ABSTRACT

HINES, ROBERT THOMAS. An Exploration of the Effects of School Improvement Planning and Feedback Systems: School Performance in North Carolina. (Under the direction Dr. James Swiss)

School districts across the nation frequently apply a formal school improvement planning process as one mechanism to improve school performance. School improvement planning is largely based on the concepts of performance management, which include strategic planning, short-term measurable goals, and frequent feedback.

The goals arising out of the planning process provide a common direction for members of the organization and offer guidance for day-to-day decisions. While a large body of research addresses the value of planning and goal-setting practices in a variety of business organizations, much less analysis considers the application of these practices to schools. This analysis attempts to add to this body of knowledge by assessing how the use of these planning and performance management tools affects school performance.

This study analyzes school improvement plans developed by 273 North Carolina elementary and middle schools (grades Pre-K through 8, inclusively), categorizing them by their number of goals, the specificity of their goals, the difficulties of their targets, and the strength of their feedback. It then uses hierarchical regression analysis to examine how these differences in goal setting and feedback relate to school performance, as measured both by a one-year measure and by school improvement over a three year period.

With respect to single-year performance, the analysis found that the full model, including controls for student demographics and other school characteristics, explained only about 10% of the variance in school performance ($R^2 = .099$). Of that 10%, about half was attributable to the independent variables analyzed while the rest was attributable to the control variables.
Regarding the three-year change model, again the overall (full model) variance explained was modest at about 17% ($R^2 = .169$). About 9.5% of this variance was attributed to the independent variables and the remainder – around 7.5% – to the controls. Some of the reasons for this finding may relate to data limitations, as will be discussed below.

However, the analysis established that some aspects of the planning process are more important than others. On one hand, some aspects were surprisingly unimportant. For example, in Model 1, the single-year performance model, the beta coefficients for the variables measuring goal clarity, feedback, and the two goal difficulty blocks all were non-significant. The introduction of each of the blocks of variables into the hierarchical model resulted in only very small, non-significant changes in $R^2$ (change in $R^2 = .008, .005$ and $0.003$, and $0.009$, respectively).

However, some performance management aspects were related to clear performance improvement. For example, in Model 1, schools with a larger number of specified goals and strategies performed at a notably higher level ($\beta = .127, p < 0.05$). This finding was contrary to some previous research, which suggested that large numbers of goals and milestones harmed effectiveness by diffusing organizational focus. However, this finding was in accord with yet other research which indicates that having too few goals can lead to goal displacement (tunnel vision) and that having a large number of strategies can strengthen the performance of agencies if they have many task immature workers.

Strength of feedback about progress also had a positive impact in the multi-year model, even though the beta coefficient was not significant. As might be expected, the feedback measure that was based on teacher reports of actual behavior showed a larger relationship than the feedback measure based on a school’s intentions.
In the second, multi-year performance model, the measure of clear targets resulted in a significant beta coefficient ($\beta = .150, p < 0.05$), suggesting that higher percentages of clear targets increase performance. The measure for easy math goals also was significant ($\beta = -.237, p < 0.05$), supporting the hypothesis that performance decreases if goals are too easy. The remaining coefficients remained non-significant.

The analysis was limited by the lack of availability of a number of potentially important factors. The study was unable to measure the extent to which each principal actively supported the system, which research indicates is often important in system success. It also lacked measures of school-level expenditure data, and of past planning efforts that preceded the one-year snapshot. The addition of this data may have improved the study’s relatively modest $R^2$.

Nonetheless, the analysis contributes to the literature on both performance management and school improvement planning by operationalizing the common planning and goal-setting concepts of organization focus, goal clarity, goal difficulty, and feedback with using school artifacts and data, ultimately achieving findings that suggest several next steps for better understanding the role of school improvement planning processes in raising school performance.
An Exploration of the Effects of School Improvement Planning and Feedback Systems:

School Performance in North Carolina

by

Robert Thomas Hines

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Public Administration

Raleigh, North Carolina

APPROVED BY:

__________________________________
Richard Clerkin, Ph.D.

_________________________________
Shevaun Neupert, Ph.D.

_______________________________
Dennis Daley, Ph.D.

_____________________________
James Swiss, Ph.D.
Advisory Committee Chair
Dedication

For my sons, Will, Wes, and Ethan. I hope that throughout their lives they will recognize and appreciate the advantages they have enjoyed because of their education, and that they will devote some small part of their lives to helping others less fortunate than them to gain a similar foundation.
Biography

Rob Hines earned a Bachelor of Arts in History from North Carolina State University in 1994. After working for a little more than a year after graduation, he moved with his wife to Syracuse, NY to attend the Maxwell School of Citizenship and Public Affairs, from which he earned his Master of Public Administration in the summer of 1996.

Following his graduation from Syracuse University, Mr. Hines worked for several years as a consultant in the for federal, state, and local government clients, including the US Customs Service, the US Army, the US Navy, the State of North Carolina, and the city of New Haven, Connecticut. His interests gravitated toward the field of education beginning with a stint working for the North Carolina Department of Public Instruction.

While at DPI, where he became a full-time employee in 2004, Mr. Hines worked in the Technology Services area as well as in Academic Services. During his time there, he completed a North Carolina Education Policy Fellowship and rose to become a director, working closely with schools districts across the state on a variety of initiatives, including school improvement planning. Mr. Hines led the internal agency council that developed the 2009 NC School Improvement Planning Implementation Guide.

Mr. Hines began his doctoral studies at North Carolina State University while working full-time at the Department of Public Instruction. He completed the Doctor of Philosophy in 2017, a little more than two years after leaving the Department of Public Instruction for a position with College Foundation, Incorporated, a non-profit organization tasked with supporting the North Carolina State Education Assistance Authority and higher education across North Carolina.
Acknowledgements

Anyone who completes a doctoral degree inevitably has people to thank for aid and encouragement along the way. The path to completing this degree was winding, sometimes disheartening, and on a few occasions, felt nearly insurmountable. It took many years to get through this and I am grateful I didn’t have to do it alone. I want to offer my heartfelt thanks to a number of people who played a role in helping me finally get through it.

The greatest thanks goes to my wife, Katie. She has supported me in all manner of endeavors during our years together, and this process has been no exception. When I needed it, she gave me a gentle nudge to keep me going, and when I really needed it, she gave me a more forceful nudge! Somehow, she could always tell when I was second-guessing this whole enterprise (which happened more than once, to be sure), and she always reminded me of the investment I’d already made in time and energy, and what a shame it would be for that to go to waste.

My dad, Jim Hines, and my grandfathers, Jay Hines and Bob Waugh, deserve thanks just because they have been wonderful examples to me in my life, and because they have always been supportive of me anytime I needed them, without question and without hesitation.

Chris Cline, Sharon Paynter, Todd Loendorf, and Stan Holt - unquestionably, they were the BEST statistics and comprehensive exam study partners I could ever have found. They helped me so many times in more ways than I can count. Thanks for making it fun…well, as fun as studying for a statistics exam can be, anyway!

I also owe immense gratitude to Dr. Rebecca Garland and to Susan Auton at NCDPI. They always had faith in me, and they never let me forget that I had their support. They were
wonderful colleagues and I’m happy still to count them as my friends. Thank you both for everything you have done for me!

Many thanks also go to Pat Kaiser, Shannon Byers, John Noschese, Marty Flournory, and Kathy Schafer for their encouragement and their willingness to pick up my slack so that when I really needed to I could focus on completing this project. Thanks especially to Shannon and Kathy for “talking me down off the ledge” a few times when it just felt like there was no way to get through!

Last, but certainly in no way least, thank you to each of my committee members, Dr. James Swiss, Dr. Shevaun Neupert, Dr. Rich Clerkin, and Dr. Dennis Daley. Each of them brought their wonderful and unique strengths to the classrooms where they taught me as well as to my committee. They were always ready to provide sound advice, incisive questions, and positive support when I needed it. They also have incredible patience, and I appreciate the support each provided. As I go forward, I will try my best to bring to my work the many lessons they taught me. Thank you!
Table of Contents

List of Tables ........................................................................................................................................... xii
List of Figures ........................................................................................................................................... xiii

Chapter 1: Introduction ............................................................................................................................... 1
  Planning for Better Learning ..................................................................................................................... 3
  Theoretical Relevance of School Planning Effects ............................................................................... 4
    Key Questions ..................................................................................................................................... 4
  Practical Relevance of Understanding School Improvement Effects .............................................. 5
    Relevance to Legislatures ................................................................................................................... 5
    Relevance to School District Administrators ..................................................................................... 6
    Relevance to School Principals and Staff ........................................................................................... 6
    Relevance to the Academy ................................................................................................................ 7
  Overview of Research Questions .......................................................................................................... 7
  Overview of Data Sources for Dependent and Independent Variables ............................................. 9
    EVAAS Data ....................................................................................................................................... 10
    North Carolina Teacher Working Conditions Survey ..................................................................... 10
    School Improvement Plans ............................................................................................................... 11
  Statistical Methods .............................................................................................................................. 11
  Overall Organization of the Study ........................................................................................................ 12

Chapter 2 – Review of the Literature ..................................................................................................... 13
  Performance Management in a School Context ................................................................................ 13
  The Purposes of Performance Measurement ..................................................................................... 15
  Building Performance Management Systems in Schools ............................................................... 16
What the Literature Suggests About Performance Management Systems ............................... 16

PM System Theory of Change: The Value of Limited Focus ........................................... 17
PM System Theory of Change: The Value of Providing Clarity........................................ 19
Clarity and Goal-setting ..................................................................................................... 20
Clarity and Target-setting .................................................................................................. 22
PM System Theory of Change: The Value of Feedback .................................................. 25
Feedback and Performance Improvement ........................................................................ 25
Feedback and Performance Improvement in Schools ...................................................... 30
Performance Management Systems and Organizational Performance ......................... 38
Summary of the literature .................................................................................................. 46

Chapter 3 - Research Design and Statistical Methods ..................................................... 47
Research Design ................................................................................................................ 47
The Model .......................................................................................................................... 49
Sample Selection ............................................................................................................... 50
Explaining the Focus on Elementary and Middle Schools .............................................. 50
Total Population and Sample Size .................................................................................. 51
Data Sources ..................................................................................................................... 51
Education Value-Added Assessment System (EVAAS) .................................................... 52
School Improvement Plan Analysis .................................................................................. 53
North Carolina Teacher Working Conditions Survey (TWC) ........................................ 54
Control Data Sources ........................................................................................................ 56
School Report Cards ......................................................................................................... 56
Grade, Race, and Sex Data ................................................................................................. 58
Free and Reduced Meals (FARMS) Data ........................................................................ 58
Variables .......................................................................................................................... 59
Dependent Variables ....................................................................................................... 59
School-wide Accountability Growth Estimate .............................................................. 59
SAGE 3-year Trend ......................................................................................................... 61
Independent Variables ................................................................................................... 61
Independent Variables Measuring Focus ..................................................................... 62
Independent Variables Measuring Clarity .................................................................... 63
Independent Variables Measuring Goal Difficulty ....................................................... 64
Operationalizing Goal Difficulty ................................................................................. 66
Independent Variables Measuring Feedback ............................................................... 67
Control Variables .......................................................................................................... 68
Average class size .......................................................................................................... 69
School size .................................................................................................................... 70
Teacher experience ........................................................................................................ 70
Percentage of teachers with an advanced degree ......................................................... 71
Composition of the student body .................................................................................. 72
Percentage of National Board Certified Teachers ...................................................... 72
Statistical Methods ........................................................................................................ 73
Error Term Gauss-Markov Assumptions .................................................................... 74
Considerations Regarding Potential Flaws in Research Design ................................ 74
Summary ......................................................................................................................... 76
Chapter 4 – Analysis and Results ................................................................................ 78
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection and Preparation</td>
<td>78</td>
</tr>
<tr>
<td>Dependent variable data</td>
<td>78</td>
</tr>
<tr>
<td>Independent variable data</td>
<td>78</td>
</tr>
<tr>
<td>School improvement plan collection and analysis</td>
<td>79</td>
</tr>
<tr>
<td>Teacher Working Conditions Survey data</td>
<td>85</td>
</tr>
<tr>
<td>School performance and target data</td>
<td>86</td>
</tr>
<tr>
<td>Missing Values and Data Imputation for Independent Variables</td>
<td>86</td>
</tr>
<tr>
<td>Control Data</td>
<td>87</td>
</tr>
<tr>
<td>Reliability</td>
<td>87</td>
</tr>
<tr>
<td>Validity</td>
<td>88</td>
</tr>
<tr>
<td>Internal validity</td>
<td>88</td>
</tr>
<tr>
<td>Selection Bias</td>
<td>89</td>
</tr>
<tr>
<td>Construct validity</td>
<td>89</td>
</tr>
<tr>
<td>Face validity</td>
<td>90</td>
</tr>
<tr>
<td>External validity</td>
<td>90</td>
</tr>
<tr>
<td>Data Screening</td>
<td>90</td>
</tr>
<tr>
<td>Final Regression Model and Regression Equations</td>
<td>97</td>
</tr>
<tr>
<td>Regression Analysis Results</td>
<td>99</td>
</tr>
<tr>
<td>Structure of the Hierarchical Regression Analysis</td>
<td>100</td>
</tr>
<tr>
<td>Hierarchical Regression Results for School Performance</td>
<td>101</td>
</tr>
<tr>
<td>Hierarchical Regression Results for Change in School Performance</td>
<td>105</td>
</tr>
<tr>
<td>Discussion of Results</td>
<td>111</td>
</tr>
<tr>
<td>Summation of the Hierarchical Regression Predicting School Performance</td>
<td>111</td>
</tr>
</tbody>
</table>
Other Limitations ............................................................................................................. 131
Summary of Limitations ................................................................................................. 133
Opportunities for Future Research ................................................................................. 133
List of Tables

Table 4.1 – Correlations of Independent Variables .......................................................... 92
Table 4.2 – Correlations of Independent Variables .......................................................... 92
Table 4.3 – Independent Variables with Near-zero Variance........................................... 96
Table 4.4 - Hierarchical Regression Results Predicting School Performance – Blocks 1-3 ..... 104
Table 4.5 - Hierarchical Regression Results Predicting School Performance – Blocks 4-6 ..... 105
Table 4.6 - Hierarchical Regression Results Predicting Change in School Performance – Blocks 1-3 ................................................................................................................................. 109
Table 4.7 - Hierarchical Regression Results Predicting Change in School Performance – Blocks 4-6 ................................................................................................................................. 110
List of Figures

Figure 1-1. Initial Overall Research Model ................................................................. 9
Figure 3-1. Initial Overall Research Model ................................................................. 49
Figure 4.1 – Plot of Residuals and Predicted Values for Dependent Variable sage1415 ....... 94
Figure 4.2 – Plot of Residuals and Predicted Values for Dependent Variable 3yrtrend .......... 95
Figure 4.3 – Final Overall Research Model ................................................................. 97
Chapter 1: Introduction

More than thirty years ago, the National Commission on Excellence in Education sounded an alarm over the state of education in America with the release of *A Nation at Risk* (1983, ANAR). Metric after metric indicated the severity of the problem: high levels of functional illiteracy, declines in standardized test scores, a lack of higher-order thinking skills, declines in science achievement, and high costs of remedial education in business and industry (ANAR, pp. 8-11). The Commission lamented the relative decline of US students in comparison to students in other industrialized nations, citing the fact that on 19 tests of academic achievement, American students never placed first or second, and on seven tests they placed last (ANAR, 1983, p. 8). The loss of American prominence internationally reappears throughout the report, and in the years since the business, education, and policy communities have opined this same concern nearly continually.

Even today the US is not competitive for the top positions in a variety of measures. The most recent Trends in International Mathematics and Science Study (TIMSS, 2015) report indicates that the average mathematics scale score of 539 for American fourth-graders placed the US 14th in the world – statistically significantly lower than six Asian and four European countries. The US average was also statistically significantly lower than that of one US state, Florida, which participated as an independent educational system, and the US was statistically tied with eight other international education systems (Mullis, Martin, Foy, & Hooper, 2016, p. 22). American eighth-grade students performed at a level that placed them 10th in the world and statistically tied with nine other nations. While the average mathematics score for US students grew by 21 points (4th grade) and 26 points (8th grade) between 1995 (the first year of the TIMSS
assessments) and 2015 (Mullis et al., p. 29), the average science scores for US fourth-grade students in 2015 were not statistically different from those in 1995 (Mullis et al, p. 37). With respect to science, the 2015 TIMSS results placed US fourth-graders 10th in the world among other countries and education systems administering the test and statistically tied with six others, two of which had a higher average scale score (Mullis et al, p. 22). American eighth-graders place 11th in the world, statistically tied with seven other educational systems.

Finally, the Program for International Student Assessment (PISA) 2015 test, which assesses the reading, math and science literacy of 15-year-olds, places the US 24th in reading with a scale score that is statistically average at 497 points, and only slightly above the Organization for Economic Cooperation and Development\(^1\) (OECD) mean of 493 points (OECD 2016, p. 151). This reflects essentially no change since the last PISA administration in 2012, when scores placed the US in exactly the same position relative to other participating nations. Twenty international education systems had higher average scores than the United States in all three areas assessed by PISA. It is within this context of mediocrity that American school systems for decades have sought ways to raise student achievement.

The value of K-12 education to American society is undeniable. In fiscal year 2013 (the latest data available as of May 2017), US federal and state governments combined spent over $600 billion on K-12 education (National Center for Education Statistics, 2017). This level of investment reflects a belief that education is critical to ensuring a bright future for the nation and its citizens. Many Americans see education as an economic development tool, providing citizens the skills and knowledge needed to become marketable in the domestic and international economies. In addition, a sound education provides the intellectual foundation Americans need

---

\(^1\) The Organization for Economic Cooperation and Development (OECD) is an organization of democratic governments around the world whose mission is “to promote policies that will improve the economic and social well-being of people around the world.” (OECD website, July 24, 2016)
to participate effectively in our democracy and in a world that is changing like never before. It can be argued that ultimately, education builds an individual's capacity to succeed and, by extension, the capacity of our society to prosper. If this is true, it helps explain our preoccupation with improving our schools and our seemingly perpetual search for better methods of increasing student learning.

**Planning for Better Learning**

In school districts across the nation, one of the methods applied for improving school performance is a formal school improvement planning process. School improvement planning is largely based on the concept of performance management, an overarching approach to management based on setting clear directions through long and short term goals, and then tracking progress regularly. Performance management begins with strategic planning, a technique used in American businesses for many years. This tool creates organizational focus and reinforces organizational culture. The annual goals arising out of the planning process map a common direction for members of the organization, guiding day-to-day decisions about how to conduct business. Inclusive and thoughtful planning processes result in shared goals and strengthen organizational culture through stakeholder interactions. Empirical studies by numerous researchers have provided evidence that goal-oriented planning, combined with feedback on progress, can often boost organization performance.

The essential research question related to school improvement planning remains “Is there a relationship between specific planning practices and student and school outcomes?” In particular, do such tools as clear, specific goals, and school-specific performance feedback lead to increased school performance? These questions are explored here in the context of North Carolina school performance using a measure of school “value-added.” Exploring these
questions and building our knowledge base around school planning and performance offers both theoretical value for the education and management fields and practical value for the public.

**Theoretical Relevance of School Planning Effects**

Theoretical questions arise in a discussion of planning and management and its applicability to schools. An effective planning and performance management process necessarily requires establishment of performance goals. Previous research shows that goal-setting – conscious and deliberate establishment of a desired performance level and strategies for achieving it – aids in raising performance. This finding is robust across a number of environments and industries. Goal-setting effects have been shown to be applicable to groups, organizational units and entire organizations (Locke & Latham, 2002), but have only recently begun to be assessed in a school setting.

**Key Questions**

Several important questions arise with respect to the specific components of planning that aid in a school performance improvement context. Does a plan that focuses effort by limiting the number of goals and strategies help to improve performance? Can clarity of measurement support school-level improvement activities? Should difficult-but-not-impossible (“stretch”) goals be expected to lead to greater effort on the part of the school community? Feedback offers the possibility of informing performance and adjusting effort; is it critical in the complex environment of our schools? In other words, do the core findings around the importance of these specific planning components hold true in a school environment in the way previous research demonstrates for other organizations? As in any environment, principals’ management approaches vary along a continuum from tightly to loosely controlled and from formal, individual leader-driven to shared management approaches. Whether or not a particular
perspective or approach is more useful than another for the specific needs of schools interested in improved performance is an important aspect of both management and education theory and is particularly important for the field.

**Practical Relevance of Understanding School Improvement Effects**

Understanding how school management approaches affect school performance offers practical relevance to politicians and agency staff, district-level school leaders, and school administrators. In North Carolina, Article 8B of General Statute 115C-105 establishes the School-Based Management and Accountability Program. This legislation, passed by the NC General Assembly in 1996, was one of the first state education accountability laws in the United States. Designed to improve student performance, the law uses the two distinct components of school-based management and accountability and codifies the implementation of school-based management, devolving significant decision-making power to the school level, and directing local school boards to provide “maximum flexibility to schools in the use of funds” to accomplish school goals (NCGS 115C-105.25). This legislation also establishes the legal requirement for each public school in North Carolina to develop a school improvement plan and defines the constitution of a school improvement team, specifying administrative, faculty, and community representation on the team.

**Relevance to Legislatures**

The same statute cited above provides broad authority to the State Board of Education to develop and implement an accountability program for public schools. The recently revised version of this accountability program is known as the “Ready model.” The legislation of improvement planning and accountability suggests a high degree of faith among state lawmakers that such programs will improve education across the state. This begs the questions as to
whether the process actually provides the expected value such that legislators should implement or maintain a planning requirement to help ensure achievement standards are met. A demonstration of clear positive effects of planning on achievement makes a powerful argument for legislating a school improvement planning process.

Relevance to School District Administrators

With respect to school improvement planning, a key interest at the school district level is the question: how can local school administrators best implement a planning process and support schools in executing their plans? In this era of near-perpetual shortage of resources, district administrators want to obtain the best possible return (student learning and school performance) on the resources they invest in schools. Understanding best practices and the value of detailed planning can be critical for focusing these resources and supporting schools throughout the process.

