balance providing adequate information about a growth report such that the participants could understand the information with the need to not influence their opinion about the use or value of the growth report.

Another concern with the instrumentation is the survey. Because the data reports had to be created, the survey also had to be created by the researcher. While reliability coefficients (Cronbach's alpha) fall in the acceptable range for the types of students (.79, .76, .86, and .70 respectively), the same cannot be said for the questions about teachers confidence in and perceived usefulness of the data. Reliability tests for these questions were run two ways for each question, only one of which produced an acceptable value.

An additional concern is differential attrition. While the administration of the survey, in theory, should have produced an equivalent number of responses, twenty-five status data surveys were returned while only 19 growth data surveys were returned. The possibility exists that those teachers who had the most difficulty understanding the growth data were those that did not return the survey. This could possibly have skewed the results in favor of those who understood and therefore used the growth data.

In addition to issues related to internal validity, a number of threats to the generalizability of the study exist. The study was conducted using math teachers from select middle schools within a single school division from a single state. Of the over 100,000 teachers in the state, only 62 had the opportunity to participate, and of those, only 45 actually participated. While many of these restrictions were necessary, they do limit the generalizability of the study.

One restriction was placed on the research by the county in which it was conducted. As a condition of participating in the research, the county limited the number schools that were accessible to the researcher in order to decrease the burden on the teachers. Teachers from less than half of the schools in the county participated.

Because this was an exploratory study, only one subject matter was examined. Math was chosen for a number of reasons; however, the generalizability to other subjects is limited. One reason for choosing math was the number of reporting categories that exist in the current system (five), however, other subjects, such as reading, have far fewer reporting categories. Additionally, the same reporting categories are used for all math tests in grades three through eight. This means that growth can be calculated for the test as a whole as well as within individual categories. In other tested areas, however, such as social sciences, the reporting categories change from year to year, such that only overall growth could be calculated. Additionally, high school math courses such as Algebra and Geometry use different reporting categories than those used in grade three through eight, limiting number of growth measures that would be available at that level. An additional advantage of using math data is that students are tested in math each year. While this is also true of reading, it is not true of the other subjects tested. For a subject such as science, where students take a test in grade five and are then not tested again until grade eight, growth data would have limited usefulness, again reducing the generalizability of this study.

Conclusions

Based on the results of this study and the existing literature, several conclusions can be drawn at this time. The first, as indicated in the findings on teacher perceptions of student success, is that using growth data in addition to status data provides a substantively different picture of student success than using status data alone. In this study, teachers were presented with data that they could not have seen before, yet they were able to distinguish the four types of students based on the data. This was evident as teachers with access to growth data consistently rated students with high growth more favorably and students with low growth less favorably than teachers with only status data. Additionally, the comparison of Type 2 and Type 3 students indicates that teachers value growth over achievement.

A great deal of attention is given to the achievement gaps in education. What is not always evident from the national discussion is that achievement gaps are not created by educators; rather, they exist the moment students walk through the doors. Given these gaps and the current education law, in order to meet the demands placed on them, educators are required to create a system in which previously underperforming students show consistently more growth than their peers. To continue to evaluate schools and teachers based on systems that reflect the status of society rather than their efforts is both unproductive and unfair. More importantly, schools and teachers who achieve significant amounts of growth with the neediest students continue to go unrecognized in a status model while potentially ineffective schools and teachers are rewarded based on the status quo. Such a system only serves to perpetuate the achievement gap, not close it.

While this study clearly shows that teachers value growth and make different decisions about student success when they have access to growth data, it is not yet clear whether the availability of growth data would impact teachers' perceptions of student gain over time. A teacher making an evaluation of an actual student or class would have access to far more than one set of standardized test scores, whether those scores included growth or not. It may be that the classroom level data and anecdotal data that teachers already have available to them, and which they have more confidence in (Kerr et al, 2006) would provide the same information as growth data.

Another conclusion that can be drawn from this study and the existing literature concerns teachers' understanding of their strengths and weaknesses. When examining growth data, teachers were able to identify different areas of need than were evident from status reports. This is similar to results found when examining growth data at the school level. The additional information that is inherent is growth reports has the potential to better identify the needs of students, teachers, and schools. Given the current charge for the educational system to overcome the achievement gap, it is necessary for districts, schools, and teachers to be able to determine what is working and what is not. This study clearly shows that this understanding varies based on the type of data received, indicating another potential use for growth data.