Relevance to School Principals and Staff

Bryson (1995) noted that organizations with planning teams that are familiar with and committed to the process can complete most of the steps in his Strategy Change Cycle using a 3-4 day process spread over a few weeks. He also stated that planning processes applied to broader communities likely require “considerably more” time because of the broad involvement required (p.23). Schools tend to reflect such broader communities, and within a school, the relevance of planning relates largely to time and to the delicate balance of priorities. Good planning and implementation consume significant time: establishing a planning team, identifying stakeholders and understanding their perspectives, looking together for specific opportunities, threats, strengths, and weaknesses, and developing long-term goals linked to short-term actions and
performance measures. The process requires additional time for data collection and analysis for measuring performance, and for adjusting plans to remedy unsatisfactory performance.

Those with first-hand experience working in a school understand that time is severely limited. Conditions can and often do change rapidly and many different factors impact student learning. Ideally, principals and assistant principals function as the instructional leaders within schools and complete a variety of activities on any given day – conducting classroom observations, providing instructional guidance to teachers, interfacing with parents and district-level staff, and addressing discipline issues, to name a few. Time represents a precious commodity in a school, and principals need to know whether the significant time investment required of planning will pay off. Their primary question is one of return on investment – similar to the investment banker's or the CEO's - “What is the performance return on this investment of time spent planning?”

Relevance to the Academy

The present study also adds to the literature by providing an analysis that covers a different context than some other studies. Sun and Van Ryzin (2012) studied similar tools, but they point out that their analysis of performance management focuses on New York City schools and that contextual factors may hamper generalizability. The present study offers a state-level view of school improvement planning.

Overview of Research Questions

Poister and Streib (2005) asserted that strategic planning “is an action-oriented type of planning that is useful only if it is carefully linked to implementation – and this is often where the process breaks down” (p.46). The current analysis assesses the link between planning and implementation for a subset of North Carolina schools to determine whether specific dimensions
of planning practice relate to overall school performance. It also attempts to demonstrate the applicability of generally accepted best practices for planning in a school environment. This analysis considers the two primary components of the process – initial establishment of the school improvement plan and execution of the plan – and whether a relationship exists between what schools do and how their students perform. Specific research questions include:

**Research question 1:** Do schools with a more focused school improvement plan, as evidenced by smaller numbers of plan goals and strategies, perform better?

**H1:** The number of goals and strategies in an improvement plan is inversely related to overall school performance.

**Research question 2:** Do schools with plans that offer greater goal clarity, shown by greater proportions of goals having clear targets and completion dates, perform better?

**H2:** A positive relationship exists between clear goals and school performance.

**Research question 3:** Do stretch goals result in higher levels of school performance than either very easy goals or extremely difficult goals?

**H3:** Stretch goals support higher performance. Relative to stretch goals, very easy goals and very difficult goals are associated with lower performance.

**Research question 4:** Do the presence of feedback mechanisms and perceptions of feedback effectiveness impact school performance?

**H4:** A positive relationship exists between both the presence and perceptions of feedback mechanisms and school value-add.

These research questions and the literature imply a model where overall school performance, as measured by the school value-added assessment, varies based upon the practices in place for
developing and executing an improvement plan within an individual school building. This model is represented in Figure 1.1.

**Figure 1-1. Initial Overall Research Model**

**Overview of Data Sources for Dependent and Independent Variables**

This quantitative analysis draws data from three primary sources: the Education Value-Added Assessment System (EVAAS), the North Carolina Teacher Working Conditions Survey (TWC), and school-level improvement plan documents. Each provides an important insight into the model components.
**EVAAS Data**

Dependent variable values are drawn from the Education Value-Added Assessment System, or EVAAS. This information system uses student-level assessment data to build a longitudinal model of student, school, and district performance. The system uses two separate sophisticated statistical models; one provides a predictive model of student performance (that is, it can be used to predict future performance on an assessment based upon a student’s own previous performance) and the other provides a growth model of student and school performance, which offers a measurement of the amount of growth a student has made over a defined period of time, or that a school has provided to its students, on average. (These are described in more detail later.) School value-add results, based on past student performance, can be compared to actual performance to assess how well schools are teaching students relative to expectations for a given student cohort. Students with as few as two assessment results can be included in the analysis. Over time, students serve as their own controls as additional assessment results are fed into the system each year.

**North Carolina Teacher Working Conditions Survey**

Results from the North Carolina Teacher Working Conditions Survey provides school-level data regarding some of the planning practices and personnel perceptions of these practices in each school. This survey, conducted every two years since 2002, is specifically designed to obtain input from licensed teachers and administrators about a variety of school and teaching conditions as well as other support for educators for the purpose of informing school and district improvement efforts. This research survey consistently yields impressive participation levels. The 2014 version, conducted online between March 10 and April 4, 2014, received over 93,000
responses – about 90% of the licensed teachers and administrators in the state. This version provides some of the data for the present analysis.

**School Improvement Plans**

Finally, analyses of the school improvement plans from each school included in the sample provide snapshots of the output produced by the planning committees at each school. These snapshots offer some insight into the degree of sophistication of the planning process at an individual school and the practices used by the planning teams to produce the plans.

**Statistical Methods**

The present study uses a quasi-experimental sampling design built on data from multiple sources that include a valid, reliable biennial survey of North Carolina teachers regarding school-level working conditions, a statistically-based school performance assessment system used by NC schools and districts (wherein school performance is based upon overall performance of students within the school), and an analysis of school improvement plans from a number of school years, based upon availability. Survey and improvement plan data provide independent variables while the dependent variable is supplied by the school assessment system.

The nature of the data for this study has important implications for the analytical methodology. For the present analysis, both dependent variables reflect a school-level score (based upon the aggregated results of the students in each school), and all other predictor variable and control variable data reflect this same unit of analysis – the school. A multivariate ordinary least squares regression model is the best option for effectively handling this data. Further details are found in the section on statistical methods within Chapter 3.
Overall Organization of the Study

Chapter 1 offered a brief introduction to this analysis.

Chapter 2 provides a review of literature relevant to the research questions identified in Chapter 1 and the specific variables to be used in the analysis.

Chapter 3 offers a description of the statistical analysis in the present study. Research questions are reviewed, data sources are revisited in greater detail, and the dependent and independent variables are described. The statistical methods are detailed.

Chapter 4 provides the analytical results.

Chapter 5 provides a detailed discussion of the analytical results, the implications of these results, and proposes further questions for future analysis based upon the findings.
Chapter 2 – Review of the Literature

The present study considers the impact of several practices critical to school improvement planning success. Such practices reflect effective implementation of performance management systems and can help ensure that teams assess progress toward goals and receive information from the assessment process that is necessary to improve. These practices, executed together in a logical progression, offer a foundation for school operations that can produce higher student performance.

Performance Management in a School Context

Public organizations generally suffer no shortage of work, especially in the current environment where many legislators and citizens expect public agencies to “do more with less.” The regular and frequent aforementioned comparisons of American schools to those in other countries and to each other (TIMSS and PISA testing as well as other test regimens such as the domestic National Assessment for Educational Progress, or NAEP) support an argument for the idea that our public schools are among the institutions most encumbered with high expectations. Yet, our school environments reflect a complicated mixture of actors and activities that affect performance outcomes. School principals act as site-level executives, responsible for managing day-to-day operations, for motivating teachers and other staff, solving teacher and student problems, and acting as customer service agents in dealing with parents as well as state and school district administrators. Teachers perform routine administrative duties daily – maintaining attendance, planning lessons, and supervising student activities – in addition to teaching students. To be effective requires them to understand students and adapt their teaching to content, constraints (such as time), and individual student learning styles. Outside of the schools, experts
abound, ready to offer the latest educational strategy that will address schools’ wide-ranging needs and ensure student learning. (One source, manta.com, a marketing firm, listed more than 15,000 educational consulting companies in the United States in May 2017). Each new initiative offers the promise of success and the risk of diminishing marginal returns. It is a logical conclusion that spreading more improvement initiatives across a finite amount of time necessarily limits the time available for any one initiative, and the resulting overload can drain effort and motivation, and reduce the chance for any given initiative to breed success.

The present analysis proposes that well-executed performance management systems, rather than an array of programs, offer significant promise in this environment. An effective performance management system is typically considered to have three simple components:

1. A set of measurable goals or objectives. (The terms “goals” and “objectives” are often treated as synonyms.) Most of these goals or objectives have annual time frames, and most should focus on outcomes – expected effects on clients – rather than on internal processes.

2. Regular reporting on progress toward these annual goals. Regular feedback often is facilitated by establishing shorter-term milestones or “strategies” that lead to the annual goal.

3. Remedial action taken when performance falls short of the goals.

A number of benefits can be expected from a performance management approach. Straightforward goals offer information about what is to be done, when it should be complete, and how it will be measured, providing clarity to each member of the school community about his or her role in, and the mechanisms to be used for, moving the school toward its vision. Providing feedback during execution adds critical knowledge to the process: directive information about whether the strategies implemented are working and how to adapt them; motivational information to aid in maintaining or raising the level of effort, or both (Payne &
Hauty, 1955; Locke, Cartledge & Koeppel, 1968; Strang, Lawrence, & Fowler, 1978). Feedback gathered at the end of the planning cycle also offers an overall assessment that can be used as input into the next planning cycle. Schools can use performance management systems to establish priorities, to limit focus, to measure impacts, to inform staff, to adapt their approach based upon results and, ultimately, to consciously drive improvement.

The Purposes of Performance Measurement

Development and use of performance management systems (hereafter, often shortened as PM systems) varies widely across the United States and the world. Nonetheless, the intent of such systems is largely the same irrespective of coordinates on a map. Behn (2003) effectively summarized the key purposes of performance management systems, which include:

- Evaluating,
- Controlling,
- Budgeting,
- Motivating,
- Promoting,
- Celebrating,
- Learning, and
- Improving.

These purposes serve multiple constituencies in the public sector (citizens, politicians, staff members, and managers themselves) and can aid in satisfying the needs of more than one. Established goals vary, as do the strategies selected to reach them necessarily vary depending upon which purpose the system is designed to serve. School improvement planning represents one variant of performance management systems and often focuses primarily on organizational
learning for the purpose of improving student performance which, when aggregated, can provide an overall indication of how well schools perform.

**Building Performance Management Systems in Schools**

Performance management systems typically include a planning process that analyzes data and results in identification of organizational goals, strategies to achieve the goals, and measurements designed to communicate progress toward goals. A review of school-level plans demonstrates that levels of sophistication vary; systems may measure inputs, outputs, outcomes, or a combination of these. Measures may be leading, lagging, or both. Within school communities, sophistication of performance management systems frequently depends upon the level of expertise held by principals and planning team members, and this regularly may be impacted by turnover at the school and even rotation of team members over time. This analysis proposes that investing in greater planning expertise within our schools will offer returns in the form of more effective school improvement planning. It supports this proposition based upon the existing body of research in strategic planning and performance management, and aims to add to this existing body of knowledge by focusing on the use of key performance management practices in North Carolina public schools and their impacts on student performance.

**What the Literature Suggests About Performance Management Systems**

Since at least the mid-1950s (Locke & Latham, 2002), research has demonstrated that careful development of performance systems can help improve organizational performance. Studies by Mace (1935) and Lewin (1944) pre-date even that time period. Performance goals provide direction, helping to focus effort on relevant activities. Goals provide motivation - an “energizing function” - and affect persistence, enticing people to work longer and harder to achieve the goal. They also encourage the discovery and use of task-relevant information and strategies. (Locke &
Latham, 2002) Measuring progress against goals offers important information about how well strategies are working; and providing feedback to staff engaged in implementing strategies provides the information necessary to realize the motivational aspects of goal-setting. This section first addresses research around the specific “theory of change” aspects of performance management systems including creating organizational focus, providing clarity, and offering feedback then broaches performance management systems and their impact on organizational performance more broadly. The literature reviewed here includes research focused on public sector entities, including public schools wherever possible.

*PM System Theory of Change: The Value of Limited Focus*

Performance management systems provide direction and help organizations generate focus. Establishing direction and focus includes limiting how much the organization tries to accomplish. A long-time management theorist, Drucker (1980) suggested that trying to do too much at once is a certain path to non-performance due to a lack of established priorities. A number of other researchers also have identified this as an important aspect of goal ambiguity theory because greater numbers of goals make it difficult to determine which are more important, leading to negative consequences for public agencies. (Jung 2011)

Chun and Rainey (2005) supported this concept in an analysis designed to explain organizational performance using measures of goal ambiguity, specifically a measure they call “priority goal ambiguity,” which they defined as the number of goals and performance targets, standardized and transformed into a composite score. The authors measured organizational performance using several dependent variables including managerial effectiveness, customer service orientation, productivity, and work quality. Their dataset included nearly 32,000 total survey responses, a fact that they recognize leads to many statistically significant results with
small effect sizes. The analysis controls for organization size, age, location, competing demands, financial publicness, type of policy responsibility, and problem complexity as well as several respondent-level controls including tenure, pay-grade, job category, and managerial level.

Chun and Rainey (2005) hypothesized a negative relationship between the number of goals and priorities and each of the perceived organizational performance variables, suggesting employees in agencies with higher numbers of goals and targets will perceive lower levels of performance across each of the four performance dimensions.

The authors’ findings were mixed. While the hypothesis predicted a negative relationship between ambiguity and each dependent variable, the analysis determined that goal ambiguity negatively impacted only the measure of managerial effectiveness and the effect size was small despite the significance of the relationship. It showed no significant relationship with the other dependent variables. Nonetheless, Chun and Rainey (2005) assert that the overall findings in combination with those of a prior study support the conclusion that three of the four ambiguity dimensions “offer viable ways of conceiving and measuring organizational goal ambiguity for government agencies.” (p.546) However, the weak effect sizes and low explained variance, as represented by low R-squared values, across the models point to the need to develop still better techniques for measuring performance of government agencies.

Like Chun and Rainey (2005), Jung (2011) hypothesized a negative relationship between number of goals and organizational performance. He, too, cited Drucker’s (1980) assertion that trying to do too many things at once is “guaranteed to produce non-performance.” Jung explained and extended Drucker’s remarks, suggesting that having greater numbers of goals not only increases the difficulty of prioritizing between them, but also divides effort and leads to lower overall performance.
Jung (2011) studied performance in 97 US federal agencies and assessed several dimensions of goal clarity, including the total number of goals. He suggested that in organizations with many goals, achievement of one goal may in fact be counterproductive to another, hindering its achievement. His analysis used actual goal attainment data (the dependent variable and an objective measure) and goal ambiguity data (independent variables) from the 2006 and 2007 Program Assessment Rating Tool (PART). Jung included several control variables, including organization size (based upon number of employees), budget size, institutional location (within or outside an executive department), the type of policy the agency dealt with (regulatory or non-regulatory), agency age, proportion of professional employees, and whether the agencies were managed by presidential appointees or by career civil servants. His analysis found a low yet significant negative correlation between the number of goals and organizational performance. Jung’s ordinary least squares regression model supported his hypothesis of a negative relationship between the number of goals and organizational performance, indicating that additional goals can be expected to reduce organizational performance.

Despite the mixed results found in the analyses discussed above, the present study proposes to find a negative relationship between the number of goals and strategies and organizational performance.

**PM System Theory of Change: The Value of Providing Clarity**

Performance management systems can aid organizations’ execution by providing a clear picture of organizational priorities and expectations. Intuitively, one can argue that having the employees and managers of any organization “on the same page” with respect to the purpose and direction of the company or other entity should support the achievement of those purposes.
Several researchers have assessed clarity as a component of PM system research, speculating that increased clarity offered by a well-designed PM system will raise organizational performance levels.

**Clarity and Goal-setting**

Moynihan, Pandey, and Wright, in their 2011 analysis of transformational leadership, suggest that transformational leaders build a foundation (“set the table,” in their parlance) for effective use of performance information in an organization, in part, by establishing clear goals, which they identify as a mediating factor in successful performance management. Using a quantitative analysis and structural equation modeling, they determined whether observed effects were consistent with their proposed theory of action, namely the presence of indirect relationships between variables through mediating factors. They suggested that a clear vision fosters clear goals and these in turn lead to use of performance information that provides leaders “a better sense of which tasks are critical, their relative importance, and how they can be achieved.” (p.148) Such knowledge can improve performance. They hypothesized that transformational leadership would positively affect the use of performance information and specifically suggested that the clarity offered by goals would mediate this relationship.

The authors conducted their analysis using data from a national survey of senior local government managers known as the National Administrative Studies Project (NASP-IV). The sample included over 3,300 individuals, of which about 47 percent responded. Their variables included performance information use as the dependent variable and measures of transformational leadership, goal clarity, and developmental culture as independent variables. Their two-stage structural equation model confirmed direct effects of transformational leadership on goal clarity and indirect effects through goal clarity on use of performance information,
theoretically leading to greater organizational effectiveness. All of their hypothesized paths were statistically significant in the predicted direction. This analysis confirmed the importance of clear goals in supporting the use of performance information, a key ingredient in effective organizational performance management.

Hu and Liden (2011) focused on the role of goal clarity in supporting team potency and performance. The discussion reinforced Moynihan, et al (2011) and the concept of organization goals cascading down to individuals. They suggested that an individual’s ability to complete his/her role and achieve goals requires knowledge of the link between individual tasks and the team’s or organization’s goals. Goal clarity “heighten[s] individuals’ understanding of task goals….but also highlights individuals’ connections to co-workers, teams, and the organization,” and this greater understanding and connection in turn leads to better communication and integration of the team’s work. (p.852) Hu and Liden hypothesized a positive relationship between goal clarity and team performance through the partial mediating effect of team potency.

Hu and Liden (2011) surveyed employees and managers across 95 teams in five Chinese banks to assess goal clarity and team potency. Responses were provided using a seven-point Likert scale. Managers rated team performance using a separate four-item survey, also with a seven-point scale. Hu and Liden controlled for average team tenure, team age, organizational membership, and organizational tenure. All analyses were conducted at the team level, but hierarchical linear modeling was used due to potential non-independence of ratings since pairs of managers were rating multiple teams’ performance. Hu and Liden’s results supported their hypotheses, indicating a significant, positive relationship between clear goals and team performance.
A study by Chun and Rainey (2005a) proposed *evaluative ambiguity* (lack of clarity) as an indicator of difficulty in evaluating progress toward organizational goals. They defined this construct as the percentage of subjective or workload-oriented performance indicators contained within an agency plan. Subjective indicators were defined as any measure based upon individual perception of performance or condition, or without a numerical measure. Workload-oriented indicators were defined as those based upon inputs or outputs as opposed to being outcome- or results-oriented. The authors hypothesize that federal employees in agencies with higher levels of evaluative goal ambiguity would perceive lower levels of organizational performance.

Results of the authors’ regression analysis indicated significant negative relationships between evaluative ambiguity and measures of organizational effectiveness, as predicted. Chun and Rainey (2005a) suggested that the finding “supports the frequent assertions…about the effects of the lack of clear performance indicators on organizational performance.” (pp.547-548) They conclude that goal clarity is good for performance and suggest one interpretation may be that well-designed performance management initiatives can benefit public sector organizations.

*Clarity and Target-setting*

Boyne and Chen (2007) addressed target setting in an analysis of performance management systems that is relevant to the present study, which includes measures of target establishment in performance planning. They assessed the impacts of target-setting programs, including the general use of targets, the number of targets, and the difficulty of targets, on student performance. Their analysis assessed student performance in 147 English localities over the period from 1998 through 2003, which included a pre-target-setting period and a post-target-setting period. Boyne and Chen classified each authority into target-oriented and non-target-oriented groups, allowing them to compare those with targets to themselves during the pre-target
period (prior to 2001) as well as to those without established targets. They hypothesized that both the presence (having targets vs. not having targets) and difficulty of targets relate positively to performance while the number of targets relates negatively to the same.

Boyne and Chen (2007) used a fixed-effects model and found seven of eight target variables positively related to performance and statistically significant at the 5% level or better, demonstrating that authorities with targets established “performed better than their peers in the LPSA [target-setting] period and better than themselves in the pre-LPSA period.” (p. 472) However, the authors’ analysis of target difficulty did not support their hypothesis; none of the “stretch,” or difficult-but-not-impossible, target models performed significantly better than their routine target-setting models. Therefore, the increased performance could not be attributed to the target difficulty.

With respect to the number of targets, the authors anticipated finding a negative relationship with performance. Their analysis found the opposite; none of their eight coefficients were significantly negative. In fact, six of the eight were significantly positive, suggesting that higher numbers of targets lead to better performance. Boyne and Chen offered two possible explanations. First, they suggested that perhaps the range of data is too small to accurately assess the relationship. The largest number of targets was six, which may not have reached the threshold at which negative effects may be seen. Second, they allowed that perhaps number of targets is not a reliable measure and instead serves as a proxy for the priority of education improvement within the local authority. The implication, therefore, is that higher numbers of education targets indicate a greater focus on improving education and perhaps a greater effort toward achieving positive outcomes. The similarity of Boyne’s and Chen’s hypothesis to the
In their analysis of public duty motivation in the US federal government, Jung and Rainey (2011) aligned with prior research defining target-specification ambiguity as unclear statement of performance targets. The authors measured target ambiguity as the proportion of all stated goals in the 2006 Program Assessment Rating Tool (PART) that do not have clear performance targets. They hypothesized that the level of target specification ambiguity relates negatively to employees’ motivation, a key mechanism by which goal-setting affects performance.

The authors used a multinomial logistic regression analysis to assess the ambiguity-motivation relationship. Their results for the target-specification ambiguity variable were not significant, indicating that increased target ambiguity has no impact on the likelihood of lower motivation, thus their hypothesis was not confirmed. However, their analysis used three non-intuitive categories of scale response, despite their survey asking a relatively direct question to assess the item. Given this measurement approach, greater weight should be placed on other studies which have more direct measurement of effects and which also show significant results.