It cannot be assumed, however, that growth data from standardized testing is the only means of reaching these conclusions at the teacher level. As indicated when examining perceptions of student success, teachers have many additional sources of data available above and beyond standardized testing data, however, teachers in this study were asked to make their decisions using only the reports available to them. There is evidence that teachers are able to successfully use data for a variety of purposes (Brunner et al, 2005) and that this use can be an effective tool in closing achievement gaps (Symonds, 2004). While growth data provide a different picture of teacher strengths and weaknesses, when considering real students and real classes, teachers may be able to reach the same conclusions from the additional data that is available to them.

This leads to the final conclusion, which is that growth data alone may not be able to overcome certain barriers, such as teachers' confidence in the data and their perceptions of the usefulness of standardized testing data as part of their decision making process. While the lack of significance found in this area may be due to limitations of the study rather than a true reflection of the value of growth data, these findings indicate that those currently implementing a growth model and those considering implementing a growth model should understand that it may have limited value at the teacher level.

This does not mean that growth data do not provide essential information. School, district, and state level decision makers cannot have the same understanding of what is happening at the classroom level as an individual teacher. It may be that growth as measured through standardized tests is the only accurate picture that decision makers at this level have available to them. When considering the effectiveness of programs, policies, and practices, growth data provide a substantially different picture than status data. Examining growth data at the school or district level has the potential to prevent the misidentification of Type 2 and Type 3 schools. Without growth data, ineffective programs at Type 2 schools may continue to be viewed as effective while effective programs at Type 3 schools may be viewed as ineffective. Such misidentification could influence practice, and the allocation of resources, moving forward, preventing progress. As such, growth data should be an important source of information for school and district level decision makers, one which provides a more comprehensive framework for school assessment than a status only model.

Recommendations

Based on this study and its findings a number of recommendations can be made. These include recommendations for practice at the state, district, and school level and recommendations for further research.

For Practice

Recommendations for practice exist at a number of levels within the educational system. These levels include state, district, school and teacher. At the state level, for those states which are developing or considering growth models, that process should include a careful examination of how the data that is obtained from these models can be used not only for accountability, but also for guiding practice. Creating a system that is used solely for judging schools and teachers will have limited ability to have a positive impact on the changes that need to be made. For those states with models that have already been implemented, those models need to be examined to determine what kinds of information is available or can be made available to schools and teachers that will assist them in moving toward the goals, not just informing them of when they have or have not been met.

Districts and schools also play an important role in the process of guiding change. Districts need to develop systems, with or without the availability of growth data, that allow schools and teachers access to data in ways that allow them to examine the data for trends as it relates to individual students, teachers, schools, and programs. At the school level, administrators need to develop their own capacity for understanding, analyzing, and using data as well as developing structures within their school that both encourage and enable the use of data by departments and individuals.

Another consideration for schools and districts is the amount of time that is available to teachers to analyze data. Research shows that schools that are successful in closing the gap have designated time set aside during the school day for examining the data (Symonds, 2004). Regardless of the type of data available, whether growth or status data from standardized testing or data available from classroom and school level assessments, teachers generally are not trained in the use of data. In order for data to become information which then becomes knowledge, teachers will need to devote time to understanding the data that is available and that time needs to be made available by schools and districts.

Additionally, schools and districts need to consider how to best guide the use of data. Although not a consideration for this research, several studies point to leadership in the use of data as central to the endeavor. Feldman and Tung (2001) and Lachat and Smith (2004) found that schools with effective data use had leadership that was committed to data use and had built a vision for such use. More specifically, Lachat and Smith (2004) found that this leadership could be distributed. School leaders, like teachers, are not necessarily trained in the use of data. Identifying leaders within the school who have an affinity for data use and who can take on the role of data coach should be a priority for school principals, especially as new and different data becomes available.

For Further Research

The era of high stakes testing is not likely to go away. Already, tests designed to measure student achievement are being used to measure school districts, individual schools, and most recently, teachers. As this study shows, growth data has the potential to provide a different picture of student progress than status data alone. If we are truly going to leave no child behind in our race to the top, it is imperative that different means of enabling teachers to become better educators be explored. Growth data should not be ruled out as a means of accomplishing this goal; however, a great deal of research still needs to be done in order to determine the most effective ways to use growth data.

One area of research should focus on the types of models currently in use or in the implementation stage. In January 2010, the Council of Chief State School Officers (CCSSO) surveyed accountability directors in all fifty states (Blank, 2010). The survey revealed that 17 states are currently using a growth model while 13 are in the process of developing a growth model. Additionally, the survey revealed that the models of growth used by states varied and included linear growth, growth to proficiency, and value-added, among others (Blank, 2010). Studies which compare the different models and specifically the usefulness of the models at the state, district, and school level should be undertaken.

Another area examined by the CCSSO was the intended purpose of the growth models. Interestingly, while 27 states indicated that the model would be used for accountability, only 20 states reported that its purpose was to identify successful improvement strategies (Blank, 2010). Given this, comparison studies of growth models according to intended purposes should also be undertaken. Seventeen states indicated that their growth models would be used to evaluate programs (Blank, 2010), which presents another area of research that can be explored.

Finally, the value of growth data to educators should continue to be examined. This study clearly demonstrates the importance teachers place on growth. Given this, there exists the potential that teachers will be able to use growth data in ways that are different from the data that they currently have. While 14 states indicated that growth data was reported to teachers, only four states indicated that individual student growth data was available through a data warehouse (Blank, 2010). Research into the use of growth data at the teacher level needs to examine the type of data available to teachers and the value of that data to teachers.

List of References

List of References

Barone, C. (2009, March). *Are we there yet?: What policymakers can learn from Tennessee's growth model.* Education Sector. Retrieved October, 30, 2009, from http://www.educationsector.org/research/

Bernhardt, V. L. (2004). *Data analysis for continuous school improvement.* Larchmont, NY: Eye on Education.

Blank, R.K. (2010). *State growth models for school accountability: Progress on developing and reporting measures of student growth*. Washington, D.C.: Council of Chief State School Officers.

Brunner, C. Fasca, C. Heinze, J. Honey, M., Light, D. Mandinach, E. et al. (2005). Linking data and learning: The grow network study. *Journal of Education for Students Placed at Risk*, 10(3) 241-67.

Carlson, D. (2001). *Focusing state educational accountability systems: Four methods for judging school quality and progress*. Retrieved October, 21, 2009, from http://www.aceeonline.org/presentations/carlson_models.pdf

Center for Public Education, (2007a, November 9). *Measuring student growth: At a glance.* Retrieved November 16, 2009, from http://www.centerforpubliceducation.org/site/apps/nlnet/content3.aspx?c=lvIXIiN0JwE&b=5130063&content_id={9BA6BD92-A81A-4891-B77E-77029506B39D}&notoc=1

Center for Public Education, (2007b, November 9). *Measuring student growth: A guide to informed decision making*. Retrieved November 16, 2009, from http://www.centerforpubliceducation.org/site/apps/nlnet/content3.aspx?c=lvIXIiN0JwE&b=5117397&content_id={29512C02-FC06-45F3-BDAE-82CDEA6641C4}&notoc=1

Davis, F. B., (1964). *Educational measurements and their interpretations.* Belmont, California: Wadsworth Publishing Company, Inc.

DePascale, C. A. (2006). *Measuring growth with the MCAS tests: A consideration of vertical scales and standards*. Paper prepared for the Massachusetts Department of Education.

Feldman, J. and Tung, R. (2001, April) *Whole school reform: How schools use the data-based inquiry and decision making process.* Paper presented at the annual meeting of the American Educational Research Association. Seattle, WA.

Gallagher, Means, & Padilla, (2008). *Teachers' use of student data systems to improve instruction: 2005 to 2007*. Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.

Goldschmidt, P. & Choi, K. (2007, Spring). *The practical benefits of growth models for accountability and limitations under NCLB* (CRESST Policy Brief 9). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S., Blank, R., & Williams, A. (2005). *Policymakers' guide to growth models for school accountability: How do accountability models differ?* Council of Chief State School Officers. Retrieved October 15, 2009, from http://www.ccsso.org/publications/details.cfm?publicationid=287

Heck, R. H. (2006). Assessing school achievement progress: Comparing alternative approaches. *Educational Administration Quarterly*, 42(5), 667-699.

Holcomb, E. L. (1999). *Getting excited about data: How to combine people, passion, and proof*. Thousand Oaks, CA: Corwin.

Huynh, H. & Schneider, C. (2005). Vertically Moderated Standards: Background, Assumptions, and Practices. *Applied Measurement in Education*, *18*(1), 99-113.

Ikemoto, G. S. & Marsh, J. A. (2008). Cutting through the "data-driven" mantra: Different conceptions of data-driven decision making. In *Evidence and Decision Making: Yearbook of the National Society for the Study of Education*, *106*(1), 105-131.

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record*, 106(6), 1258-1287.

Jones, J. M. (2007). Improving the use and effectiveness of data among teachers. *Dissertation Abstracts International, 69* (01A), 132-48.