Overall, analytical results related to clarity were mixed, with researchers finding support for the idea that goal clarity supports higher performance whereas analyses of target clarity offered less emphatic support. The present analysis is expected to provide results that support the importance of clear targets and to add further evidence and support for the importance of goal clarity to improving performance.
PM System Theory of Change: The Value of Feedback

Many data points are established during the course of a school year. Schoolteachers encounter daily opportunities to establish how well students are learning course content and whether or not their knowledge, skills, and abilities are meeting established standards. Data result from formative assessments, which can occur multiple times for every student during the course of a class period, from benchmark assessments (periodic assessments designed to measure learning over the course of a learning unit), and from summative assessments that typically occur at the end of large periods of time and assess against a large number and variety of objectives (these are frequently end-of-year assessments). For an individual student these data provide a source of feedback about what she knows and is able to do with what she knows. It offers teachers input about how to help the student achieve learning objectives and standards. By aggregating the data from its students, one can glean a picture of a school and its overall effectiveness. These student and school-level data provide valuable feedback for input to the school improvement process.

Feedback and Performance Improvement

A significant volume of research suggests the relevance and the importance of feedback for improving performance. Much of the early work in assessing feedback focused on “clear technology” types of activities - psychomotor functions and other short-duration actions assessed based upon time and/or accuracy measures and where feedback was simple to give and easy for participants to interpret. Despite their simplicity, they offer insights into the foundations of how feedback impacts performance. A smaller volume of research has focused on longer-term studies and fields where cause-and-effect relationships are less clear. This section addresses feedback by
first summarizing the general case for the importance of feedback for improving performance and then focusing on recent research specifically related to feedback and its role in schools.

Taken together, the data resulting from the numerous studies over many years indicate that providing feedback can help improve performance and that feedback in combination with goal-setting offers one of the best mechanisms for improving performance. Payne and Hauty (1955) identified two key mechanisms through which feedback operates: by providing directive information (guiding or cueing factors that suggest what is wrong and how to correct it) or motivational information (competitive factors that induce people to work harder or longer). Their study assessed the effects of both kinds of feedback on performance – specifically on work decrement, or performance reduction – and their results indicated that both types significantly affected performance, demonstrating that performance under a treatment condition with non-specific feedback was significantly higher than under control conditions and that performance with specific feedback was still higher. Similar results were found for motivational feedback – performance variation attributable to this kind of feedback was highly significant and statistically independent from the other experimental conditions.

A later meta–analysis by Locke, Cartledge, and Koeppel (1968) focused on the motivational effects of feedback, hypothesizing that these effects actually reflect goal-setting by subjects in response to feedback received. While the authors suggested that prior studies finding feedback effects were measuring the wrong thing, their analysis of previous studies demonstrated a pattern showing that feedback improves performance. Across a series of studies presenting a variety of conditions, study participants who received feedback typically performed better, as defined in the studies, than those who received no feedback.
Hamner and Kim (1976) assessed the external validity of Locke’s theory of goal setting using a quasi-experimental analysis designed to investigate feedback and goal setting effects on performance. Three experimental groups received goals related to cost performance, safety, and service, and either extrinsic (supervisory) feedback, intrinsic (self-generated) feedback, or both types of feedback. A fourth control group received goal setting instructions only. The authors’ analysis indicated effects of goal setting alone, thus yielding support for Locke’s theory. However, the results indicated greater effects, beyond those of goal setting alone, when feedback of either type was added. This finding held true across all cost and safety goals, but not service goals. However, Hamner and Kim (1976) point out in their discussion that “while service improved across all groups, the greatest amount of improvement occurred in the external feedback plus praise groups; again indicating that goal setting plus external feedback and praise is superior to goal setting alone in bringing about improvements in performance.” (p.55)

Erez (1977) went further, suggesting that feedback is a required component of improvement efforts. Citing Locke and other colleagues, she stated that previous research implied “that goals cannot be set meaningfully unless subjects have some knowledge of how hard the goals are and how they are performing in relation to their goals.” (p.624) Erez tested her hypothesis using an experimental design in which the experimental group received performance feedback while the control group received none. Erez predicted that goals would correlate to performance only, or at least more strongly, under conditions of high feedback, and that goals would not correlate to performance, or would correlate less strongly, under conditions of low knowledge of results (goals and feedback interact to affect performance). Her empirical tests confirmed her hypotheses and led her to conclude that feedback is a necessary condition for goals to affect performance. Her findings offer some confirmation of a logical relationship between feedback,
goals, and performance: even in the presence of clearly-defined goals, regulating action and effort toward goal achievement requires information about the results.

While the above studies make a general case for the value of feedback in improving performance, none are especially well-suited to generalizing for an environment like a school, where the tasks are more complicated and time horizons are longer and stretch across months rather than hours. Becker’s (1978) study more nearly addressed the time factor in his study of energy conservation. Becker’s experimental study better addressed the impact of feedback on activity over time by examining feedback in a field setting, engaging one hundred households in reducing energy consumption over a month’s time during the summer. Becker asked 80 families to reduce electricity consumption for several weeks during the summer. Forty families had the difficult goal of a 20% reduction; the other half received an easy 2% reduction goal. Within each of the two groups, half of the families were given feedback three times a week about their consumption. The remaining twenty families served as a control. Becker’s results confirmed his prediction. The difficult goal/feedback group conserved the most electricity and was the only group that used significantly fewer kilowatt hours than the control group leading Becker to the conclusion that rather than goal-setting alone leading to a performance improvement, goals along with feedback relative to the goals leads to results.

Becker’s analysis is relevant because it suggests that feedback is important to performance over time and making improvements in the educational quality of a school is a long-term proposition. In fact, given the volume of research and related findings, the impact of feedback seems clear and one might expect them to be applicable to the school improvement environment. Coe (2002) however, suggests that the reality is decidedly less clear, offering the following apt description of the state of feedback research: “…the plausible view that feedback generally
enhances performance is still prevalent in the literature. Nevertheless, a closer examination of the
evidence reveals a far more complicated picture: feedback is by no means always beneficial in its
effects, and identifying the conditions under which it may be expected to improve performance is
far from straightforward.” (p.4) Coe, whose own doctoral research suggested positive effects of
feedback on student performance, also states that his “main finding...seems to be that the effects
of feedback are extremely complicated.” (Coe, 1998, p.243).

Kluger and DeNisi (1996) conducted a meta-analysis which reinforced the inconclusive
nature of the feedback research. The authors’ assessment included a review of more than 600
effect sizes over more than 23,000 observations. Inclusion in their analysis required several
things: at least one treatment group that received feedback but no other confounding
manipulation; at least one control group that received no feedback intervention; measurement of
performance as opposed to mere discussion of performance; ten or more participants; and,
enough information to calculate a Cohen’s statistic ($d$): means and standard deviations, $t$ values,
or exact F values and all relevant means. While their initial body of works reviewed was
approximately 3,000 papers, the authors’ inclusion requirements led to 131 papers (5%) being
included. Further review, however, indicated that this number represented 10-15% of the
empirical feedback intervention literature. The authors calculated Cohen’s statistics and
categorized the results. Positive $d$’s reflected performance gains due to a feedback intervention
whereas negative values reflected performance declines. Their assessment is that while many
positive effects have been found in prior research, there were also “large and often ignored
variability in FI [Feedback Intervention] effects.” (p.274-275) Their results indicated that
feedback interventions overall improved performance on average by about .4 of a standard
deviation ($d = .41$), but that more than a third of the interventions actually decreased
performance.

**Feedback and Performance Improvement in Schools**

In the years since Kluger and Denisi (1996) and Coe (2002) provided their assessments of the state of feedback research, investigation has continued but little has occurred that has altered the state of things, particularly with respect to operations of feedback in schools. While volumes of research on feedback have been conducted over many decades, empirical research on feedback and school performance is limited (Coe, 2002; Schildkamp & Teddlie, 2008, Verhaeghe et al, 2010). Of the research that has been done specific to the school environment in recent years, these results, too, have been mixed. A few studies have shown that feedback can serve a valuable role by providing information that allows targeted intervention with individual students or whole-school approaches that treat the entire school community. (Hammond & Yeshanew, 2007; Schildkamp & Visscher, 2010) Other studies have been less clear about the usefulness of feedback. On balance, research on feedback in schools remains nuanced based upon the complexities of feedback systems and the state of the art. Recent research around feedback conducted in the Netherlands, the United Kingdom, and the United States has focused on use of feedback, and perceived effects of using feedback data; only a very few have assessed direct relationships between feedback and performance. Schildkamp and Visscher (2010) conducted case study research into the use of performance feedback in Louisiana schools. Specifically, the authors analyzed the perceived effects of performance feedback as part of the School Analysis Model (SAM). This improvement framework is designed to improve school quality, including student achievement. (p. 1390) The authors drew five schools from a sample of 27 schools in a larger study that divided the schools into three groups: a low-/non-use of SAM feedback group, an “average” use of feedback group where results were discussed but did not drive action, and an
intensive use of feedback group where SAM results were used to take action in the school. Each group was represented within the smaller 5-school sample chosen for detailed analysis in this study.

For each of the five schools, the authors gathered information about school context, School Analysis Model feedback, and use of that feedback (including a questionnaire and interviews with staff) and perceptions about the effects of feedback. At all five schools, staff discussed SAM feedback and wrote an improvement plan, but only three implemented the feedback-based improvement activities. Schildkamp & Visscher (2010) found that in these three schools, staff experienced positive effects: new insights into the school, improvements in student assessment, instructional improvements, management improvements, and greater cooperation in the school. In the two schools where improvement plans were not implemented, few positive effects were perceived by staff; in fact, the staff in these schools pointed to the negative effects of greater stress as a result of the SAM feedback. The positive impacts of feedback may help promote better performance and the authors cite this knowledge of results as “a performance promoting factor” in line with many other studies. (p. 1401)

Verhaeghe et al, (2010) assessed the results of feedback use as part of a larger analysis designed to develop a new school performance feedback system in the Netherlands. The authors based this study on trial versions of school-level feedback reports that offered schools information about student and class performance in mathematics and reading during the first 2 years of primary education. The qualitative analysis design, “recommended when the knowledge base is limited and the nature of variables, processes, and interrelations is less clear” (p.172), focused on principal perceptions of performance feedback use. Data gathering took the form of semi-structured interviews of principals using open-ended questions based upon a conceptual
framework that included school context, school – and user – related characteristics, feedback characteristics, support to feedback users. The principal sample was selected using theoretical sampling based on conceptual rather than representative grounds and based upon degree of feedback use, number of non-special needs students, self-evaluation experience, and school performance as defined in the feedback reports. Interviews were transcribed verbatim and independently coded by two researchers (inter-rater reliability of coding was .90). With respect to the results of feedback use, the authors found that instrumental use of feedback, defined as taking action based upon the information, was rare. However, the information was used to inform reflection and discussion within some schools, occasionally affirming existing impressions. Feedback was used symbolically “in highlighting existing opinions and underlining various problems in the school’s functioning.” (p.177) There was also use of feedback for strategic purposes – developing self-evaluation reports required by education authorities. Finally, the authors also found that some principals used report data to motivate school staff. One important point made by principals in this study was that feedback results offered little to teachers with respect to student-focused action, that is, the aggregate-level nature of the data did little to inform classroom instruction and therefore, presumably, to affect student performance. On balance, Verhaeghe et al concluded that feedback, at least in this case, was little used and had little impact.

Vanhoof et al, (2011) conducted an analysis of effects of feedback use by gathering self-reports of school improvement effects, a method “successfully applied in previous studies on data use.” (p.143) Specifically, the authors established experimental and control groups with random assignment of schools to a group. Each of the schools in the two groups received standardized feedback reports as part of their participation in a larger, longitudinal Dutch study.
Feedback reports addressed math, reading and other subjects and included learning gain and value-added information. The experimental group (23 of 45 invited schools) received support to interpret and use the feedback, including principal professional development on interpreting, understanding and explaining the feedback. The control group received no support. Using a follow-up questionnaire several months later, the authors assessed data literacy, the use of feedback, and the perceived effects of feedback and finding a “rather low level of perceived impact” of feedback (p.149). However, the authors confirmed that feedback and support for its use resulted in greater data literacy and self-efficacy among the experimental group principles. This suggests that feedback with support may open the door to greater improvement effects by creating potential for better use of feedback, as well as greater confidence in its use, to support learning.

Tymms (1995) investigated the effects of feedback on teacher behavior and attitudes in an analysis that reported on the results of experiments conducted using feedback from the A Level Information System (ALIS), an education monitoring system used in the United Kingdom beginning in the mid-1980s. (p.123) In the relevant experiment, Tymms provided both short and long versions of an ALIS report to determine whether the form of feedback made a difference in how the report was used. Tymms then surveyed 74 schools and colleges regarding teacher perceptions of ALIS and used the results to develop composite scores in three areas focused on the reading of ALIS reports, the use of data to change behaviors, and attitudes about the value and potential of the data in the reports. Tymms found that the short form of feedback resulted in more thorough review of the ALIS report across all curriculum areas. Regarding changes in teacher behavior, Tymms’ multi-level analysis found no difference in the effects of short and long reports in the mathematics and science curriculum areas, however, among humanities
disciplines, “long reports were associated with more changes, although differences did not reach statistical significance ($p > .1$).” (p.136) Finally, regarding attitudes about the ALIS program, the languages curriculum area respondents were significantly more positive towards ALIS when they received the shorter reports. Tymms’ analysis and results, while not directly related to student performance, serve the useful purpose of demonstrating the complexity of the issues around determining feedback effects on performance. For feedback to lead to changes that impact student performance, the feedback has to be understood and used effectively, a factor that can be influenced by the form and perceptions of the feedback itself.

Schildkamp, Visscher, and Luyten (2009), intent on determining feedback impacts on student achievement, analyzed the use of a school self-evaluation instrument designed to improve the quality of education in Dutch schools. The authors used a sample of 79 Dutch schools that provided a representative sample of students with respect to socioeconomic status of student families. Within this sample, two cohorts of students were followed for five years. Average student achievement was measured using results from a spelling and mathematics testing system that provided nine measurements per student during the five-year monitoring period. The authors’ also included student information thought to influence achievement such as gender, socioeconomic status, age, class size, and perceived intelligence. The authors conducted a multi-level analysis with repeated measures and estimated linear and non-linear growth models. Their Cohort 1 results indicated that students in schools with higher use of feedback did not make more growth over time than students in schools with lower use of feedback; these results held true for both spelling and mathematics. Cohort 2 results were similar to those of Cohort 1 and showed no differences in growth in either spelling or mathematics for students in high-feedback-use schools versus low-feedback-use schools.
Fuchs and Fuchs (1986) analyzed one form of feedback, formative evaluation, on student achievement. The authors conducted a meta-analysis of 21 studies, selecting studies based upon several criteria. First, studies had to use a control group to assess formative evaluation effects. Second, “formative evaluation” had to include at least two curriculum-based data collections each week. Third, program adequacy decisions based upon the formative evaluation had to occur at the individual as opposed to the group level. The authors excluded studies that assessed non-academic behaviors, provided feedback only to students, employed college-age subjects, or focused on behavior modification and employed time series to test experimental effects. Study results were transformed to a common measure of effect size, Cohen’s $d$ (difference between the treatment means divided by the control group standard deviation). Each effect size was converted to an unbiased effect size. The overall weighted average unbiased effect size was .7, which suggests “the integration of formative evaluation [feedback to the teacher, not the student] with instruction would raise the typical achievement outcome score from 100 to 110.5, or from the 50th to the 76th percentile.”\(^2\) These results suggest that feedback of formative evaluation information to teachers within the body of studies analyzed significantly increased student performance. Additionally, the authors noted that “neither quality of study nor year of publication appeared to mediate or moderate evaluation effects. Only publication type yielded a statistically significant difference,” with effect sizes of published studies being higher than those found in unpublished manuscripts. (p.205)

Coe (1998) conducted an experiment with UK teachers included in the A Level Information System (ALIS)\(^3\), wherein half of the participants received value-added performance and student perception information for multiple years. The control group received no feedback.

---

\(^2\) The authors specified this result in terms of the standard normal curve and assuming a test scale with a population mean of 100 and a standard deviation of 15.

\(^3\) The “A Level” is a standard examination for secondary students in the United Kingdom.
Coe’s analysis identified differences between the two groups’ students’ performance, with the students of teachers in the experimental group averaging about “one third of an A level grade” better than students of control group teachers. (p. 243)

Hammond and Yeshanew (2007) conducted a study in the United Kingdom that assessed the impact of feedback provided to schools on student performance and progress in English and mathematics during primary age schooling. Using descriptive statistics and multi-level modeling, the authors analyzed longitudinal pupil-level data on three measures, school-level contextual information, and school-level feedback provided as part of school participation in a variety of programs. The descriptive analysis investigated whether differences exist between schools that received feedback and those that did not, including whether differences exist between those who paid for feedback as part of a specific program and those who received feedback as part of a different, unpaid feedback mechanism. The authors’ multi-level analysis focused on Key Stage 2 English and mathematics test scores and the Key Stage 2 average score.\(^4\)

Hammond & Yeshanew’s (2007) analysis showed that schools receiving feedback produced a higher mean performance score than schools receiving no feedback. In fact, “all three outcome variables were positively associated with the schools who participated in PASS [a feedback program in which schools paid to participate], and those schools who received feedback as part of another project.” (p.103) Their analysis also showed differences between feedback programs with one program in particular, PASS, “showing a significant difference (\(p<.05\)) in attainment compared to those who received feedback as part of another project.” (p.103) The authors’ multi-level model serves to reinforce the findings cited above. For schools participating in the PASS (paid feedback) program, feedback had significant positive impacts on

---

\(^4\) Key Stage 1 (KS1) refers to the period of primary school when students are six and seven years of age. Key Stage 2 (KS2) refers to the time period of primary school when students are 8-11 years of age.
Key Stage 2 average performance. Positive impacts were demonstrated also for other schools receiving feedback (unpaid), but the relationship was not significant. The authors cite one other variable of particular interest – the interaction between student Key Stage 1 performance and unpaid feedback, which was significant but negative. They explain this by suggesting “these schools had a slightly lower relationship between KS1 and KS2 performance than others.” (p.105) Finally, their analysis of the other outcome variables, KS2 English and KS2 Math performance, showed similar results: significant positive relationships between paid feedback and performance in both English and math, a positive, non-significant relationship between unpaid feedback and performance in English, and a significant positive relationship between unpaid feedback and performance in math, albeit a smaller effect than that for paid feedback and math performance.

Hammond & Yeshanew’s (2007) study is important as it represents a close analog to the present study within the body of research presented with respect to the impact of feedback. The dependent variables include school performance measures based upon student performance on standardized testing. The feedback variables are not identical in nature, but the present study includes feedback-related variables chosen to represent the receipt and use of feedback, similar to what Hamond and Yeshanew accomplish with their choice of variables. This will add to the relatively small body of research that looks specifically at school performance and how performance management systems affect such performance.

Schildkamp and Teddlie (2008) included two different school performance feedback systems, one in the Netherlands named ZEBO (a Dutch acronym for primary school self-evaluation) and one in the United States, the School Analysis Model, or SAM. Their analysis focused on feedback effects on variables “which seem to be positively associated with pupil
To assess effects of the feedback systems, the authors developed a questionnaire with versions for both principals and teachers. The authors analyzed questionnaire data to determine perceived effects on school leadership and to what degree negative effects were perceived. Descriptive statistics revealed several perceived effects that were consistent across the two feedback systems. First, the results showed increased consultation on the quality of education, which Schildkamp and Teddlie cite as important for inducing change in an organization. Second, the results showed moderate improvements in professional development for staff due to identification of individual needs for additional professional development. Teachers and principals also indicated that feedback resulted in changes to classroom instruction, “another requirement for improved school performance.” These changes also included increased collaboration and sharing among teachers as well as increases in “hands-on” activities, technology use, and a focus on critical thinking. With respect to negative effects, the analysis indicated slight differences between the Dutch and American perceptions but no major perceived effects, with less than ten percent each of American teachers and principals indicating negative effects of the feedback system, primarily related to time required and to stress induced.

**Performance Management Systems and Organizational Performance**

Frazier (2008) studied whether nine different management practices, including strategic planning, performance measurement, and process improvement, produce effects on a government agency’s performance. While Frazier (2008) points out “there is little, if any, empirical evidence regarding the effect of strategic planning…on organizational performance in public sector agencies,” he notes that numerous private sector analyses indicate positive relationships between strategic planning and organizational performance (pp.79-80) as well as goal setting and performance in both public and private sector organizations. His hypothesis
suggests that planning practices deployed in public organizations should lead to higher performance as well, and he sets out to test this assertion.

Frazier’s (2008) analysis design included a 44-item survey of eight state revenue agencies. Thirty-nine items measured employee perceptions of management practices in numerous areas, including planning and performance measurement, both important facets of performance management systems. The five additional items measured perceptions of organizational results. (Frazier notes on page 82 that, while not as good as objective data, evidence suggests a high correlation between perceptual and objective data.) Frazier’s planning measurement used six of the 39 management practice items while he measured performance measurement practices using five others. The items included in the strategic planning and performance measurement constructs yielded Cronbach’s alpha coefficients of .8633 and .8625 respectively, suggesting that the items reliably measure what they purport to measure. The author’s exploratory procedures included calculating Kendall’s tau and generating multiple regression equations.