Kerr, K. I., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006, August). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education 112,* pp. 496-520.

Lachat, M., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk, 10*(3), 333–49.

Light, A., Wexler, D. H., & Heinze, J. (2005). *Keeping teachers in the center: A framework of data-driven decision-making.* New York: Center for Children and Technology.

Lissitz, R. W. & Huynh H. (2003). Vertical equating for state assessments: Issues and solutions in determination of adequate yearly progress and school accountability. *Practical Assessment, Research & Evaluation, 8*(10). Retrieved November 7, 2009, from http://pareonline.net/getvn.asp?v=8&n=10

Mandinach, E., Honey, M., & Light, D. (2006). *A theoretical framework for data-driven decision-making*. Paper presented at the annual meeting of American Education Research Associational, San Francisco, CA.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education*. Santa Monica, CA: Rand Education.

McMillan, J. M. (2004). *Educational research: Fundamentals for the consumer.* Boston, MA: Pearson Education, Inc.

Means, B. Gallagher, L. & Padilla, C. (2007). *Teachers' use of student data systems to improve instruction*. Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers.* Boston: National Board on Educational Testing and Public Policy.

Petrides, L. & Guiney, S.H. (2002). Knowledge management for school leaders: An ecological framework for thinking schools. *Teachers College Record 104*(8) 1702–1717.

Planty, M., Hussar, W., Snyder, T., Provasnik, S., Kena, G., Dinkes, R., KewalRamani, A., and Kemp, J. (2008). *The condition of education 2008* (NCES 2008-031). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Schafer, W. D. (2006). Growth scales as an alternative to vertical scales. *Practical Assessment Research & Evaluation, 11*(4) 1-6. Retrieved May 9, 2009, from http://pareonline.net/pdf/v11n4.pdf

Stecher, B.M. & Hamilton, L.S., (2006). *Using test-score data in the classroom* (WR-675-EDU). RAND Corporation. Retrieved November 7, 2009, from http://www.rand.org/pubs/working_papers/WR375/

Symonds, K. W. (2004, August). *After the test: Closing the achievement gaps with data.* San Francisco, CA: Learning Point Association and Bay Area School Reform Collaborative.

U.S. Department of Education (2009, January 8). Secretary Spellings approves additional growth model pilots for 2008-2009 school year (Press Release). Retrieved November 23, 2009 from http://www.ed.gov/news/pressreleases/2009/01/01082009a.html

Zvoch, K., & Stevens, J. J. (2006). Successive student cohorts and longitudinal growth models: An investigation of elementary school mathematics performance. *Education Policy Analysis Archives, 14*(2) 1-25.

Zvoch, K. & Stevens, J.J. (2008). Measuring and evaluating school performance: An investigation of status and growth-based achievement indicators. *Evaluation Review,32*, 569-595.

Appendices

Appendix A

Student Test Score Information Distributed to Participants

Student Achievement Test Status Information

Total Test Information

Achievement Levels

Student performance on each test will be reported as one of three achievement levels:

Pass/Advanced
Pass/Proficient
Fail.

Scaled Scores Total Test Information

In addition to the achievement levels, student performance on all tests will be reported as scaled scores. For ease in interpretation, the same scaled score range of 0 to 600 will be used on each test at each grade level.

The following summarizes the test scaled score information:

Pass/Advanced	500-600
Pass/Proficient	400-499
Failure	0-399

Reporting Category Information

Scaled Scores for Reporting Categories

Student performance will also be reported for the reporting categories that make up each test, as outlined in the test blueprints. Student performance on the reporting categories will be reported as scaled scores so that progress over time may be measured.

Student performance will be reported on a scale ranging from 0 to 50, with a scaled score of 30 on each reporting category representing approximately the level of achievement necessary to attain a proficient score on the total test. Students with reporting category scaled scores below 30 may need additional instruction in those content areas.

Student Achievement Test Growth Information

Description of Growth Data

Growth is determined by differences in pre-test/post-test performance and is calculated only for those students who took the pretest at the beginning of the academic year.

Expected Growth (EXP) is defined as 10 months progress in a school year.

Student growth is categorized by the number of months increase for the present score over the pretest score as follows:

Symbol	Description
++	Student demonstrates significantly more than expected growth.
+	Student demonstrates more than expected growth.
EXP	Student demonstrates expected growth.
-	Student demonstrates less than expected growth.
--	Student demonstrates significantly less than expected growth.
NDA	Results from pre-test unavailable at time of report.