The initial multiple regression analysis indicated results were most strongly related to process improvements, strategic planning, and performance measurement. Frazier (2008) suggested that these three systemic and analytic factors serve to mediate the additional variables. The author’s second SEM model confirmed eight of the nine tested management practices positively affect employee-perceived results and that, in fact, planning, performance measurement, and process improvements have greater effects than the other six management practices. Frazier conducted follow-up interviews with employees, who attributed to these practices tangibility, demonstration of progress and demonstration of clear causes and effects. Specifically, with respect to strategic planning, employees pointed to the importance of aligning
work plans to agency goals. Performance measurement, according to employees, helped to
“improve the streamlining and fine-tuning of processes” and in demonstrating progress. (p.94)

Walker and Boyne (2006) provided “the first empirical test of a comprehensive program
of management reforms that is intended to improve organizational performance” (p.371) in the
UK public sector. The authors intended to build upon previous analyses showing that managerial
actions and strategies can affect organizational results. Their analysis included a survey of over
4000 managers and staff at 386 English local governments responsible for delivering education,
land planning, housing and other services. Organizational performance, the authors’ dependent
variable, was measured using a “core service performance” score, part of a measure constructed
by an external commission, and by internal measures obtained using eight items included in the
survey. Their analysis focused on several key factors of the government’s reform program: use
of minimum standards, devolution of authority, flexibility, and choice. Each factor included
several independent variables. Of particular relevance to the present analysis were variables for
target setting as well as for target ownership. The final analysis showed target setting was not
significantly related to performance as measured by external core service performance or internal
outcome perceptions, while target ownership was significantly and positively related to both
measures, suggesting that target setting is a necessary but not sufficient condition and that targets
“are effective only if owned by those directly responsible for providing services.” (p.386)

Andersen (2008) analyzed the effect of performance management reforms in Danish
public schools. Using a mailed questionnaire, he surveyed all public and private schools in
Denmark to ask about a variety of performance management procedures, including development
of “steering documents,” written goals, and evaluations (which imply measurement).
Respondents were asked to indicate whether these performance management techniques were in
use and, if so, to specify how long they had been in place. As part of assessing effects of these practices on performance, Andersen used these data to create indexes for scoring the schools on the basis of the performance management tools and the duration of use.

Results of Andersen’s (2008) multi-level analysis indicate that with the exception of one of the years analyzed, the reforms do not correlate significantly with student performance. The single year with a significant finding of effect should be considered to be a small effect given the size of the dataset (performance data for over 80,000 students), according to Andersen. Evaluating the reform indexes, Andersen noted that they demonstrate “a clear tendency toward improvements over time,” (p. 549) but summarized the overall results by stating that they provide “minimal support to the expectation that performance management systems improve performance,” (p. 552) though he offers some thoughts about why this may particular to this study. Of concern to Andersen, however, were findings regarding the interaction effects between performance management reforms and student socioeconomic status. These results indicated “students with low socioeconomic status perform worse at reforming schools than at similar non-reforming schools.” (p.541) On the surface at least, this may suggest that reform implementation in these schools might have added additional stress to an already difficult and under-performing environment. Andersen’s overall results confirm the need for additional research to better evaluate the value of performance management in schools.

Walker, Damanpour, and Devece (2010) considered innovation and performance management effects on organizational performance in another analysis of English local governments. The authors assert a number of hypotheses, among them that performance management is positively associated with organizational performance, their dependent variable in this study. Organizational performance data were provided by an externally-developed dataset
used for audit inspections and investigations. Performance management measures came from a survey of local government units and was designed to capture perspectives from multiple levels within the target organizations, including both corporate-level and service area-level officers. Performance measures consisted of several components related to mission and goals, targets and measures, objective clarity, and communications designed to allow for corrective action.

Walker, Damanpour, and Devece (2010) employed hierarchical regression to assess effects of innovation and performance management on performance and used structural equation modeling to assess mediation of effects of performance management between innovation and organization performance. Their analysis found direct effects of performance management practice on organizational performance, and their structural equation models also determined that innovation, while not directly affecting performance was, in fact, mediated through performance management. They summarized their results by pointing out that they demonstrate “that performance management mediates the effect of innovation on performance and have [sic] a positive influence on its own,” and that performance management systems do influence “the likelihood that employees will behave in ways that lead to the attainment of organizational objectives.” (p.381)

In another analysis directly relevant to the present one, Sun and Van Ryzin (2012) test the idea that performance management improves public sector outcomes in their analysis. Using testing data from New York public schools, the authors analyze the relationship between performance management practices and educational outcomes.

Sun & Van Ryzin (2012) gathered data from a New York Department of Education quality review of about 1000 schools. They developed four independent variables from the testing data. Two of the variables focus on the proportion of students proficient in math and
English-Language Arts (ELA) while the other two focus on math and ELA growth of the lowest performing third of students. Independent variables were developed from an explanatory factor analysis that loaded 35 quality review criteria onto seven factors. The authors also controlled for a variety of common school, student and teacher characteristics including student-teacher ratio, student socio-economic status and teacher experience and education.

Sun and Van Ryzin’s (2012) OLS regression analysis found with respect to student proficiency, higher performance management index scores related significantly to better student achievement, holding constant the 12 control variables. In fact, the authors estimated a two- to three-percent gain in student achievement for each standard deviation improvement in performance management. Their models fit the data very well, with the ELA model explaining 72% of the variation in student proficiency and the math model explaining 53% of the variation in student proficiency.

Despite an overall model fit that was not as good ($R^2_{ELA} = .17$, $R^2_{Math} = .06$), the student progress models also showed a significant association between the performance management index and the ELA and math performance of the lower third of students. The authors’ calculated that this finding suggests about “a 1 percentage point increase in the proportion of students demonstrating progress in their ELA and Math scores” for a one standard deviation increase in the performance management index. (p.9) In combination with the proficiency finding, this leads the authors to conclude that “consistent with the theory and aims of the performance movement in public management, those schools that engage more fully and effectively in performance management practices do indeed perform better as measured by both the level and gain in standardized reading and math test scores.” (p.10)
It is a finding such as this that the present study aims either to reinforce or to refute, for as Sun and Van Ryzin (2012) also point out in this recent analysis, while there is “widespread acceptance of the performance management approach” in public organizations, “whether it is actually associated with better organizational outcomes has rarely been tested empirically.” (p.2) This is especially true in educational environments and the present analysis will add to the small body of academic performance management literature specific to public education by providing an additional source of empirical results focused on performance management effects on school-level outcomes.

Fernandez (2009) conducted an exploratory analysis that examined the effectiveness of school improvement planning in 252 Clark County, Nevada schools. The dependent variable consisted of a school-level measure of student growth on the Iowa Test of Basic Skills (ITBS) from one school year to the next. School improvement plan quality measures for each school provided the independent variable. The quality measure for each plan was based upon a scoring rubric used by the Leadership and Learning Center to evaluate the plans (The author was not a plan evaluator; he obtained the scores form the Leadership and Learning Center). The author also included control variables for several school demographic and resource factors for each school, including measures of minority representation, limited English proficiency, and school resources (per pupil spending). The analysis produced two models – one for mathematics scores and another for reading scores.

As hypothesized, Fernandez (2009) found plan quality positively related to school improvement as measured by growth in student scores on the ITBS test. In both reading and mathematics, regression coefficients were significant (math at the $p < .01$ level and reading at the $p < .05$ level), even when controlling for school demographics and resources. Fernandez (2009)
concluded that these results show “that the relationship between SIP quality and school performance is not just a product of wealthy schools” (p.354) producing higher quality improvement plans that lead to better results.

Huber and Conway (2015) set out to assess whether the quality of a school’s improvement plan relates to the level of student achievement in a school. Their analysis focused on low-performing school districts in Connecticut, and began with the hypothesis that they would find a positive relationship between plan quality and student achievement. Like the present analysis, the authors included schools serving grades K-8.

Huber and Conway (2015) applied a method that scored school-level plans using a rubric called the PIM, or Planning, Implementation, and Monitoring School Improvement Audit. The PIM rubric includes assessment of planning processes and use of SMART goals (specific, measurable, attainable, realistic, and timely). PIM scores provided the independent variable, while School Performance Index scores based upon state assessments including reading, writing, mathematics, and science supplied the student achievement data (the dependent variable). Additionally, the authors included variables to control for prior year achievement, per pupil spending, minority representation, socio-economic status, English language learners, special education students, and school size.

Huber and Conway (2015) found a “small but statistically significant positive relationship between SIP score and the outcome variable,” student achievement. (p.61). The regression analysis demonstrated “a high proportion of shared variance between the predictor and outcome variables, with $R^2 = .861$ ($p < .01)$” (p.63). However, the relationship between student achievement and plan scores was not significant ($p = .052$), though it was extremely close. The analysis indicated that multicollinearity between variables was not a factor. According
to Huber and Conway (2015), the results were mixed, as “the correlational analysis and descriptive analysis suggested support, but the multiple regression analysis did not reach statistical significance” (p.63). Given these results – the nearly-significant relationship between the overarching plan scoring index and student achievement – the present analysis attempts to take a more fine-grained approach by looking at which specific aspects of plans seem to have an effect. By using specific components of individual plans, rather than an overall score for the entire plan, this more detailed approach should offer additional insight into the portions of the planning and goal-setting process that offer the most opportunity to influence student achievement and school performance.

**Summary of the literature**

The academic literature around key aspects of performance management systems offers a few key suggestions. First, it suggests that goal setting, measurement, and feedback have an important role to play in producing desired outcomes in the public sector. Second, it suggests that that role is not always as clear-cut as it may seem. While the primary features of performance management systems seem to affect performance, the mechanisms are complex and require further study to gain a more complete understanding. Finally, the literature suggests that certain public sector environments with especially complicated relationships and missions, including educational environments, need additional research to determine whether the mechanisms by which performance may be improved are more or less similar than those mechanisms that can be used to produce results in other environments. The present study offers an opportunity to increase the research base and our understanding in the fields of public administration and education.
Chapter 3 - Research Design and Statistical Methods

Chapter 3 organizes the research design and methodology into three sections. The first provides the overall research design and the research model, including sample selection. The second portion discusses data sources and the variables drawn from these sources. The final section reiterates the main hypotheses of the analysis, discusses the statistical procedures to be used, and considers potential design limitations.

Research Design

This analysis uses a quasi-experimental sampling design built on data from multiple sources that include a valid, reliable biennial survey of North Carolina teachers regarding school-level working conditions, a statistically-based school performance assessment system used by NC schools, and an analysis of school improvement plans. Survey and improvement plan data provide independent variables while the dependent variables are supplied by the school assessment system. Both dependent variables reflect a school-level score (based upon the aggregated results of the students in each school), and all other predictor variable and control variable data reflect this same unit of analysis – the school. These variables, together as part of the hierarchical regression analysis presented here, offer an opportunity to gain insight into the research questions discussed in Chapter 1 and briefly summarized here. Two primary components of the process are assessed – initial establishment of the school improvement plan and execution of the plan – and whether a relationship exists between what schools do as part of planning and execution and how their students perform. Conducting a hierarchical analysis, in which variables are entered in blocks, allows assessment of the impact of the various components of the model on school performance.
First, with respect to initial establishment of the school improvement plan, the present analysis assesses whether a significant relationship exists between school performance, as indicated by a measure of school value-added\(^5\), and key plan factors such as the number of goals and strategies, the clarity of goals and strategies, the use of stretch goals (those which are difficult-but-not-impossible) and the inclusion of planned feedback loops.

Second, regarding school improvement plan execution, the study seeks to determine the impact of implementation approaches, especially feedback to teachers, on overall school performance. The present study assesses these links between planning and implementation for a subset of North Carolina schools to determine whether specific dimensions of planning practice relate to overall school performance. The specific hypotheses were first stated in Chapter 1, and they are listed again below.

**Research question 1:** Do schools with a more focused school improvement plan, as evidenced by smaller numbers of plan goals and strategies, perform better?

**H1:** The number of goals and strategies in an improvement plan is inversely related to overall school performance.

**Research question 2:** Do schools with plans that offer greater goal clarity, shown by greater proportions of goals having clear targets and completion dates, perform better?

**H2:** A positive relationship exists between clear goals and school performance.

**Research question 3:** Do stretch goals result in higher levels of school performance than either very easy goals or extremely difficult goals?

---

\(^5\) Value-added is a statistical analysis used to measure the impact of districts, schools and teachers on the academic progress rates of students from year-to-year. A value-added score is calculated by comparing current achievement to all prior achievement as measured by an assessment. In North Carolina, these are end-of-grade (EOG) and end-of-course (EOC) assessments. Additional detail can be found at [http://ncdpi.sas.com](http://ncdpi.sas.com).
**H3:** Stretch goals support higher performance. Relative to stretch goals, very easy goals and very difficult goals are associated with lower performance.

**Research question 4:** Do the presence of feedback mechanisms and perceptions of feedback effectiveness affect school performance?

**H4:** A positive relationship exists between both the presence and perceptions of feedback mechanisms and school value-add.

**The Model**

These questions and the literature imply a model where overall school performance, as measured by the school value-added assessment, varies based upon the practices in place for developing and executing an improvement plan within an individual school building. This model is presented again in Figure 3.1.

![Figure 3-1. Initial Overall Research Model](image-url)
Sample Selection

As mentioned previously, all North Carolina public schools, regardless of grade levels served, are required to develop a school improvement plan. The present study is based upon an analysis only of elementary and middle schools that were open during the 2013-14 and 2014-15 school years. This analysis draws its sample from elementary and middle schools because of differences in high school testing programs that result in different inputs to the school value-add measure. The elementary and middle school testing regimes are explained in more detail below.

Explaining the Focus on Elementary and Middle Schools

Elementary and middle school students in North Carolina public schools take annual End-of-Grade (EOG) tests developed by the NC Department of Public Instruction. These exams are given for mathematics, reading, social studies and, for fifth and eighth grade students only, for science. These tests are grade-specific, may not be given to students “off-grade-level” (a seventh grader cannot take the eighth-grade test) and serve as the measurement instruments for the North Carolina school accountability model. The annual results are included in the EVAAS system for each student who takes the test, providing repeated measures over time for all students who attend a NC elementary or middle school for more than one school year. (See the Data Sources section for additional details.) Students who are accelerated in a particular subject such as math still take the grade-specific EOG for their grade level, even though they may also take a subject-specific examination as well (an End-of Course test, or EOC). In elementary schools in particular, students routinely receive instruction in self-contained classrooms, that is, they do not see different teachers for different subjects. (Magnet schools, which typically offer an expanded library of courses designed to draw students from across the district, are often an
exception to this.) Generally, in a self-contained class students receive instruction in multiple subjects from the same teacher, and nearly every teacher in the school provides the instruction in those subjects (excluding Special Needs teachers, counselors, and other specialists in the school).

**Total Population and Sample Size**

For the 2013-14 school year, North Carolina had a total population of 1,836 schools serving students in pre-kindergarten through eighth grades. Because the present analysis includes a detailed review of individual school improvement plans, a sample of this population is used to facilitate analysis within a reasonable period of time. Sample size for this population was determined using the standard formula for sample size determination. Based upon this calculation, the sample requires a minimum of 318 schools to achieve a 95% confidence level at a 5% confidence interval. A sample of 318 schools was randomly selected from the previously mentioned population of 1,836 elementary and middle schools. To ensure a random selection, school information was placed in an Excel spreadsheet and sequentially numbered. Schools were randomly drawn from the numbered list using the built-in RANDBETWEEN function, then school names were matched to their number using the VLOOKUP function.

**Data Sources**

Multiple sources provide data for the present study. Dependent variable values are drawn from the Education Value-Added Assessment System, or EVAAS. Independent variable data about planning practices are drawn from an analysis of school improvement plans from North Carolina elementary and middle schools. Additional independent variable data regarding personnel perceptions of these practices are drawn from the North Carolina Teacher Working

---

6 Source: Education Directory and Demographical Information Exchange (EDDIE), a resource provided by the North Carolina Department of Public Instruction at http://apps.schools.nc.gov/pls/apex/f?p=125:1:
Conditions Survey (TWC). Control variable data are drawn from a variety of sources including the NC School Report Card system and other data collection and reporting systems managed at the NC Department of Public Instruction. Each of these sources is described in greater detail below.

**Education Value-Added Assessment System (EVAAS)**

Dependent variable values for each school in the sample are drawn from the Education Value-Added Assessment System, or EVAAS. This information system uses student-level assessment data to build a longitudinal model of student, school, and district performance using two different statistical models. Initially, EVAAS provided a predictive “univariate response model” (URM) for all students. This model was designed to estimate individual student performance on standardized tests based upon each student’s past performance on such tests. Predictions required a test history of at least three tests. These predicted scores are compared to students’ actual scores to determine how well the students performed against expectations. These performance data were used to develop value-add measures for schools. Recently, the EVAAS developer, SAS Institute, altered the underlying statistical model for elementary and middle schools - the unit of analysis for this paper. This “multivariate response model” (MRM) offers a growth-based analysis of student performance. The MRM is a longitudinal, multivariate, linear mixed model that SAS describes in their documentation as essentially a multivariate, repeated-measures model.”\(^7\) The model simultaneously fits observed test scores for each student. Students with as few as two assessment results can be included in the analysis. Over time, students serve as their own controls as additional assessment results are added into the system each year. Irrespective of the model employed, all EVAAS analyses use scores from standardized tests

---

\(^7\) See SAS White Paper “SAS EVAAS for K12 Statistical Models,” p.3
using reliable measures that highly correlate with state standards and that measure both high-achieving and low-achieving students, and these student performance data are used to develop value-add measures for schools.

Each public school in the state receives school value-added reports based upon the test results of its students, as described above. The results permit EVAAS users to observe the average progress of students within a school and to compare a school's achievement level to the state's average achievement. For elementary and middle schools, end-of-grade (EOG) reading, math, and science test scores for students in grades four through eight provide the basis for estimating a school’s effect on student progress for each student in a school.

A school’s Accountability Growth Estimate is based upon the aggregation of its students’ scores. Therefore, when extrapolated to the school level, a positive value indicates that students in that school, on average, made more academic progress than their peers, on average, across the state. A negative value means they did not make as much progress as their peers; any score within two standard errors of average is deemed maintaining the same level of achievement. Scores will rise in a school where students, on average, exhibit academic growth during the school year.

School Improvement Plan Analysis

North Carolina General Statute §115C-105.27 requires every North Carolina public school to develop a school improvement plan. Improvement plans must consider both State Board of Education goals as established in the mission statement and the school-specific annual performance goals established by the Board under GS §115C-105.35. To support schools in the planning process, the North Carolina Department of Public Instruction (DPI) provides a School Improvement Planning Implementation Guide as well as a template aligned to the best practices
found in the Guide. School districts and schools may use these tools to the extent they wish; while every school must have a school improvement plan, it does not have to follow the concepts or format recommended by the DPI. As a result, there is significant variation in the form and quality of plans around the state.

Analysis of the school improvement plans from each school in the sample provides a snapshot of the output produced by the planning committees at each school. These snapshots offer some insight into the degree of sophistication of the planning process at an individual school, and they provide specific information relevant to the research questions.

**North Carolina Teacher Working Conditions Survey (TWC)**

The North Carolina Teacher Working Conditions Survey (TWC) provides school-level data regarding some of the planning practices and personnel perceptions of practices in each school. This survey, conducted every two years since 2002 by the New Teacher Center (NTC),\(^8\) is specifically designed to obtain input from licensed teachers about school and teacher leadership as well as other supports for educators such as facilities and resources, community support, and professional development. Survey results are intended to inform school and district improvement efforts. The statewide survey, conducted anonymously via the Internet during a pre-defined time-period, consistently yields impressive participation levels. The 2014 version, conducted online between March 10 and April 4, 2014, received over 93,000 responses – about 90 percent of the licensed teachers and administrators in the state. Several questions within the survey provide response data that can offer insight into some of the research questions in the present study.

---

\(^8\) The New Teacher Center, originally part of the University of California at Santa Cruz, is an independent, non-profit, non-partisan group with a mission to support the development of an effective, dedicated and inspired teaching force. NTC has broad experience conducting similar surveys around the US. Details on NTC may be found at www.newteachercenter.org.
The survey design focuses on several key areas of interest to schools and school districts, which can use the results as part of the on-going improvement process. Topic areas include:

- Community Engagement and Support
- Teacher Leadership
- School Leadership
- Managing Student Conduct
- Use of Time
- Professional Development
- Facilities and Resources
- Instructional Practices and Support
- New Teacher Support

The data drawn from the TWC survey for this analysis come from the question area that focuses on teacher and school leadership and relates to Research Question 4 and how feedback mechanisms operate in the schools. Details of the survey data and associated research questions are summarized below. Most, but not all, of the survey questions ask teachers to respond by selecting a level of agreement ranging from strong disagreement (“1”) to strong agreement (“4”). “Don’t know” responses are assigned a value of “5.” The present analysis focuses on teachers who indicate agreement (3) or strong agreement (4) with the questions that are used to define the variables, addressed in greater detail later.
Control Data Sources

The present study uses several data points to control for various aspects of the schools included in the analysis. These data are focused on variables commonly cited as important controls as part of educational production functions, and detailed discussion is provided in the section related to variables below. Control variable data are drawn from information systems and reporting mechanisms developed and used by the North Carolina Department of Public Instruction to execute the management of the state education system and to ensure that state and federal reporting requirements are met. All data are considered public information and are available on the DPI website. Some data may only be reported at the school or other summary level to ensure compliance with privacy laws.

School Report Cards

Much of the control data were drawn from NC School Report Cards. The report cards, initially developed during the first administration of Governor Mike Easley (2001-2005) in response to federal legislative requirements, provide basic school demographic information such as school size, as well as data regarding student performance, school safety, and teacher quality. The purpose of school report cards is to provide an easily available and public source of information to help parents gain a better understanding of their school(s) and to make informed decisions about what schools they wish their child(ren) to attend. Under the Elementary and Secondary Education Act legislation (formerly known as No Child Left Behind), parents must be offered school choice when certain school performance standards are not met for three consecutive years.

Every North Carolina public school - traditional, charter, and alternative - receives an annual report card. There are some minor differences between school types (schools are assigned
to a category - elementary; middle; high; combined elementary, middle and high; combined elementary and middle; or combined middle and high - based upon grade range), but generally all schools receive the same basic comparisons. Because charter schools have greater operational flexibility, their teacher quality data are slightly different than other schools.

School data are based on information from all grades within the school. For some data, including School Size, School Performance, School Safety, Attendance, and all information in the Quality Teachers section, all schools in the same grade range category are averaged to produce district and state comparison results. Where data are not available to be reported, or where student numbers do not allow reporting based on federal or state privacy requirements, notations of “N/A,” “>95%,” or “<5%” are provided.

School report card data are presented using a variety of “tabs” accessible from an individual school’s School Report Card page. The School Profile tab provides basic school information including average class size, a number that is provided for each grade level within the school. For this analysis, grade-level averages are themselves averaged to provide a school-level value. Individual classroom sizes are not available. The School Profile also offers school size data, based upon the final Average Daily Membership (ADM) for the school. ADM for each school month is based upon the sum of each student’s number of days in membership, divided by the number of days in the school month. (Student membership begins upon enrollment and continues until the student is withdrawn for any reason.) ADM is an important part of determining school funding allotments. School size is a calculated number based upon the monthly membership day figures and the number of days in the school year.

The Quality Teachers tab offers school-level teacher data that includes teacher experience data. These data are reported as percentages of teachers within an experience range (0-3 years, 4-
10 years, and 10 or more years). The percentages for the latter two categories are added together to provide each school’s percentage of teachers with more than 3 years of experience. The same portion of the School Report Card also provides school-level data for the percentage of teachers with an advanced degree and the number of teachers who have obtained certification from the National Board of Professional Teaching Standards, also reported at the school-level.

**Grade, Race, and Sex Data**

Student grade, race, and sex data are collected from each school at the end of the first school month. These data are reported annually in February for the current school year. The report is provided as a spreadsheet workbook and organizes the data in four ways: by local education agency (LEA, or school district), by LEA and grade level (first grade, second grade, and so on), by school, and by school and grade level. The data are organized into columns that include numbers of males and numbers of females for each of six racial categories: American Indian, Asian, Black, Hispanic, Pacific Islander, and White. A seventh column is included for counts of those who specify two or more categories.

For the present analysis, numbers for six of the seven categories, by school, are used in combination with school size data to calculate the proportional representation of American Indian, Asian, Black, Hispanic, Pacific Islander, and multi-race/ethnicity students within the overall student population.

**Free and Reduced Meals (FARMS) Data**

The Child Nutrition Services section at the Department of Public Instruction operates state and federal child nutrition programs designed to support economically disadvantaged students and their families. Staff process applications for free and reduced cost meals and collect data on applications received. A school-level report of the percentage of economically
disadvantaged families provides an annual snapshot of economically disadvantaged students at every school. Data for elementary and middle schools are generally reliable. High school data are often under-reported as many teenage students are reluctant to apply for free or reduced meals.

FARMS summary-level data are commonly used as a proxy for economically disadvantaged students and families. For the present study, school-level data drawn from the Child Nutrition Services database are used for this purpose.

**Variables**

This analysis draws upon school-level data from multiple sources as described briefly above and detailed in the following section. This section describes each variable in greater detail and includes an assessment of each variable against Gauss-Markov assumptions. For a multivariate ordinary least squares (OLS) regression analysis to offer the best linear unbiased estimation and, in turn, be suitable for statistical inference, dependent and predictor variables must conform to these assumptions.

**Dependent Variables**

The current analysis models two different dependent variables: the School-wide Accountability Growth Estimate for 2014-15, and the three-year trend (2013-14 through 2015-16) of this same indicator for each school in the sample.

**School-wide Accountability Growth Estimate**

The School-wide Accountability Growth Estimate (SAGE) is a composite index calculated by the EVAAS system for every public school in the state. The SAGE is a quantitative
value measured on a continuous, unbounded scale, meeting key Gauss-Markov assumptions for the dependent variable in an OLS regression.\(^9\)

For elementary and middle schools, this index score includes all end-of grade (EOG) test scores (reading, math, and science) for each student in the school and offers a school-level “value-added” assessment to add to traditional student achievement indicators. This is useful because whereas traditional achievement indicators inherently capture effects of students’ backgrounds, value-added measures capture effects of the school on student achievement (Bell & White, 2016).

Value-added indicators accomplish this by using all available student scores to follow progress of a student over time and provide a more reliable assessment of students’ achievement upon entry into a given grade level. The school-level indicators then compare students’ change in achievement over a school year to an expected amount of change. This change in achievement reflects a school’s value-added, which may be positive, negative, or neutral.

It is important to note these value-added measures are not connected to proficiency. A student or group of students may achieve a high level of growth within a given school year and yet still not reach a level considered proficient simply because they began the year far below the proficiency level. The opposite may also occur. Students may begin the year proficient, achieve effectively no growth, and still be proficient. The growth component is important because it provides a basis for understanding how well a school has performed at helping students’ progress academically (SAS EVAAS Methodology).

---

\(^{9}\) Within the sample used for this analysis, the SAGE ranges from about -11 to about 14.
**SAGE 3-year Trend**

The second dependent variable assessed is the SAGE 3-year trend for each school in the sample and is used to assess change in performance over a three-year period. This measure uses the same SAGE variables described above and is calculated by subtracting a school’s 2013-14 SAGE value from the 2015-16 SAGE value. For example, one school in the sample has a 2013-14 SAGE value of 2.01 (indicating positive school impact on student growth for that school year), a 2014-15 value of 2.55, and a 2015-16 SAGE value of -.21 (reflecting negative school impact on student growth). The overall 3 year trend for that school is -2.22 (Ending value of -.21 minus the starting value of 2.01), meaning that over the three year time period when the school’s improvement plan(s) were expected to improve student and school performance, the school’s performance (impact on student growth) actually decreased.10

**Independent Variables**

Poister and Streib (2005) assert that strategic planning “is an action-oriented type of planning that is useful only if it is carefully linked to implementation – and this is often where the process breaks down” (p.46). The present analysis assesses the link between planning and implementation for a subset of North Carolina schools to determine whether specific dimensions of planning practice relate to overall school performance. This may offer insight into the issues noted earlier and to demonstrate the applicability of generally accepted best practices for planning in a school environment. The study considers two primary components of the process – initial establishment of the school improvement plan and execution of the plan – and whether a relationship exists between what schools do and how their students perform. A detailed

---

10 Within the sample, the SAGE 3-year trend ranges from about -14 to 13.
description of the independent variables used to operationalize the analysis is provided for each research question.

**Independent Variables Measuring Focus**

The level of focus generated by a school improvement plan is measured using a variable that aligns to research question one. This variable reflects the total number of goals plus the total number of strategies contained within a school’s school improvement plan.

The present analysis includes both goals and strategies to ensure a comprehensive picture of plan focus. While working closely with school improvement teams across North Carolina over several years (2008-2014), the author found often that shifts between goals and strategies can occur that do not reduce the breadth of initiatives within a plan and therefore do not increase plan focus. These shifts include changing “goals” into “strategies” so as to meet a pre-determined maximum number of goals, an approach that does not result in greater focus because the overall number of items to be attended to remains the same.

A few words should be said about the semantics of improvement planning. The use of the words “goal” and “strategy” comply with a hierarchy defined in the North Carolina School Improvement Planning Implementation Guide. Goals reflect the highest-level intentions or outcomes at any level (state, district, or school). Strategies reflect one level deeper, that is, they are the means by which the goals are to be achieved. The experiences of team members as well as local protocols or other factors may influence the language used within an individual plan (for instance, the word “objectives” or perhaps “outcomes” may be used in place of “goals”), but it is this hierarchical structure within a plan that is used to generate the counts that provide evidence of the plan’s degree of focus. Additional details of how goals and strategies were counted for the present analysis are included with the procedures in Chapter 4.
Like the individual counts of goals and strategies used to construct it, this variable is quantitative and interval. While the state model for school improvement planning recommends that plans limit the number of goals to three-to-five and the number of strategies to no more than three, in practice not all school districts use the state planning model, and even those that do may deviate from the recommendations based upon local preferences or needs. Consequently, both variables are anticipated to meet the Gauss-Markov assumption of nonzero variance. These data also will be screened for multicollinearity, but no reason exists to expect perfect, or even high, multicollinearity between this variable and the others contained in the analysis, or with the error term. (Data screening details are included in Chapter 4.)

**Independent Variables Measuring Clarity**

Plans containing clear goals can be expected to result in better performance because the goals provide precise information about what needs to be achieved and how and when to measure success. (Moynihan et al, 2016; Hu et al, 2011) Specifically, clear goals offer unambiguous targets (typically quantitative), measurable indicators, and precise dates for reaching the target. Two variables associated with research question two provide measures of goal clarity for the improvement plans included in the sample.

“Target” provides a continuous quantitative measure of the percentage of plan goals that include an unambiguous quantitative target or measurable status or state of being, i.e, “a 10 percent increase in…,” or “a 75 percent proficiency level…” Terms such as “all” or “every”, although not numerical, nonetheless were treated as clear targets.

“Date” is another quantitative, continuous measure and indicates the percentage of goals within a plan containing an unambiguous date by which a target is to be achieved. To receive credit for a goal’s end date, the plan must clearly define the date (“December 15, 2014”) or an
established and clearly identifiable end date ("end of first semester," “end of third quarter”). Unacceptable end dates include anything identified as “on-going,” or any other ambiguous date because the school can never be judged to have missed the goal. There will be close calls for a few goals. The determining factor in every case will be whether the date is sufficiently clear to provide accountability – a point at which the school can clearly be said to have missed the goal. A date that provides accountability also provides an incentive for teachers to meet the goal.

A goal which listed just a month and year is an example of a close call. On one hand, “November 2014” is ambiguous because it could mean the first day of the month or the last day of the month, or any other date between the two. At the same time, though, it does allow an outsider observer to say a goal has been missed if it has not been reached by December 1 and it would be counted (with some misgivings) as having a date.

Each of these variables is anticipated to meet the OLS regression assumptions of nonzero variance, a lack of multicollinearity, and a lack of correlation with the error term.

**Independent Variables Measuring Goal Difficulty**

In a setting where many people are involved in influencing results, goal difficulty (degree of “stretch”) can be hard to assess. Many different factors must be considered including the starting point, the amount of growth or change expected, and the ability of individual actors to influence outcomes. The simplest measure of goal difficulty is distance from a given target.

During the time period included in this analysis, an Elementary and Secondary Education Act (ESEA) waiver granted by the US Department of Education allowed North Carolina to replace the traditional Adequate Yearly Progress (AYP) measure with “more ambitious performance targets, known as Annual Measurable Objectives (AMOs).”

---

11 See “Understanding North Carolina’s Annual Measurable Objectives,” p.1
academic performance as well as participation targets. For reading and math, included in the present analysis, AMOs are academic proficiency level targets. Schools must meet these targets on designated assessments both overall (whole school) and for specific subgroups. AMOs levels are set at the state level, that is, all schools have the same whole school target, irrespective of their starting point.

The use of AMOs in the present study solves a problem particular to school improvement plans, namely that some schools do not identify a specific goal and/or target for reading and math proficiency in their plan, introducing the possibility of missing data for some schools. Because findings of impact for goal difficulty may suggest important implications for current practice, operationalizing goal stretch provides the potential for adding practical value to school planning teams.

Annual Measurable Objectives are set at the state level, and all schools have the same proficiency target. By definition, then, some schools have easier-to-achieve goals than other schools. Schools that are currently performing well as measured by proficiency levels may be very close, or even already above, the AMO target for the next year. Schools that are performing less well are further away from that target, perhaps even so far away that reaching it in the designated school year is nearly impossible. Understanding these relationships has the potential to be important because of its relevance to practice. If a weak relationship exists between easy goals and school performance (the presence of easy goals is not impacting performance) and a strong relationship exists between stretch goals (10-20 point increases) and performance, it suggests that AMO targets should not be established across the board, but instead should be customized for each school based upon their current proficiency levels so as to impact
performance at the school level – a change to current practice. Minimally, such a finding would indicate a need for further study.

**Operationalizing Goal Difficulty**

Rather than attempt a subjective evaluation of goal difficulty, whole school AMO targets for 2014-15 serve as the foundation for six variables associated with reading and math proficiency. The simplest measure of goal difficulty is reflected in each school’s distance from the 2014-15 state AMO target specified for all schools, based upon actual proficiency levels for the prior school year (2013-14). According to theory and prior research on goal difficulty, one would anticipate a non-linear relationship. In cases where achieving the target is very easy or very hard, one might expect the relationship to be negative. In cases where the target is difficult, but not impossible to achieve, one would expect a positive relationship. To operationalize this for purposes of the present analysis, the difference between the state AMO target and prior year performance is used to create 6 dummy variables for each school, three each for reading and math. These variables are devised to assess the relationship between performance and goal difficulty, as described below.

Variables *easyread* and *easymath* are designed to assess the relationship between school performance and easy goals. These two variables were coded such that a school required to raise reading and/or math proficiency less than 10 points (an “easy” goal) received a 1 and all others received a 0. A negative relationship is anticipated such that school performance is lower for schools with easy goals.

Variables *stretchread* and *stretchmath* are designed to assess the relationship between school performance and more difficult, but still not impossible, goals. These two variables were coded such that any school required to raise proficiency between 10 and 20 points (inclusively)
received a 1 and all others received a 0. The current analysis anticipates a positive relationship between school performance and these goal stretch variables, indicating that schools required to achieve difficult but not impossible gains perform better. The stretch goals serve as the reference categories.

Variables \textit{vdifread} and \textit{vdifmath} are designed to assess the relationship between school performance and goals that are unlikely to be achieved within the time span of a school improvement plan. These two variables were coded such that any school required to raise proficiency more than 20 points received a 1 and all others received a 0. The current analysis anticipates a negative relationship between school performance and these “impossible goal” variables, indicating that schools required to achieve extremely high gains perform less well.

Each of these variables are expected to meet necessary assumptions with respect to non-zero variance and multicollinearity. None are anticipated to correlate strongly with the error term.

\textit{Independent Variables Measuring Feedback}

Precise data regarding feedback mechanisms tied directly to a school’s improvement plan is the ideal way to measure feedback effects. The present analysis approaches feedback in two ways. The first involves measuring the presence of defined feedback loops within the school improvement plan; within the plan itself, is a regular “check” cycle proposed either for all the goals or for each goal individually? A specific frequency must be defined – for instance, quarterly checks. The definition of a specific mechanism also would be acceptable, such as teacher–principal meetings or data team meetings with a defined time or regularity – weekly, bi-weekly, quarterly, or a specific listing of dates throughout the year, for example. The presence or lack of these defined feedback loops provides the data for the dichotomous variable \textit{feedback}. 
Of course, the presence of a defined feedback approach does not offer much value unless implemented. The TWC survey provides two questions specifically related to feedback and these questions populate the data for two separate variables, datause and recvfeedbk. The first variable focuses on use of data, asking teachers to rate the school leadership on how well it facilitates use of data “to improve student learning.” The second variable focuses on teacher perceptions of the value of feedback received by asking them to assess how well the feedback helps them improve teaching. Each of these ordinal variables takes on a whole number value between 1 and 5 where 1 corresponds to strong disagreement, 4 corresponds to strong agreement, and 5 reflects a lack of knowledge or opinion about the question. For purposes of this analysis, the data for these two variables reflect the percentage of teachers who indicate agreement (3) or strong agreement (4) with the each question.

The variable datause is expected to satisfy all Gauss-Markov assumptions, whereas the potential exists for high or perfect multicollinearity between the variables feedback and recvfeedbk14. For example, for schools for which the value of feedback is “0” (no feedback loop identified in the plan), one may expect consistent “strongly disagree” or “don’t know” responses for the variable recvfeedbk14. Such a situation may also indicate a violation of the assumption of non-zero variance, if responses are consistently the same. Both conditions will be tested fo as part of the analysis.

Control Variables

Sun and Van Ryzin (2012) point out that analysis of factors related to school performance must control for appropriate school, teacher, and student characteristics. Their analysis included “12 control variables…tested extensively in the existing literature,” (p.8) The control variables include data commonly identified as part of the vast literature on education production functions
(see Hanushek, 1996) and tracked in schools and districts across the country, including North Carolina. The present analysis uses most of these same control variables, in part because of their use in prior analyses and in part because of the likely logical relationships between these variables and student and/or school performance. Control variables included in the analysis are as follows: average class size, school size, teacher experience, the percentage of teachers with advanced degrees, and composition of the student body. Also included as a control is the percentage of board-certified teachers for each school, a control relevant to NC and described in greater detail below, along with each of the other control variables. Variables for student stability and immigration status that were included in Sun and Van Ryzin’s (2012) analysis are excluded here due to lack of availability of data.

*Average class size*

Sun and Van Ryzin (2012) and others use school student-teacher ratio as a control variable, however, when measured at the school level this has the potential to be misleading. A school may employ certified staff who are not responsible for direct student instruction, however these staff members can technically be included in a student-teacher ratio calculation. If such staff members are included in the denominator of the ratio calculation, their presence would artificially lower the student-teacher ratio whereas in practice they would not actually be responsible for reducing the day-to-day teaching load on the instructors in the school.

Instead, this analysis uses another readily available value as a proxy for this ratio - a school’s average class size. Ultimately, class size can affect outcomes by affecting the level of attention a teacher can offer to each student. High class sizes mean less teacher time available per student within a given class, which may affect how well students understand course/class material, thereby affecting academic performance. By extension, a school with higher average
class size might be expected to measure lower on school performance scores relative to a school with lower average class sizes. Therefore, the relationship between class size and school performance is expected to be negative.

**School size**

School size is a demographic for which effects on school performance remain controversial. Bradley and Taylor (1998) found increasing exam performance with increasing school populations, though the relationship was non-linear, with exam performance increasing at a lower rate as population grew. However, in their study of Welsh schools, Foreman-Peck et al (2006) asserted “reducing very large schools to…at the most 500 for schools restricted to education between the ages of 11 and 16, in the long run could yield significant gains in exam performance.” (p.157) Sun and Van Ryzin (2012) measure school size using total enrollment in their analysis. The present analysis uses a slightly different measure - the final Average Daily Membership (ADM) as reported in the Principal’s Monthly Report. This has been the accepted method of calculating school size in North Carolina for many years and it is this number that is used when school size measurements are needed for official reporting purposes.

**Teacher experience**

In teaching, as in many professions, more experience is generally expected to mean better performance. Research has shown that teachers with more experience are more skilled in the classroom, and previous studies have found positive relationships between experience levels and student performance. The present study operationalizes teacher experience by including each school’s percentage of teachers with 4 or more years of teaching experience. This data is reported annually via the North Carolina School Report Card. As in other analyses, teacher experience levels are expected to correlate positively with school performance. This particular
factor is rapidly becoming more relevant in North Carolina as recent legislative changes and a sustained period of stagnated teacher salaries are beginning to take a toll. The state is just beginning to see an overall decrease in teacher experience due to retirements and to an exodus of experienced teachers. Anecdotally, it is being reported that many of these teachers leave for other states where teacher pay is better or leave the profession entirely in pursuit of other professions and there are concerns that over the next several years the situation may worsen.

**Percentage of teachers with an advanced degree**

Attainment of an advanced degree typically represents an increase in specialized knowledge in a particular area, whether the area is content-specific, such as in math or science, or teaching-specific. According to the Center for Educator Compensation Reform at the US Department of Education, “the preponderance of evidence suggests that teachers who have completed graduate degrees are not significantly more effective at increasing student learning than those with no more than a bachelor’s degree,” at least at the elementary and middle school levels. Some of these same studies found that secondary teachers with a master’s degree were more effective at increasing student achievement than those without advanced degrees, especially in the areas of mathematics and science.\(^\text{12}\)

This factor is not anticipated to have a significant relationship with school performance in the present analysis, however this is nonetheless an especially relevant control variable. Additional pay for teachers with an advanced degree has been a long-standing policy in North Carolina. Recently the General Assembly moved to stop this practice and created a groundswell

of opposition in the process. A finding of a significant relationship in either direction between this variable and school performance, while not a primary motivation or consideration of this analysis, would add an interesting data point to some of the on-going discussions about advanced degree pay for North Carolina teachers.

Composition of the student body

Meier and O’Toole (2002), Sun and Van Ryzin (2012), Fernandez (2009) and others point out that socioeconomic as well as racial and ethnic factors have been shown to correlate with student performance. This previous research has indicated that schools with higher concentrations of minority students and/or poverty tend perform less well when it comes to standardized testing. Consistent with prior research, including the studies cited above, these factors are included in the present study. Two variables are included in the analysis. The first, \%minority, reflects each school’s percentage of Native American, Asian, black, Hispanic, Pacific Islander, and multi-race/ethnicity students. The second demographic variable, \%econdis, accounts for the percentage of economically disadvantaged students at each school and is based upon the numbers of students at each school who are eligible for free and reduced meals. Both are anticipated to be negatively related to overall school performance.

Percentage of National Board Certified Teachers

The National Board for Professional Teaching Standards (NBPTS) operates a national, voluntary system to assess and certify teachers. The process requires candidates to gather a portfolio of evidence of their work that includes student work samples, lesson plans, and videos and complete a detailed analysis of that evidence. In addition, all candidates complete a full day of assessments focused on content knowledge in their main teaching area. A recent analysis by Cantrell, et al. (2008) studied the ability of the NBPTS process to recognize effective teachers by
comparing the performance of those identified for certification to those rated poorly by the process. The authors found that students randomly assigned to highly rated teachers performed better than those assigned to teachers who rated poorly in the NBPTS process. The present analysis uses each school’s percentage of NBPTS-certified teachers to control for this factor across schools. This is especially relevant to the analysis given the fact that North Carolina consistently leads the nation in the overall number of National Board-certified teachers. Currently, the state employs 20,873 NBPTS-certified teachers (nearly one-fifth of all NC teachers), and the state’s largest school district, Wake County Public Schools, employs more board-certified teachers than any other school district in the country.\textsuperscript{13} It is expected that a higher percentage of NBPTS-certified teachers will relate positively with school performance.

\textbf{Statistical Methods}

This analysis draws upon data from multiple sources. Data from all sources are measured at the school level. As a result, this analysis uses a multiple regression model, an appropriate mechanism for measurements at a single unit of analysis, to test the fit of a model designed to predict school performance based upon several independent variables associated with school improvement planning and execution practices.

This analysis tests whether School-wide Accountability Growth Estimates, a measure of school value-added, is related to a series of planning practices at the school level. A hierarchical regression is used to introduce four separate blocks of variables and to determine whether each block offers any significant predictive value, and which independent predictors in the block, if any, are significant. The blocks correspond to variable categories identified from the literature.