Student Growth Data Reported

Overall Growth Rating

Indicates overall growth for the test for which information is reported.

Growth Summary

Indicates the number of reporting categories in which students exceeded, met, or fell short of expected growth.

Reporting Category Growth

Indicates the amount of growth a student demonstrated for each individual reporting category.

Class Growth Data Reported

Reporting Category Tally

Indicates the number of students in each category:

(++/+) above normal growth,
(EXP) expected growth, and
(--/-) less than normal growth.

Appendix B

Sample Data Reports Distributed to Participants

Standardized Testing Results **Group Code: Test Teacher 01**
Admin: Spring 2008 Non-Writing
Grade Level Mathematics Test - Status Report

	Overall		Reporting Category Scaled Score				
Name	Scaled Score	Proficiency	RC_1	RC_2	RC_3	RC_4	RC_5
Student 01	463	Pass/Proficient	50	29	36	27	50
Student 02	416	Pass/Proficient	50	22	36	27	34
Student 03	443	Pass/Proficient	32	39	50	24	29
Student 04	488	Pass/Proficient	38	50	50	24	34
Student 05	433	Pass/Proficient	50	31	33	27	31
Student 06	398	Fail	25	28	25	50	35
Student 07	398	Fail	22	33	30	50	28
Student 08	433	Pass/Proficient	38	34	31	50	27
Student 09	488	Pass/Proficient	38	50	50	24	34
Student 10	475	Pass/Proficient	38	29	50	31	39
Student 11	488	Pass/Proficient	38	50	50	24	34
Student 12	424	Pass/Proficient	29	34	31	24	50
Student 13	416	Pass/Proficient	32	29	28	50	31
Student 14	463	Pass/Proficient	50	29	36	27	50
Student 15	475	Pass/Proficient	38	29	50	31	39
Student 16	416	Pass/Proficient	50	22	36	27	34
Student 17	463	Pass/Proficient	50	50	28	35	31
Student 18	424	Pass/Proficient	29	34	31	24	50
Student 19	400	Pass/Proficient	29	50	26	27	29
Student 20	424	Pass/Proficient	29	34	31	24	50
Student 21	453	Pass/Proficient	50	39	28	31	34
Student 22	398	Fail	22	33	30	50	28
Student 23	424	Pass/Proficient	50	29	28	35	31
Student 24	398	Fail	22	33	30	50	28
Student 25	463	Pass/Proficient	50	29	36	27	50
Student 26	416	Pass/Proficient	50	22	36	27	34
Student 27	475	Pass/Proficient	38	29	50	31	39
Average	**439.1**		**38.4**	**34.1**	**36.1**	**32.5**	**36.4**

Standardized Testing Results
Admin: Spring 2008 Non-Writing
Grade Level Mathematics Test - Growth Report

Group Code: Test Teacher 01

	Overall		Growth Summary			Reporting Category Growth				
Name	Scaled Score	Growth Rating	++/+	EXP	-/--	RC_1	RC_2	RC_3	RC_4	RC_5
Student 01	463	EXP	1	4	0	EXP	EXP	EXP	EXP	++
Student 02	416	++ / +	4	1	0	++	EXP	+	+	+
Student 03	443	EXP	0	5	0	EXP	EXP	EXP	EXP	EXP
Student 04	488	EXP	0	4	1	EXP	EXP	EXP	-	EXP
Student 05	433	EXP	1	3	1	EXP	EXP	--	EXP	+
Student 06	398	EXP		1	3	1	EXP	--	EXP	+
Student 07	398	- / --	0	1	4	--	-	-	EXP	-
Student 08	433	EXP	0	4	1	EXP	EXP	--	EXP	EXP
Student 09	488	- / --	0	2	3	-	EXP	EXP	--	-
Student 10	475	- / --	0	1	4	-	--	EXP	--	-
Student 11	488	++ / +	4	1	0	+	++	++	EXP	+
Student 12	424	- / --	0	1	4	-	-	-	--	EXP
Student 13	416	EXP	1	3	1	EXP	EXP	--	EXP	+
Student 14	463	++ / +	4	1	0	++	+	++	EXP	++
Student 15	475	++ / +	4	1	0	+	EXP	++	+	++
Student 16	416	EXP	1	3	1	EXP	--	EXP	EXP	+
Student 17	463	EXP	1	4	0	EXP	EXP	EXP	EXP	+
Student 18	424	++ / +	3	2	0	+	+	EXP	EXP	++
Student 19	400	EXP	0	4	1	EXP	++	--	EXP	EXP
Student 20	424	EXP	1	3	1	EXP	EXP	-	EXP	++
Student 21	453	EXP	1	3	1	EXP	EXP	--	EXP	+
Student 22	398	EXP	1	2	2	--	EXP	-	+	EXP
Student 23	424	EXP	1	3	1	EXP	EXP	--	EXP	+
Student 24	398	++ / +	3	2	0	EXP	+	EXP	++	+
Student 25	463	- / --	0	2	3	EXP	--	-	--	EXP
Student 26	416	- / --	0	2	3	EXP	--	--	-	EXP
Student 27	475	EXP	1	3	1	-	EXP	EXP	EXP	++
	Reporting Category Growth Frequency Counts		**++ / +**			**5**	**5**	**4**	**4**	**11**
			EXP			**15**	**16**	**10**	**17**	**8**
			- / --			**6**	**6**	**13**	**6**	**3**