\textsuperscript{13} Source: http://www.nbpts.org/sites/default/files/2016_staterankings_all_nbcts.pdf; accessed May 29, 2017
and included in the model diagram and are designed to lend greater clarity to the model. The blocks include variables related to plan focus, goal clarity, stretch goals, and feedback.

**Error Term Gauss-Markov Assumptions**

For the model as a whole, the necessary assumptions with respect to the error term are anticipated to hold true. There is no reason to suspect heteroscedasticity, autocorrelation, or non-normal distribution of the error term, but the analysis will check for these.

**Considerations Regarding Potential Flaws in Research Design**

Generally speaking, the present analysis attempts to incorporate standard social science research practices to ensure accurate, reliable and informative results. However, as with nearly all social science research, the researcher cannot control every aspect of the study and so a few words should be said about potential flaws in the research design and about the generalizability of the analysis.

With respect to the population studied, the sample size was based upon the standard sample size calculation for the population under study – elementary and middle schools in North Carolina – that would ensure a 95% confidence level at a 5% confidence interval. However, two separate models were run – one that included imputed data for several critical variables, and one that excluded the schools with missing data, which reduces confidence in the results and, therefore, reduces generalizability. Chapter 4 provides details of data imputation and exclusion of schools from the sample.

The sample was drawn randomly from all schools in the population, using standard automated spreadsheet functionality to number make the selection. Because the sample includes only elementary and middle schools, any inferences that can be made about the impact of school
improvement cannot be extended to high schools in North Carolina or elsewhere. Regarding such potential inferences, several additional research design factors are considered below.

The external validity of the present study potentially is affected by several factors related to the nature of the analysis and the data used. First, the study is non-observational and with respect to the school improvement plans it cannot be known to what extent the information in the plans reflects what occurs in the schools. A school team may develop what on paper is a “perfect” plan that follows best practices for planning, but we cannot know from the plan how well it is integrated into the life of the school. A strong improvement plan built using good planning practices will be ineffective if put on a shelf and not discussed again until the next planning cycle. The present analysis attempted to address this with the inclusion of data points regarding feedback to teachers. External validity is addressed in more detail in Chapter 5.

Second, data for the analysis comes from multiple existing sources of information about North Carolina schools, and NCDPI-provided data is the same data used by the agency for state and federal reporting and can be considered the best available data. However, the usual caveats apply. Much of the data originates at the school or district level and inconsistencies in collection are possible for some data due to local conditions and/or human error. The state agency has refined its collection processes over many years, however, and routinely provides support and professional development for these procedures and the potential impact of data errors is small. Along the same lines, the Teacher Working Conditions Survey, another significant data source, has been shown to be a valid, reliable instrument and has been in use for nearly fifteen years. Nonetheless, like other survey instruments it cannot be considered a perfect data source due to the usual pitfalls. Respondents may misunderstand questions, misremember facts, interpret
events at the school level subjectively and therefore differently, or even embellish or lie in their responses.

Additionally, it is best to reiterate some information originally discussed very early in the biographical notes. The author of the present study spent 10 years working at the North Carolina Department of Public Instruction, and for part of this tenure was engaged in working closely with school districts and individual schools on their school improvement planning processes. While this represents a potential source of bias with respect to the analysis of individual school-level plans, the author has made every effort to guard against such bias and has included detailed information about how plan analysis was conducted and how results were coded to yield the raw data related to school improvement plans.

Finally, this single-state analysis significantly reduces generalizability. The study is based upon North Carolina conditions whereas a multi-state analysis may offer better generalizability due to differences across states. For example, school improvement planning in North Carolina is a legal requirement for every school. This analysis offers no insight into whether the compulsory nature of the planning process impacts effectiveness. It may be that where such planning is voluntary, those who undertake it are more invested in its success and therefore exert greater effort during planning and implementation. Nonetheless, the present study may offer clues as to whether or not certain variables are worthy of further examination here and in other states.

Summary

Chapter 3 is designed to offer an overview of the research design and provide details regarding the study population and sample, the sources for the data used in the analysis, and an exposition of the specific variables assessed. A brief overview of the statistical methods
employed was included, as well as a discussion of research design limitations. The following chapters will provide the additional methodological details and results of the statistical analyses.
Chapter 4 – Analysis and Results

The present analysis includes data from multiple sources, as described previously. Obtaining primary data from the school improvement plans necessitated establishing procedures for conducting the necessary counts in a consistent manner. Some secondary data required transformation as well. This chapter provides information about data collection and preparation, analytical procedures, and finally, results of the analysis.

Data Collection and Preparation

Once the schools were selected and data sources were identified, data had to be obtained from each source. Because much of the data comes from official State of North Carolina collections and reports, most data sources offered easy access to this public data.

Dependent variable data

Dependent variable data for each school in the analysis were readily available from the public section of the EVAAS website discussed previously and found at https://ncdpi.sas.com/. Each school was accessed individually and School Accountability Growth Estimates for each school year from 2013-14 through 2015-16 were entered into a spreadsheet. The 2014-15 value in its raw form provides one dependent variable. The three-year trend measurement provides the second dependent variable. This was calculated by subtracting the starting point value (2013-14) from the endpoint value (2015-16).

Independent variable data

Independent variable data derived from a variety of sources including individual school improvement plans, NC school accountability results, and the NC Teacher Working Conditions
survey. Details of these sources are included in Chapter 3. The collection and preparation of each is addressed below.

**School improvement plan collection and analysis**

School improvement plan collection presented the first hurdle to data-gathering for the present analysis. Upon identification, the principals of each school in the sample were contacted via email and asked to provide an electronic copy of their school’s improvement plan for use in this research. Improvement plans were requested after schools received the required local board approval for their plans. Following two separate requests of all the schools in the sample, 80 plans were received in response (a 23% response rate). Due to this very low response rate, the remaining plans for the schools were then sought on a school-by-school basis via Internet searches of individual school websites and, in some cases, contact with district level personnel. Ultimately, 273 plans were obtained (86%) and comprised the final sample used in the analysis. This number is lower than the planned-for 318 school improvement plans and results in a 5.47% confidence interval. These plans represent 84 of North Carolina’s 115 school districts, or 73%, and include urban, suburban, and rural school districts.

The plans obtained for inclusion in the sample cross multiple years. State law mandates that plans be completed no less often than every two years. However, in practice school plans vary in terms of the time period covered (specific years range from the 2012-13 school year to the 2015-16 school year) and also vary in the length of time covered (some plans are two-year, in accordance with the current state cycle, while others are annual, and a few even covered a three-year cycle, as required when the ABCs legislation initially was passed). This is not ideal from the standpoint of measuring plan impact on a given outcome for a given school year, but is mitigated somewhat by the fact that the plan format and approach are typically guided by the district office
and tend to be relatively consistent over time. As such, this analysis assumes that a given plan may be considered representative of a school’s planning process over a 3-5 year period of time. Periodically, the state department will offer a new template and/or revised guidance and districts may change their approach as a result. The last two major revisions to NCDPI recommendations occurred in 2009 and 2016, effectively bookending the period from which this analysis draws its plans, because one dependent variable measures change in growth across the school years from 2013-14 to 2015-16.

Plan analysis was essentially a tedious but straightforward counting exercise, and required some initial analysis and preparation. All counts were conducted by the author. To avoid subjectivity in conducting the counts and to address concerns about reliability, the counting process was “calibrated” in advance. Calibration attempts to increase reliability by defining “rules” for how the count is conducted. (Garson 2002) For this process, the author randomly selected 60 plans from the final sample of 273 plans. These 60 plans were reviewed to calibrate and to define specific “rules” for counting and to develop count sheets used to record counts for each plan. These rules, discussed in greater detail below, were then used to assess the full group of 273 plans.

**Goal count rules**

Goals are the highest-level outcomes specified, and identifying goals in the plans was a straightforward exercise. The vast majority of the plans defined goals using exactly that term. However, a few used terms such as “outcome” or “objective” instead of “goal.” The initial review resulted in two key rules used for counting goals, described below.

First, goal breakdowns are counted as established by the school planning teams. Even if a single goal contained multiple targets (for instance, a different target for different subgroups of
students) the goal was counted as only one goal. The following goal from a Guilford County elementary school illustrates this scenario:

“By June 2016 Jefferson Elementary School will increase the number of SWD students scoring proficient in reading from 23% to 43.6%; AA students from 45% to 57%; and LEP from 25% to 51.6%.”

This goal identifies three separate subgroups of students, each of which have different targets defined. The specific strategies chosen by the planning team apply to achieving all three targets and this was counted as one goal for the school rather than as three separate goals.

Second, when multi-year school improvement plans (for example, a plan inclusive of school years 2014-15 and 2015-16) contain both a full-term goal as well as an interim goal (typically the goals reflect a building block approach), each is counted separately as a different timeline applies to each. The goal shown below, in connection with the goal immediately above, illustrates this concept.

“By June 2015 Jefferson Elementary School will increase the number of SWD students scoring proficient in reading from 23% to 35.5%; AA students from 45% to 51%; and LEP from 25% to 44.6%.”

In this case, though listed together, the goals are counted as two. Both of these rules were applied to the goal count for each plan, as appropriate. The resulting counts yielded a range of 1 to 18 goals across all 273 school improvement plans. The number of goals averaged just fewer than 5 and the median value for all plans was 5.

---

14 SWD means Students With Disabilities; AA means African-American; LEP means Limited English Proficiency; these acronyms are commonly used in schools and districts across North Carolina.
Strategy count rules

As discussed previously, the present analysis includes both goals and strategies to ensure a comprehensive picture of plan focus; because the researcher experience indicates that school planning teams may introduce “shifts” between goals and strategies so as to meet external, pre-determined criteria such as a maximum number of goals.

Strategies reflect one level deeper than goals; they are the means by which school teams intend to achieve the goals they have identified. The initial 60-plan analysis indicated no need for special counting rules for strategies as these typically reflected common, formal individual programs (i.e., PBIS, or Positive Behavioral Intervention and Support) as well as actions such as “Train teachers in the use of math manipulatives.” Strategies were commonly separated from one another and identifiable as separate and distinct. The strategy counts yielded a range of 0 to 77 strategies goals across all 273 school improvement plans. The average number of strategies fell just under 16 and the median value was 13.

As previously discussed, the strategy counts were combined with the goal counts to arrive at a total count for each school, which served as the school-level measure of plan focus.

Target count rules

The preliminary plan analysis also indicated that counting targets for the calculation of the associated variable would not be complicated, and it allowed again for defining specific counting rules. As discussed earlier, the continuous, quantitative variable “target,” reflects the calculated percentage of plan goals that include an unambiguous quantitative target or measurable status or state of being. The initial analysis confirmed that for a target to count as unambiguous, it must identify at least one of the following:
• a specific, measurable quantity or amount to be achieved, i.e., “achieve a 75 percent proficiency level,”
• a specific, measurable amount of increase or decrease, i.e., “achieve a 10 percent increase,”
• a specific, measurable status or recognition, such as “achieve NC School of Excellence status”

Based upon these rules, terms such as “all” or “every” were treated as valid, unambiguous targets equivalent to 100%, whereas targets using language limited to “increase” or “decrease” (as in “increase proficiency…” or “increase participation…”) without a specific level or amount were ambiguous and therefore were not counted.

The specific procedure for arriving at the variable value for a school involved looking at each goal in the plan and determining whether it contained a specific target as defined above. A plan’s target count was divided by its goal count to arrive at a percentage value for the variable.

For example, consider the following school goals:

• Goal 1: Our school will increase our percentage of students who show proficiency on the reading EOG by 10% to 62.4%.

• Goal 2: Our school will increase proficiency on the math EOG by 10% to 62.8%.

• Goal 3: Our school will maintain a safe and supportive school environment for all students by setting high behavior expectations and ensuring that protocols are in place in the event an emergency situation would occur.

• Goal 4: Our school will strive to increase the knowledge and skill level of all teachers so that technology can be effectively integrated into instruction.

15 The NC state accountability model offers several recognition categories for schools based upon student achievement.
The plan that contains these 4 goals has a target count of 2 because only goals 1 and 2 have specific, measurable targets. The school’s final “target” variable value is .5 (2/4). For each school, the variable values were calculated in a spreadsheet based upon entered counts. Values for this variable ranged from 0 to 100 percent, with a mean of just above 73%.

**Date count rules**

“Date” is another quantitative, continuous measure. Preliminary plan analysis identified the clear rules for counting a goal as having a clear, unambiguous date for measuring achievement. For the goal to be counted as having a date, it must provide one of the following:

- a specific target date, such as “June 10, 2015”
- a specific, measurable point in time, such as “the end of the third quarter”
- a window with a definable endpoint, i.e., “December 2015”\(^{16}\)
- an identifiable timeline that occurs at a defined point, i.e., “….on the EOG test”

Unacceptable end dates include anything identified as “on-going,” or any other ambiguous timeline because the school can never be judged to have missed the goal. The determining factor in every case was whether the date is sufficiently clear to provide accountability – a point at which the school can clearly be said to have achieved or missed the goal. A date that provides accountability also provides an incentive for teachers to meet the goal.

As with targets, the date count for a plan was used to calculate the percentage of goals containing an unambiguous date by which a target is to be achieved. The calculated percentage provided the value for the variable, which ranged from 0 to 100%. The mean percentage for this variable was about 87%.

---

\(^{16}\) A window can be considered ambiguous because it could mean the first day of the month or the last day of the month, or any other date between the two. However, it allows an outside observer to say a goal has been missed if it has not been reached by January 1 and therefore is counted as having a date.
Feedback Loops

The dichotomous feedback variable obtained from plan analysis reflects the presence or absence of a defined feedback cycle in a school’s improvement plan. As previously described, the rules for counting a plan as having feedback loops include:

- identification of a specific feedback frequency, such as “quarterly”
- specification of a feedback mechanism with a defined time, i.e., “meetings on October 1, 2014, January 5, 2015…”
- identification of a feedback mechanism with a defined regularity, such as “weekly meetings”

Because this variable is dichotomous, plans were given credit for having feedback loops (coded as a “1”) even if only one of its goals included an indicator of regular feedback. Plans with no indication of a feedback loop anywhere were coded as “0.” As mentioned earlier, other variables from the NC Teacher Working Conditions Survey were included in the analysis to assess both the presence and value of feedback. About 70% of the plans contained at least one feedback loop within the plan and therefore were coded as a “1.”

Teacher Working Conditions Survey data

Teacher working conditions survey data was provided by the New Teacher Center and is also publicly available on the North Carolina Teacher Working Conditions website at https://ncteachingconditions.org/results/68. The data were gathered from survey responses to two statements in the School Leadership portion of the 2014 survey. These statements, from Question 7.1 of the overall survey, included:

Please rate how strongly you agree or disagree with the following statements about school leadership in your school.
f. The school leadership facilitates using data to improve student learning.

g. Teachers receive feedback that can help them improve teaching.

Data for each school were used without transformation as the results provided the combined percentage of teachers who either agreed or strongly agreed with the statements.

**School performance and target data**

Individual school performance data and annual measurable objective targets were obtained from the NCDPI website at http://www.ncpublicschools.org/accountability/reporting/ and were used to develop the dummy variables required to analyze goal difficulty. Creating these variables required identification of 2013-14 actual school performance and 2014-15 Annual Measurable Objective targets. Goal difficulty variables were calculated by subtracting 2013-14 performance levels from 2014-15 targets to determine each school’s distance from target achievement. This raw distance-from-target result was used to categorize each school’s reading and math goals as easy (distance less than 10 points), difficult, or stretch (distance between 10 and 20 points, inclusively), and very difficult (distance greater than 20 points).

**Missing Values and Data Imputation for Independent Variables**

The present analysis includes several instances of data imputation to address missing values. Data imputation occurred based upon missing data in two sources – the Teacher Working Conditions Survey and school improvement plans.

With respect to the Teacher Working Conditions Survey, which provided school-level feedback data for two variables, six schools were missing data for the school year included in the analysis, 2013-14. Data were imputed for these six schools (a total of 12 data points in the overall analysis) using the sample average for each variable.
Regarding the school improvement plans, a total of 45 schools in the sample had no plan available for review. Data for these schools were imputed using the sample average for each variable. However, because the present analysis is focused on assessing the value of school improvement plans and the planning process, the author believed this lack of plan availability to be significant and therefore, conducted the final analysis of the two models using both samples – one that includes all schools with imputed data and a second that includes 273 schools, excluding each school with a missing plan. The results reported reflect the 273-school analysis, though a very brief indication of the differences is reported for each dependent variable analyzed.

**Control Data**

Control data for the 2014-15 school year were obtained from the NC Department of Public Instruction sources cited previously: School Report Cards, the Grade, Race and Sex data collection, and Free and Reduced Meals data. However, some raw data required minor transformations to yield the variables included in the analysis.

Because the data sources for control variables are based upon long-established collection procedures and are used for numerous analysis and reporting activities, collectively these data contained very few missing values. However, where missing values were identified, the sample average for that variable was used to impute a value.

**Reliability**

Reliability addresses how well a particular item or scale will yield similar results when administered or measured by a different person or at a different time. Measures are reliable if they are consistent, irrespective of when they are measured or by whom.

Most of the data used for this analysis comes from external sources that produce the data over and over again for reporting and analysis and emphasize the reliability of the data. In this
study, reliability concerns relate primarily to the review of school improvement plans. The counting of goals and strategies that measure each plan’s degree of focus, identification of what qualifies as a clear target and a clear date for the measures of goal clarity, and awarding credit for the presence of feedback loops during the plan reviews must be completed in a consistent manner. As mentioned previously, all counts were conducted by the author after a calibration process that identified specific rules to be used for review of the plans and to capture data in a consistent manner.

**Validity**

Like many social science concepts, the concepts being measured in the present study can be difficult to measure in a concrete way. They are very much unlike measuring the length of a piece of wood or some amount of liquid, for both of which we have standard tools and practices for making accurate and consistent measurements. Assessing the validity of research allows the researcher to ensure that any threats or potential biases that may affect the research are considered and, if necessary, addressed appropriately. Achieving validity means that a study measures what it purports to measure and that the conclusion drawn from it are logical (Garson 2002) While there are many kinds of validity, they do not all apply to every study. Based upon the topic and concepts and the research design, the present analysis does not need to consider several types of validity such as mortality bias, evaluation apprehension, or control group-related concerns such as control awareness or unintended treatments. However, there are a few types of validity to consider. They are addressed individually below.

**Internal validity**

Internal validity is concerned with the introduction of unknown sources of bias. If an analysis lacks internal validity, observed effects may be the result of variables other than those
being considered in the study. There are many kinds of internal validity, but the primary consideration for the present study is selection bias, which is described below.

**Selection Bias**

Selection bias considers how closely the sample used in a study approximates a random sample. As discussed in Chapter 3, the present study determined sample size for the population of NC elementary and middle schools using the standard formula for sample size determination. The sample was randomly selected by placing a sequentially numbered list of the schools into an Excel spreadsheet. Numbers were randomly drawn from the list using Excel’s built-in RANDBETWEEN function, then school names were matched to their number using the VLOOKUP function.

**Construct validity**

Construct validity addresses the logic of the items or measures of a concept. Good constructs have clear definitions that allow specific indicators to be selected to measure them. (Garson 2002) As with internal validity, there are multiple types, but the present study is concerned mainly with discriminant validity, a type of construct validity concerned with how highly variables correlate with one another. Variables that highly correlate essentially are measuring the same thing; high Pearson’s r values between two variables, which measure bivariate correlations, suggest that one of the two variables should be dropped from the analysis.

The present study will assess bivariate correlations as an early step in the statistical analysis process.
**Face validity**

Face validity is concerned with whether a concept or measurement item logically seems to measure what it claims to measure. For the present analysis, each of the measures used to operationalize the concepts being assessed were carefully considered prior to being finalized to minimize threats to face validity.

**External validity**

External validity is concerned with the generalizability of analysis results. The current study has some limitations on generalizability. The sample is drawn from a single state and so is representative only of that state. Generalizing results to other states, nationally or internationally would not be a valid use of the results. However, the concepts being assessed and the effects that may be discovered nonetheless may contribute to the broader bodies of research in education and public management.

**Data Screening**

Data were screened for all schools in the final sample using standard techniques. Statistics and plots reported below reflect first the smaller 273-school sample, then the 308-school sample. Deviations from this are noted.

Frequency distributions provide information about how individual variables are distributed, and help to organize and understand the data. This analysis showed that variables generally are normally distributed. The independent variable for the percentage of plans with clear dates was just outside the normal acceptable range at -2.3 (small sample) and -2.4 (larger sample), skewing slightly to the left. A high proportion of plans included clear due dates for achievement of their goals and so this is not entirely unexpected.
Kurtosis values indicate how peaked or flat data are; leptokurtic distributions are more peaked in the center, while platykurtic distributions are flatter than a typical normal distribution. In both the 273-school and 308-school samples, high kurtosis values were found for the variables representing the total number of goals and strategies and again for the percentage of plans having clear due dates. High kurtosis can be addressed with transformations to the data, but because such transformations complicate the understanding of effects, and because the kurtosis values were not exceedingly high, no such transformations were applied to the data for this analysis.

A basic assumption of regression analysis is that independent variables are not significantly correlated with one another. Analysis of bivariate correlations between independent variables provides an indication of whether two variables are closely related. Pearson correlations ($r$) greater than .7 suggest a need to drop one of the two related variables from the analysis (Tabachnick & Fidell, 2001). None of the independent variables in the analysis exceeded this threshold; only one correlation – that between easy reading goals and easy math goals – approached the threshold ($r = .646$ [small sample] and $r = .640$ [larger sample]). Correlations for independent variables are displayed in Tables 4.1 and 4.2.

Field (2005) stated that assessing correlations “is a good ‘ball park’ method but misses more subtle forms of multicollinearity.” (p. 175) He suggested assessing the variance inflation factor (VIF) and/or its reciprocal, the tolerance statistic, as alternatives. VIF values of 10 or higher are cause for concern; tolerance values below .2 are also worthy of concern. Within the present analysis, the highest VIF value encountered for any variable was 2.557, and the lowest tolerance value found was .391. Both of these results suggest that multicollinearity is not a problem.
Table 4.1 – Correlations of Independent Variables

<table>
<thead>
<tr>
<th></th>
<th>goalstrtotal</th>
<th>target</th>
<th>date</th>
<th>feedback</th>
<th>easyread</th>
<th>stretchread</th>
</tr>
</thead>
<tbody>
<tr>
<td>goalstrtotal</td>
<td>1</td>
<td>-.137*</td>
<td>-.170**</td>
<td>-.102</td>
<td>-.087</td>
<td>.014</td>
</tr>
<tr>
<td>target</td>
<td>-.137*</td>
<td>1</td>
<td>.359**</td>
<td>.309**</td>
<td>-.080</td>
<td>-.035</td>
</tr>
<tr>
<td>date</td>
<td>-.170**</td>
<td>.359**</td>
<td>1</td>
<td>.216**</td>
<td>-.082</td>
<td>-.039</td>
</tr>
<tr>
<td>feedback</td>
<td>-.102</td>
<td>.309**</td>
<td>.216**</td>
<td>1</td>
<td>.053</td>
<td>-.133*</td>
</tr>
<tr>
<td>easyread</td>
<td>-.087</td>
<td>-.080</td>
<td>-.082</td>
<td>.053</td>
<td>1</td>
<td>-.516**</td>
</tr>
<tr>
<td>stretchread</td>
<td>.014</td>
<td>-.035</td>
<td>-.039</td>
<td>-.133*</td>
<td>-.516**</td>
<td>1</td>
</tr>
<tr>
<td>vdifread</td>
<td>.080</td>
<td>.119*</td>
<td>.126*</td>
<td>.068</td>
<td>-.586**</td>
<td>-.391**</td>
</tr>
<tr>
<td>easymath</td>
<td>-.029</td>
<td>-.025</td>
<td>-.111</td>
<td>-.021</td>
<td>.646**</td>
<td>-.161**</td>
</tr>
<tr>
<td>stretchmath</td>
<td>.006</td>
<td>-.042</td>
<td>.018</td>
<td>-.103</td>
<td>-.191**</td>
<td>.163**</td>
</tr>
<tr>
<td>vdifmath</td>
<td>.026</td>
<td>.069</td>
<td>.106</td>
<td>.125*</td>
<td>-.536**</td>
<td>.019</td>
</tr>
<tr>
<td>datause14</td>
<td>-.102</td>
<td>.008</td>
<td>.048</td>
<td>-.126*</td>
<td>.031</td>
<td>.035</td>
</tr>
<tr>
<td>recvfdbk14</td>
<td>-.080</td>
<td>-.066</td>
<td>.069</td>
<td>-.064</td>
<td>.083</td>
<td>-.027</td>
</tr>
</tbody>
</table>

N=273
*Correlation is significant at the 0.05 level
**Correlation is significant at the 0.01 level.

Table 4.2 – Correlations of Independent Variables

<table>
<thead>
<tr>
<th></th>
<th>vdifread</th>
<th>easymath</th>
<th>stretchmath</th>
<th>vdifmath</th>
<th>datause14</th>
<th>recvfdbk14</th>
</tr>
</thead>
<tbody>
<tr>
<td>goalstrtotal</td>
<td>.080</td>
<td>-.029</td>
<td>.006</td>
<td>.026</td>
<td>-.102</td>
<td>-.080</td>
</tr>
<tr>
<td>target</td>
<td>.119*</td>
<td>-.025</td>
<td>-.042</td>
<td>.069</td>
<td>.008</td>
<td>-.066</td>
</tr>
<tr>
<td>date</td>
<td>.126*</td>
<td>-.111</td>
<td>-.018</td>
<td>.106</td>
<td>.048</td>
<td>.069</td>
</tr>
<tr>
<td>feedback</td>
<td>.068</td>
<td>-.021</td>
<td>-.103</td>
<td>.125*</td>
<td>-.126*</td>
<td>-.064</td>
</tr>
<tr>
<td>easyread</td>
<td>-.586**</td>
<td>.646**</td>
<td>-.191**</td>
<td>-.536**</td>
<td>.031</td>
<td>.083</td>
</tr>
<tr>
<td>stretchread</td>
<td>-.391**</td>
<td>-.161**</td>
<td>.163**</td>
<td>.019</td>
<td>.035</td>
<td>-.027</td>
</tr>
<tr>
<td>vdifread</td>
<td>1</td>
<td>-.542**</td>
<td>.051</td>
<td>.558**</td>
<td>-.066</td>
<td>-.065</td>
</tr>
<tr>
<td>easymath</td>
<td>-.542**</td>
<td>1</td>
<td>-.558**</td>
<td>-.569**</td>
<td>.070</td>
<td>.027</td>
</tr>
<tr>
<td>stretchmath</td>
<td>.051</td>
<td>-.558**</td>
<td>1</td>
<td>-.365**</td>
<td>.111</td>
<td>.071</td>
</tr>
<tr>
<td>vdifmath</td>
<td>.558**</td>
<td>-.569**</td>
<td>-.365**</td>
<td>1</td>
<td>-.189**</td>
<td>-.101</td>
</tr>
<tr>
<td>datause14</td>
<td>-.066</td>
<td>.070</td>
<td>.111</td>
<td>-.189**</td>
<td>1</td>
<td>.551**</td>
</tr>
<tr>
<td>recvfdbk14</td>
<td>-.065</td>
<td>.027</td>
<td>.071</td>
<td>-.101</td>
<td>.551**</td>
<td>1</td>
</tr>
</tbody>
</table>

N=273
*Correlation is significant at the 0.05 level
**Correlation is significant at the 0.01 level.
Because this analysis includes two models with dependent variables that measure school performance – one for a single school year and a second for the change over a three-year time period that includes the single year measure – the researcher chose also to assess the correlation of these variables and found no multicollinearity between 2014-15 school performance and the 3-year performance trend from 2013-14 through 2015-16 ($r = .084$ for the 273-school sample; $r = .073$ for the 308-school sample), suggesting that differences exist between the two measures and that the models are not measuring the same impacts.

An assumption of linear regression analysis is that error distribution is constant, or homoscedastic. If this condition is not met for the independent variables, the beta-weight of an independent variable may be overestimated and appear to be more significant than it truly is, a condition known as Type I error. To test for this, one can plot standardized residuals against standardized predicted values. The resulting scatterplot helps assess how well the data meet this assumption.

Figure 4.1 displays the plot for the dependent variable measuring school performance in the smaller 273-school sample. The plot indicates that the data meet the assumption of homoscedasticity; it appears neither funnel-shaped (which would indicate heteroscedasticity) nor non-linear. The scatterplot for the second dependent variable measuring change in performance, also from the 273-school sample, is shown in Figure 4.2.
Figure 4.1 – Plot of Residuals and Predicted Values for Dependent Variable *sage1415*
Figure 4.2 – Plot of Residuals and Predicted Values for Dependent Variable \(3yrtrend\)

Like the plot for 2014-15 performance, the plot above indicates that the data for the change in performance meet the assumption of homoscedasticity.

A further assumption of linear regression is that all independent variables have some variation among their values in the data. Three variables in the analysis should be considered to have near-zero variance, and one in particular, \(datause14\), which measures teachers’ ratings of school leadership on how well it facilitates use of data “to improve student learning,” is very near zero. Table 4.3 summarizes the details for the 273-school sample.
Table 4.3 – Independent Variables with Near-zero Variance

<table>
<thead>
<tr>
<th></th>
<th>date</th>
<th>datause14</th>
<th>recvfdbk14</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Valid Missing</td>
<td>273</td>
<td>273</td>
<td>273</td>
</tr>
<tr>
<td>Mean</td>
<td>.878</td>
<td>.959</td>
<td>.865</td>
</tr>
<tr>
<td>Median</td>
<td>1.00</td>
<td>.971</td>
<td>.889</td>
</tr>
<tr>
<td>Mode</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>.286</td>
<td>.049</td>
<td>.118</td>
</tr>
<tr>
<td>Variance</td>
<td>.082</td>
<td>.002</td>
<td>.014</td>
</tr>
<tr>
<td>Minimum</td>
<td>.000</td>
<td>.773</td>
<td>.316</td>
</tr>
<tr>
<td>Maximum</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

One possible solution is to drop these variables from the analysis. However, doing so may lead to the loss of informative results. The variable `date` addresses an important aspect of goal clarity and no other variable offers a similar measure. The range of values for `datause14` is reasonable, but 110 cases (about 40%) are at the maximum value. Despite having a near-zero variance, the variable `recvfdbk14`, which measures teacher perceptions of the value of feedback received, is more broadly distributed and fewer cases are focused on any one value, including 34 cases (12.5%) at the maximum value. Because its variance is so close to zero and because it is one of three variables in the model related to feedback, `datause14` is excluded from the final analysis.

A final assumption regarding variables in a regression analysis is a lack of autocorrelation, or that error terms for observations are uncorrelated. Autocorrelation is more likely in time-series models (Berry 1993, Schroeder et al. 1986), and is not likely to be a problem for the present analysis, but the presence of autocorrelation can indicate important problems with data, including the omission of an important explanatory variable (Schroeder et al, 1986). The Durbin-Watson statistic provides a test for autocorrelation and generally acceptable values should fall between 1.5 and 2.5 (Garson 2012). The Durbin-Watson coefficients for the two
models included in the present analysis each fell within this range, indicating a lack of autocorrelation and meeting this regression assumption.

Final Regression Model and Regression Equations

The removal of the datause14 variable due to its lack of variance resulted in a slight revision to the final overall model, shown in Figure 4.3.

![Figure 4.3 – Final Overall Research Model](image)

The equations below provide the specific regression models tested and expected relationships between the dependent variables and all other variables.

Model 1:  
\[ Y_1 = \beta_0 - \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 - \beta_5 X_5 - \beta_6 X_6 - \beta_7 X_7 - \beta_8 X_8 + \beta_9 X_9 - \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} - \beta_{15} X_{15} - \beta_{16} X_{16} + \alpha \]

Model 2:  
\[ Y_2 = \beta_0 - \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 - \beta_5 X_5 - \beta_6 X_6 - \beta_7 X_7 - \beta_8 X_8 + \beta_9 X_9 - \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} - \beta_{15} X_{15} - \beta_{16} X_{16} + \alpha \]
In these models, $Y_1 =$ school performance in 2014-15 and $Y_2 =$ the change in school performance from school year 2013-14 to school year 2015-16, and independent and control variables are as shown below:

- $X_1 =$ the total number of plan goals and strategies,
- $X_2 =$ the percentage of plan goals having clear targets,
- $X_3 =$ the percentage of plan goals having clear due dates,
- $X_4 =$ the presence of feedback loops in the school improvement plan,
- $X_5 =$ easy reading goals,
- $X_6 =$ very difficult reading goals,
- $X_7 =$ easy math goals,
- $X_8 =$ very difficult math goals,
- $X_9 =$ the percentage of teachers who agree that feedback received helps to improve teaching,
- $X_{10} =$ average class size,
- $X_{11} =$ school size,
- $X_{12} =$ the percentage of teachers at a school with four or more years of teaching experience,
- $X_{13} =$ the percentage of teachers at a school with an advanced degree,
- $X_{14} =$ the percentage of teachers at a school certified by the National Board of Professional Teaching Standards,
- $X_{15} =$ the percentage of economically disadvantaged students at a school, and
- $X_{16} =$ the percentage of minority students at a school
Regression Analysis Results

The present analysis is a hierarchical regression. A hierarchical regression allows a researcher to introduce one block of predictors into the analysis at a time. This permits the researcher to make several assessments:

- Does a block of variables contribute to a change in explained variance? Statistics for the significance of the change in $R^2$ attributable to the block (F-change and significance of the F-change) provide this information for each block in the analysis.
- Is a block of variables a significant predictor of a dependent variable? Significance of the F-statistics for each block provide this detail.
- Are the individual variables within a block significant predictors of a dependent variable? For each coefficient, t-scores and significance levels are provided to allow assessment of each predictor.
- Does the complete model with all variables entered significantly predict the dependent variable?

In other words, a goal of hierarchical regression is to determine whether new variables that are added significantly improve the proportion of variance in the dependent variable explained by the model. If it does, as evidenced by a significant change in the value of $R^2$, the new variables increase the explanatory value of the model. Such an approach also allows the researcher to assess the complete model. Because a hierarchical regression builds by adding all the selected predictor variables, the final block always results in a model that reflects the results that would be achieved by simultaneously entering the selected variables into a regression analysis.

Based upon the overall model presented in this analysis, a hierarchical approach is useful because of the presence of groups of variables that measure planning concepts suggested by the
literature, such as having a focused plan, developing clear goals, choosing goals that are appropriately difficult, and including and implementing effective feedback mechanisms.

**Structure of the Hierarchical Regression Analysis**

Two separate hierarchical regression analyses were conducted to examine the extent to which five sets of predictor variables accounted for differences in school performance in a single school year and differences in change in school performance over a period of three school years. For both analyses, blocks were identically composed and introduced.

Block 1 introduced the 7 control variables – many of the relevant school demographic factors suggested by the literature review. These included measures related to the school, its staff, and its students.

Block 2 included a single variable measuring the total number of goals and strategies included in the plan, the measure of plan focus described previously.

Block 3 included the two variables measuring goal clarity. These included measures of the percentage of goals having clear targets and the percentage of goals having clear due dates or end dates.

Block 4 introduced the two variables related to feedback. The first measured inclusion of feedback loops as part of the planning process and the second addressed the value of feedback to teachers for improving instruction.

Block 5 added two dummy variables measuring the difficulty of a school’s reading goal based upon the Annual Measurable Objective target. The first variable measured easy goals and the second measured very difficult goals. “Stretch” goals serve as the reference category.

Finally, Block 6 included two additional dummy variables, again measuring goal difficulty, in this case for school math goals, and again based upon the AMO targets described
earlier. As with Block 5, variables were included for easy and very difficult goals, with “stretch”
goals as the reference category.

The results presented here provide a closer look at how each block of variables help to
explain the variance in school performance and the change in school performance as
operationalized by the dependent variables. For each model, school performance and change in
school performance, the analysis below uses the results to provide the assessments noted above.

Hierarchical Regression Results for School Performance

This hierarchical regression model examined the extent to which the six blocks of
predictors described above accounted for school performance during the 2014-15 school year. As
mentioned previously, the results presented here reflect those for the 273-school sample.
Identical regression analyses were conducted for the larger sample, which contained imputed
school improvement plan data for the schools with no plan. The results were similar, with sign
coefficients for all variables the same and with the same significant relationships for the number
of goals and strategies and for average class size. Results showed the larger model explained 9%
of the variance in the dependent variable. Table 4.4 provides results for blocks 1 through 3;
Table 4.5 provides results for blocks 4-6.

As shown in Table 4.4, the school demographic factors in Block 1 accounted for about
5% of the variance in school performance during for the 2014-15 school year. The variable
representing average class size significantly and negatively related to school performance,
indicating decreases in school performance with increasing average class sizes. This is consistent
with expectations of the model. The remaining variables in the block were not significantly
related to school performance, and several – variables for teacher experience, teachers with
advanced degrees, and board-certified teachers – showed negative relationships to school performance and, therefore, were not in accordance with expectations of the model.

Block 2 introduced the variable related to school improvement plan focus, specifically a measure of the number of goals and strategies contained in a plan. Block 2 resulted in a significant change in explained variance of a little over 2%, indicating that the addition of the variable improved on the original model. Unexpectedly, this variable was positively related to school performance and significant, suggesting that increases in the number of goals and strategies can be expected to yield increases in school performance. The main effects of average class size seen in Block 1 were essentially unchanged, remaining negative and significant.

Goal clarity measures added in Block 3 resulted in no significant increase in variance explained. Neither the clear target measure nor the clear date measure was significant and, again unexpectedly, both showed a negative relationship, suggesting that increasing percentage of goals with clear targets and due dates in a school improvement plan leads to decreased school performance. Class size and goal number measures remained significant and of roughly the same magnitude.

Block 4, shown in Table 4.5, added the two feedback-related variables. The variable measuring the presence of feedback loops in the school improvement plan was negatively related to performance, but the second feedback variable was positively related, in accordance with the hypothesis that receiving feedback useful for improving instruction leads to higher performance. Unfortunately, neither variable showed a significant relationship to school performance. Again, class size and goal number measures remained significant and have roughly the same magnitude.

The “easy” reading goal and “very difficult” reading goal variables were added in Block 5. “Stretch” goals serve as the reference variable for these two dummy variables. The
relationship between “easy” reading goals and school performance was negative – the expected direction – indicating that “easy” reading goals produce lower performance than “stretch” goals. While this reflects the expected relationship, the coefficient was not significant. The “very difficult” reading goal’s non-significant relationship to school performance was positive and therefore counter to expectations, suggesting that very difficult goals lead to higher performance.

Block 6 added the final two variables, both dummy variables measuring goal difficulty and analogous to those in Block 5, but for math goals instead of reading goals. Again, stretch goals serve as the reference for these variables. Neither of the two variables showed a significant relationship to school performance, and the “easy” goal variable’s positive sign again ran counter to expectations, suggesting that easier goals lead to better performance than stretch goals. The “very difficult” variable’s negative sign aligned with expectations, however. Class size and goal number variables remained significant, as they did in every model in which they were included. Within this final, fully specified model the magnitude of their coefficients fell a small amount compared to their original values, in both cases.

The unexpected directions for the “very difficult” reading and “very easy” math goals may be indicative indicate that the threshold for defining an “easy” versus a “stretch” versus a “very difficult” goal was set at the wrong level. In other words, defining different parameters each level (for example, establishing a higher threshold for what equals a “very difficult” goal - a 25 or 30 point gain instead of a 20 point gain), may offer more informative (and possibly significant) results.

Finally, as shown in Tables 4.4 and 4.5, five of the six blocks were significant at the $p < 0.05$ level and one, Block 2, was significant at the $p < 0.01$ level.
Table 4.4 - Hierarchical Regression Results Predicting School Performance – Blocks 1-3

<table>
<thead>
<tr>
<th>Variable</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>β</td>
</tr>
<tr>
<td>avgclass</td>
<td>-.184</td>
<td>.072</td>
<td>-.182*</td>
</tr>
<tr>
<td>schlsize</td>
<td>.001</td>
<td>.001</td>
<td>.049</td>
</tr>
<tr>
<td>tchrexp</td>
<td>-2.033</td>
<td>1.889</td>
<td>-.072</td>
</tr>
<tr>
<td>advdegree</td>
<td>-.430</td>
<td>2.266</td>
<td>-.013</td>
</tr>
<tr>
<td>nbpts</td>
<td>-1.652</td>
<td>1.974</td>
<td>-.056</td>
</tr>
<tr>
<td>%econdis</td>
<td>-1.416</td>
<td>1.279</td>
<td>-.091</td>
</tr>
<tr>
<td>%minority</td>
<td>-1.287</td>
<td>.920</td>
<td>-.107</td>
</tr>
<tr>
<td>goalstrattotal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>target</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>date</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rcvfdbk14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>easyread</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vdifread</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>easymath</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vdifmath</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R^2$          | .052    | .073    | .081   |

$R^2$ change   | .052*   | .021*   | .008   |

$F$           | 2.089*  | 2.605** | 2.323* |

N             | 273     | 273     | 273    |

*p < 0.05 **p < 0.01
Table 4.5 - Hierarchical Regression Results Predicting School Performance – Blocks 4-6

<table>
<thead>
<tr>
<th>Variable</th>
<th>Block 4</th>
<th>Block 5</th>
<th>Block 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>β</td>
</tr>
<tr>
<td>avgclass</td>
<td>-.192</td>
<td>.072</td>
<td>-.189*</td>
</tr>
<tr>
<td>schlsize</td>
<td>.001</td>
<td>.001</td>
<td>.081</td>
</tr>
<tr>
<td>tchrexp</td>
<td>-1.787</td>
<td>1.899</td>
<td>-.062</td>
</tr>
<tr>
<td>avgdgree</td>
<td>-1.801</td>
<td>2.006</td>
<td>-.024</td>
</tr>
<tr>
<td>nbpts</td>
<td>-1.760</td>
<td>1.284</td>
<td>-.113</td>
</tr>
<tr>
<td>% econdis</td>
<td>-1.088</td>
<td>.937</td>
<td>-.091</td>
</tr>
<tr>
<td>% minority</td>
<td>.030</td>
<td>.014</td>
<td>.132*</td>
</tr>
<tr>
<td>goalstrattotal</td>
<td>.001</td>
<td>.682</td>
<td>.000</td>
</tr>
<tr>
<td>target</td>
<td>-1.057</td>
<td>.753</td>
<td>-.094</td>
</tr>
<tr>
<td>feedback</td>
<td>-.215</td>
<td>.451</td>
<td>-.030</td>
</tr>
<tr>
<td>revfdbk14</td>
<td>1.751</td>
<td>1.689</td>
<td>.064</td>
</tr>
<tr>
<td>easyread</td>
<td>-.359</td>
<td>.575</td>
<td>-.055</td>
</tr>
<tr>
<td>vdiffread</td>
<td>.312</td>
<td>.579</td>
<td>.045</td>
</tr>
<tr>
<td>easymath</td>
<td>.348</td>
<td>.579</td>
<td>.054</td>
</tr>
<tr>
<td>vdfmath</td>
<td>-.765</td>
<td>.592</td>
<td>-.105</td>
</tr>
</tbody>
</table>

$R^2$     | .086     | .089     | .099     |
$R^2$ change | .005     | .003     | .009     |
$F$     | 2.044*   | 1.808*   | 1.751*   |
N     | 273      | 273      | 273      |

*p < 0.05   **p < 0.01

Hierarchical Regression Results for Change in School Performance

The second hierarchical regression model examined how well the same six blocks of predictors, identically composed and introduced, accounted for the change in school performance from the 2013-14 school year to the 2015-16 school year, inclusively. As with the Model 1 analysis, the results presented here reflect the 273-school sample. Identical regression analysis
was conducted for the larger sample with imputed data, and again, the results were similar. Only the variable related to feedback that helps improve teaching (recevfdbk14) experienced a change in the coefficient sign and with the same significant relationships were present. Results showed the larger model explained 16% of the variance in the dependent variable. Results are provided in two tables; Table 4.6 provides results for blocks 1 through 3 and Table 4.7 provides results for blocks 4-6.