Appendix C

Survey

Survey: Teachers Use of Data in Instructional Decision Making

SECTION I: Demographic Information

Gender ☐ Male ☐ Female

Ethnicity ☐ While ☐ Asian/Pacific Islander
☐ Black ☐ American Indian/American Native
☐ Hispanic ☐ Unspecified

Highest degree obtained ☐ Bachelor's ☐ Post-Graduate

SECTION II: Data Analysis

Based on the data you received, indicate what level of success you believe each of the following students achieved. Circle one number for each student.

	Level of Success					
Student Number	Very Successful	Moderately Successful	Somewhat Successful	Somewhat Unsuccessful	Moderately Unsuccessful	Very Unsuccessful
Student 02	1	2	3	4	5	6
Student 07	1	2	3	4	5	6
Student 09	1	2	3	4	5	6
Student 10	1	2	3	4	5	6
Student 11	1	2	3	4	5	6
Student 12	1	2	3	4	5	6
Student 14	1	2	3	4	5	6
Student 15	1	2	3	4	5	6
Student 18	1	2	3	4	5	6
Student 24	1	2	3	4	5	6
Student 25	1	2	3	4	5	6
Student 26	1	2	3	4	5	6

Based on the data, which reporting category do you believe represents a strength for you as a teacher? (Please select only one category)

☐ RC 1 ☐ RC 2 ☐ RC 3 ☐ RC 4 ☐ RC 5

Based on the data, which reporting category do you believe represents an area of potential development for you as a teacher? (Please select only one category)

☐ RC 1 ☐ RC 2 ☐ RC 3 ☐ RC 4 ☐ RC 5

SECTION III: Confidence in and Usefulness of the Data (Please choose one answer)

1. Compared to annual assessment data that you have received in the past, how confident are you that these data are an accurate portrayal of student achievement?
   - ☐ Much more confident
   - ☐ Somewhat more confident
   - ☐ Same level of confidence
   - ☐ Somewhat less confident
   - ☐ Much less confident

2. Compared to annual assessment data that you have received in the past, how confident are you that these data are an accurate portrayal of teacher effectiveness?
   - ☐ Much more confident
   - ☐ Somewhat more confident
   - ☐ Same level of confidence
   - ☐ Somewhat less confident
   - ☐ Much less confident

3. Compared to annual assessment data that you have received in the past, how useful are these data for making decisions regarding individualizing instruction in the classroom?
   - ☐ Much more useful
   - ☐ Somewhat more useful
   - ☐ Same level of usefulness
   - ☐ Somewhat less useful
   - ☐ Much less useful

4. Compared to annual assessment data that you have received in the past, how useful are these data for guiding your personal professional development?
   - ☐ Much more useful
   - ☐ Somewhat more useful
   - ☐ Same level of usefulness
   - ☐ Somewhat less useful
   - ☐ Much less useful

SECTION IV: Open-Ended Response
Please add any additional comments you may have regarding the data.

____

____

____

____

Vita

Patricia Louise Fox was born on March 19, 1967, in New Haven, Connecticut. She graduated from Thomas Dale High School, Chester, Virginia, in 1985. She received her Bachelor of Arts in Biology from the University of Virginia, Charlottesville, Virginia, in 1989. She worked in medical research from 1989 until 1994. She received her Master's of Teaching from Virginia Commonwealth University, Richmond, Virginia, in 1995, at which time she began teaching in Chesterfield County Public Schools. She received her post-graduate certificate in administration from Virginia Commonwealth University, Richmond, Virginia, in 2002. She currently works as an assistant principal in Chesterfield County Public Schools.