The school demographic factors in Block 1 accounted for about 7.5% of the variance in the change in school performance for the three school years between 2013-14 and 2015-16. The variable representing the percentage of minority students in a school showed a significant positive relationship to school performance, indicating increases in school performance with increasing numbers of minority students in a school. This finding is not consistent with expectations of the model. None of the remaining variables in the block were significantly related to school performance, and once again several variables showed relationships that were not in accordance with expectations of the model. These included the percentage of economically disadvantaged students (positive; expected negative), the percentage of teachers with advanced degrees (negative; expected positive), and the percentage of teachers with 4 or more years of teaching experience (negative; expected positive). The topic of unexpected relationships, which have appeared for several variables in both models, is addressed more fully in the discussion of results later.

Block 2 added the variable measuring of the number of goals and strategies contained in a plan, and resulted in a small and non-significant change in explained variance – only about .4%. The goals measure added in the block was not significant. Again, the variable
for the percentage of minority students was positive and significant, with no change in effect size. No other variables in the block were significant.

Goal clarity measures added in Block 3 resulted in a significant 2.9% increase in variance explained. Only the clear target measure was significant. Both the clear target measure and the clear date measure showed positive relationships with the dependent variable, suggesting that increasing percentage of goals with clear targets and due dates in a school improvement plan leads to increased school performance, as hypothesized. Remaining variables in the model all were nonsignificant, including the previously-significant measure of minority students in a school. Its relationship to the dependent variable nonetheless remained positive.

Block 4, shown in Table 4.7, again introduced the two feedback-related variables. Unlike the first hierarchical regression, the variable measuring the presence of feedback loops in the school improvement plan was positively related to changes in performance as expected, but the second feedback variable was negatively related (also a change from Model 1), counter to the hypothesis that receiving feedback useful for improving instruction leads to higher performance. Unfortunately, neither variable showed a significant relationship to the change in school performance, but none of the other variables were significant either. Interestingly, however, the direction of the relationship for both the minority percentage and economically disadvantaged percentage remained positive and counter to the expected direction, just as they had in the first three blocks.

Block 5 provided another significant increase in the proportion of variance explained, adding another 3% to that explained by the first 4 blocks. As with Model 1, Block 5 introduced the “easy” reading goal and “very difficult” reading goal variables. “Stretch” goals again serve as the reference variable for these two dummy variables. The relationship between “easy” reading
goals and school performance was both significant and negative – as expected – indicating that “easy” reading goals produce lower performance than “stretch” goals. The “very difficult” reading goal was once again non-significant and positive, therefore, counter to expectations and implying that very difficult reading goals lead to higher performance than less difficult “stretch” goals. Additionally, the variable measuring the percentage of clear targets in an improvement plan, non-significant in the fourth model, re-gained significance in this fifth model. Its relationship again was positive as expected, and implied increases in school performance go along with increases in the percentage of plan goals that provide clear targets. Finally, the measure of economically disadvantaged students in a school changed signs in this model, indicating the expected negative relationship with school performance despite a weak and non-significant effect size.

Block 6 added the final two variables, and in doing so it added an additional 2.9% increase in the percent of variance explained, a significant increase over the variance explained by the prior model. This block added the dummy variables measuring mathematics goal difficulty. Again, stretch goals served as the reference for these variables. Only the “easy” math variable showed a significant relationship to changes in school performance and, unlike the results for model measuring single year school performance, this variable’s negative direction coincided with expectations, suggesting that easier goals lead to lower performance than stretch goals. The “very difficult” variable’s positive sign did not meet the expectations, however. The variable for clear targets retained its significance and its positive relationship in this final, fully specified model. The variable measuring economically disadvantaged students also maintained its negative direction, meeting expectations and suggesting that increases in the percentage of economically disadvantaged students in a school lead to lower school performance.
The unexpected directions for the “very difficult” reading and math goals may be further evidence to indicate that the threshold for defining an “easy” versus a “stretch” versus a “very difficult” needs further development.

Finally, as shown in Tables 4.6 and 4.7, all six blocks were significant at the $p < 0.01$ level. These tables are presented below.

**Table 4.6 - Hierarchical Regression Results Predicting Change in School Performance – Blocks 1-3**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B$</td>
<td>$SE$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>avgclass</td>
<td>-.011</td>
<td>.089</td>
<td>-.009</td>
</tr>
<tr>
<td>schlsize</td>
<td>-.000</td>
<td>.001</td>
<td>-.001</td>
</tr>
<tr>
<td>tchrexp</td>
<td>-1.039</td>
<td>2.330</td>
<td>-.029</td>
</tr>
<tr>
<td>advgree</td>
<td>-3.117</td>
<td>2.796</td>
<td>-.076</td>
</tr>
<tr>
<td>nbpts</td>
<td>3.940</td>
<td>2.435</td>
<td>.107</td>
</tr>
<tr>
<td>%econdis</td>
<td>1.969</td>
<td>1.578</td>
<td>.101</td>
</tr>
<tr>
<td>%minority</td>
<td>2.797</td>
<td>1.134</td>
<td>.187*</td>
</tr>
<tr>
<td>goalstrattotal</td>
<td>.018</td>
<td>.017</td>
<td>.062</td>
</tr>
<tr>
<td>target</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>date</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rcvfdbk14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>easyread</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vdiffread</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>easymath</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vdiffmath</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R^2$ | .073 | .077 | .106 |

$R^2$ change | .073** | .004 | .029* |

$F$ | 2.986** | 2.744** | 3.109** |

$N$ | 273 | 273 | 273 |

* $p < 0.05$ ** $p < 0.01$
Table 4.7 - Hierarchical Regression Results Predicting Change in School Performance – Blocks 4-6

<table>
<thead>
<tr>
<th>Variable</th>
<th>Block 4</th>
<th></th>
<th></th>
<th></th>
<th>Block 5</th>
<th></th>
<th></th>
<th></th>
<th>Block 6</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>β</td>
<td>B</td>
<td>SE</td>
<td>β</td>
<td>B</td>
<td>SE</td>
<td>β</td>
<td>B</td>
<td>SE</td>
</tr>
<tr>
<td>avgclass</td>
<td>.002</td>
<td>.088</td>
<td>.002</td>
<td>-.024</td>
<td>.088</td>
<td>-.019</td>
<td>-.081</td>
<td>.089</td>
<td>-.064</td>
<td></td>
<td></td>
</tr>
<tr>
<td>schlsize</td>
<td>.000</td>
<td>.001</td>
<td>.002</td>
<td>.001</td>
<td>.001</td>
<td>.033</td>
<td>.001</td>
<td>.001</td>
<td>.038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tchrexp</td>
<td>-.534</td>
<td>2.337</td>
<td>-.015</td>
<td>-1.267</td>
<td>2.322</td>
<td>-.036</td>
<td>-1.024</td>
<td>2.296</td>
<td>-.029</td>
<td></td>
<td></td>
</tr>
<tr>
<td>advgdgree</td>
<td>-3.064</td>
<td>2.781</td>
<td>-.075</td>
<td>-2.733</td>
<td>2.746</td>
<td>-.067</td>
<td>-2.652</td>
<td>2.712</td>
<td>-.065</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nbpts</td>
<td>2.693</td>
<td>2.469</td>
<td>.073</td>
<td>3.055</td>
<td>2.440</td>
<td>.083</td>
<td>3.133</td>
<td>2.424</td>
<td>.085</td>
<td></td>
<td></td>
</tr>
<tr>
<td>%econdis</td>
<td>2.158</td>
<td>1.580</td>
<td>.111</td>
<td>-.032</td>
<td>1.722</td>
<td>-.002</td>
<td>-1.348</td>
<td>1.773</td>
<td>-.069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>%minority</td>
<td>2.246</td>
<td>1.153</td>
<td>.150</td>
<td>.373</td>
<td>1.330</td>
<td>.025</td>
<td>.246</td>
<td>1.332</td>
<td>.016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>goalstrattotal</td>
<td>.028</td>
<td>.017</td>
<td>.100</td>
<td>.027</td>
<td>.017</td>
<td>.095</td>
<td>.030</td>
<td>.017</td>
<td>.106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>target</td>
<td>1.587</td>
<td>.840</td>
<td>.126</td>
<td>1.642</td>
<td>.829</td>
<td>.130*</td>
<td>1.885</td>
<td>.823</td>
<td>.150*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>date</td>
<td>.788</td>
<td>.927</td>
<td>.056</td>
<td>.417</td>
<td>.923</td>
<td>.030</td>
<td>.159</td>
<td>.915</td>
<td>.011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>feedback</td>
<td>.621</td>
<td>.555</td>
<td>.070</td>
<td>.605</td>
<td>.553</td>
<td>.069</td>
<td>.471</td>
<td>.555</td>
<td>.053</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rcvfdbk14</td>
<td>-.098</td>
<td>2.078</td>
<td>-.003</td>
<td>.308</td>
<td>2.056</td>
<td>.009</td>
<td>.079</td>
<td>2.041</td>
<td>.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>easyread</td>
<td>-1.739</td>
<td>.697</td>
<td>-.214*</td>
<td>-1.968</td>
<td>.739</td>
<td>-.119</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vdiifread</td>
<td>.736</td>
<td>.702</td>
<td>.084</td>
<td>.294</td>
<td>.724</td>
<td>.034</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>easymath</td>
<td>-.191</td>
<td>.693</td>
<td>-.237**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vdiifmath</td>
<td>.209</td>
<td>.709</td>
<td>.023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.110</td>
<td></td>
<td>.140</td>
<td></td>
<td>.169</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ change</td>
<td>.004</td>
<td>.030*</td>
<td>.029*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$</td>
<td>2.689**</td>
<td></td>
<td>3.012**</td>
<td></td>
<td>3.254**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>273</td>
<td></td>
<td>273</td>
<td></td>
<td>273</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $p < 0.05$  ** $p < 0.01$
Discussion of Results

Two different models of school performance were tested in the present study. Each used the same sets of predictor variables to assess their impact on school performance in a single school year as well as changes in performance over a three-year period. A brief discussion of each as well as findings gleaned from across the two models are presented here. Also presented here is a section addressing the implications of the results for the research questions and hypotheses associated with the analysis.

Summation of the Hierarchical Regression Predicting School Performance

The hierarchical regression analysis of Model 1 provided little in the way of significant beta coefficients either for independent or even control variables. This model analyzed school performance for a single school year, 2014-15, using 9 different variables developed to measure several different aspects of school improvement planning, including plan focus, goal clarity, goal difficulty, and feedback.

While the final overall model was significant ($F(16,272) = 1.751, p < 0.05$), it accounted for a very small proportion of the variance in school performance – just under 10%. A review of the analysis indicates several items of interest.

Average class size, represented with the control variable *avgclass* and introduced in Block 1 was a significant predictor of school performance in every block. There were no dramatic changes in effect size over the course of the analysis, though effect sizes did change from one block to the next. None of the other control variables were significant predictors of the dependent variable.
Three other control variables *tchrexp*, *advgdgree*, and *nbpts*¹⁷, consistently demonstrated relationships with school performance that ran counter to expectations. This occurred in every block in which they were included.

The variable measuring the count of goals and strategies in a plan, first introduced to the model in Block 2, was significant in that block and every block thereafter. While this suggests that the number of goals and strategies in a plan impacts performance, the direction of the relationship was positive, the opposite of the direction expected in Hypothesis 1, indicating that increases in the number of goals and strategies are associated with increases in school performance.

Goal clarity measures also suggested relationships that were opposite of those described in Hypothesis 2. Across the entire analysis and in the final model, none were significant. The coefficient for the variable *target*, which measured the percentage of plan goals that had clear targets, changed noticeably in magnitude with the introduction of the feedback variables, suggesting a possible relationship between the two. It changed again with the introduction of the goal difficulty variables. It makes sense that clear targets, feedback, and difficulty of goals would interact to affect performance, and a closer parsing out of the relationships between these variables would be a useful topic for future research.

Directions of the coefficients for the feedback goals, first introduced in Block 4, were mixed with respect to hypotheses about their relationships with school performance. One, *feedback*, which measured the presence of feedback loops in the plans, was negative across blocks 4 through 6, contrary to expectations. The second variable, measuring teacher perceptions of the value of feedback received, however, aligned with expectations of Hypothesis 4. These results

¹⁷ The variable *tchrexp* measures the percentage of teachers in a school who have 4 or more years of teaching experience; *advgdgree* measures the percentage of teachers with an advanced degree; *nbpts* measures the percentage of teachers in a school who are certified by the National Board of Professional Teaching Standards (NBPTS).
are important in that it indicates that actual feedback is more important than simply including feedback loops in a document.

Goal difficulty measures offered no conclusive picture of the relationship between difficulty and school performance. Each of the four “easy” and “very difficult” measures were expected to relate negatively to performance, relative to stretch goals. Only the two related to easy reading goals and very difficult math goals aligned with expectations, and none of the regression coefficients were significant for any of the four variables.

**Summation of the Hierarchical Regression Predicting Change in School Performance**

The hierarchical regression analysis of the change in school performance over three school years presented a number of interesting findings, including some that relate to specific components of the model tested and others related to specific variables included and their relationship to each other.

First, this analysis found that the explanatory variables included in this model accounted for a proportion of the variance in school performance that was larger than in Model 1 and yet was still only a modest proportion of overall variance. Although the final, fully specified model was significant overall ($F(16,272) = 3.254, p < 0.01$), this suggests that many of the factors impact how school improve over time are missing from the analysis. As pointed out in Chapter 1, and as any person who has visited a variety of schools can attest, schools reflect a complicated mixture of actors, activities, and circumstances. School performance may be affected by any number of internal and external influences. The small explained variance supports the notion that the tested model accounts for only a few of the possible influences and a few of the actions that can be taken to affect performance.
As mentioned in Chapter 3, school size is a demographic for which effects on school performance remain controversial. Represented in the present study by the control variable \textit{schlsize}, this variable was never an important aspect of this analysis. It was included in every block, was never significant, and its effect size remained at or near zero throughout. This at least suggests that school size is not particularly relevant to changes in school performance and that another variable may be more informative.

The direction of the relationship between the variable \textit{\%econdis}, which reflects the percentage of economically disadvantaged students in a school, and the dependent variable shifted in Block 5 from positive to negative, and it remained negative in the Block 6 analysis. This shift suggests a possible relationship between this economically disadvantaged students goals and reading goals that may deserve exploration in another study.

Three other control variables \textit{tchrexp, advdegree, and \%minority}\textsuperscript{18}, consistently demonstrated relationships with school performance that ran counter to expectations. This occurred in every block in which they were included.

Unlike in Model 1, the variable measuring the count of goals and strategies in a plan, introduced to the model in Block 2, was nonsignificant in that block and every block thereafter. While this suggests that the number of goals and strategies in a plan is not a strong predictor of performance, as with Model 1, the direction of the relationship was positive, the opposite of that anticipated in Hypothesis 1, again indicating that increases in the number of goals and strategies are associated with increases in school performance.

In three of the four models in which the variable \textit{target} was included, higher percentages of clear targets significantly predicted increases in school performance, in accordance with

\textsuperscript{18} Just as a reminder: the variable \textit{tchrexp} measures the percentage of teachers in a school who have 4 or more years of teaching experience; \textit{advdegree} measures the percentage of teachers with an advanced degree; \textit{\%minority} measures the percentage of minority students in a school.
expectations described in Hypothesis 2. The variable date, a second indicator of goal clarity, maintained a positive relationship with changes in school performance from the point of its introduction, though it was not significant in any of the blocks. Nonetheless, this offers some additional support for Hypothesis 2.

Two independent variables related to goal difficulty (\textit{vdifread} and \textit{vdifmath}, representing the “very difficult” goals) had relationship directions that were positive, each time they were included in a block, contrary to the expectations laid out in Hypothesis 3.

Finally, the analysis demonstrated positive relationships between each of the two feedback variables and changes in school performance, providing support for Hypothesis 4 and suggesting that including feedback loops as part of the planning process and receiving feedback focused on improving instruction each support increases in performance over time.

\textit{Implications for Hypotheses}

The current study focused on two primary components of the school improvement planning process – initial establishment of the school improvement plan and execution of the plan – and whether a relationship exists between what schools do as part of planning and execution and how their students perform.

With respect to initial establishment of the school improvement plan, the present analysis assessed whether a significant relationship exists between school performance, as indicated by a measure of school value-added, and key plan factors such as the number of goals and strategies, the clarity of goals and strategies, the use of stretch goals and the inclusion of planned feedback loops. Regarding school improvement plan execution, the study seeks to determine the impact of implementation approaches, especially feedback to teachers, on overall school performance. The
specific research questions and hypotheses were initially presented in Chapter 1, and are listed again below as part of a discussion of the implications of the results for each.

**Plan Focus – Implications for Hypothesis 1**

**Research question 1:** Do schools with a more focused school improvement plan, as evidenced by smaller numbers of plan goals and strategies, perform better?

**H1:** The number of goals and strategies in an improvement plan is inversely related to overall school performance.

Some of the previously cited research by Chun and Rainey (2005) and Jung (2011) returned mixed results for the assessment of the effects of the number of goals on performance. Results were similarly mixed in the present analysis. Only Model 1 resulted in a finding of significance for the variable measuring plan focus, which measured the number of goals and strategies in each school improvement plan in the sample. In Model 2, the variable was found to be nonsignificant. Hypothesis 1 anticipated an inverse relationship, however, in both models, the final direction of the relationship between the plan focus variable and school performance was positive, indicating that increases in the number of goals and strategies are associated with increases in school performance. This suggests that the number of goals and strategies in a plan, as operationalized by this analysis, is a predictor of questionable strength, and that Hypothesis 1 is not supported.

**Goal Clarity – Implications for Hypothesis 2**

**Research question 2:** Do schools with plans that offer greater goal clarity, shown by greater proportions of goals having clear targets and completion dates, perform better?

**H2:** A positive relationship exists between clear goals and school performance.

Hypothesis 2 anticipated a positive relationship between school performance and goal clarity measures that included clear targets and clear due dates for goal achievement. Model 1 results
indicated a negative relationship between both goal clarity variables and school performance for 2014-15, the opposite of expectations. However, Model 2 results indicated both variables had the expected positive relationship, and the variable measuring clear targets was significant at the \( p < 0.05 \) level. The same variable happened to be significant in the penultimate block, also at the \( p < 0.05 \) level. While this is not unequivocal, it does offer support for Hypothesis 2. Because the finding of significance coincided with the introduction of the goal difficulty variables for reading and math, future exploration of the relationship between clear targets and goal difficulty may be worthwhile.

**Goal Difficulty – Implications for Hypothesis 3**

**Research question 3:** Do stretch goals result in higher levels of school performance than either very easy goals or extremely difficult goals?

**H3:** Stretch goals support higher performance. Relative to stretch goals, very easy goals and very difficult goals are associated with lower performance.

Hypothesis 3 stated that relative to stretch goals, very easy goals and very difficult goals lead to lower performance, thus anticipating a negative relationship. In Model 1, no variables were found to be significant and the results were mixed. The “easy” reading and “very difficult” math variables both showed negative relationships as expected, however, the “very difficult” reading and “easy” math variables displayed positive signs.

In Model 2, the “easy” reading and “easy” math variables related negatively to school performance relative to stretch goals, and the “easy” math variable was significant, but both of the “very difficult” variables had nonsignificant, positive relationships to school performance relative to stretch goals, contrary to the expectations laid out in Hypothesis 3. As such, support for Hypothesis 3 is questionable.
Feedback – Implications for Hypothesis 4

Research question 4: Do the presence of feedback mechanisms and perceptions of feedback effectiveness impact school performance?

H4: A positive relationship exists between both the presence and perceptions of feedback mechanisms and school value-add.

Hypothesis 4 anticipates a positive relationship between both the presence of and perceptions of feedback mechanisms and school performance. In Model 1, directions of the coefficients for the feedback goals were mixed. The first, feedback, which measured the presence of feedback loops in the plans, had a negative coefficient, contrary to expectations. The second variable, rcvfdbk14, which measured teacher perceptions of the value of feedback received, had a positive coefficient, thus aligning with expectations of Hypothesis 4. Neither relationship was significant.

Model 2 demonstrated positive relationships between both feedback variables and changes in school performance, providing support for Hypothesis 4 and suggesting that including feedback loops as part of the planning process and receiving feedback focused on improving instruction each support increases in performance over time. As in Model 1, neither relationship was significant. Overall, the findings offered modest support for Hypothesis 4, with three of the four variables in question displaying the expected relationship between feedback and school performance, despite a lack of significant coefficients.

Summary of Support for Hypotheses

This hierarchical analysis was conducted to assess the effects of variables related to school improvement plan focus, goal clarity, goal difficulty, and feedback. Because of the limited achievement of significant coefficients as well as findings of differences between actual and predicted directions of the relationships, the analysis offers only modest support for two of
four hypotheses. The first hypothesis supported by the analysis suggests that increasing goal clarity in school improvement plans can help to improve school performance as indicated by school-level value-added measures. The second hypothesis supported suggests that feedback, when properly planned and executed, can contribute to improvements in school performance.

As mentioned, two additional hypotheses were generally unsupported by the analysis. The first suggested that a school improvement plan’s level of focus as measured by the number of goals and strategies included in the plan is negatively related to school performance such that increasing numbers of goals and strategies lead to decreasing school performance as measured with the school-level value-added indicators. One model found a significant, positive relationship between the number of goals and strategies and school performance. A second model found a non-significant, but also positive effect, thus not supporting the original hypothesis.

The second unsupported hypothesis suggested that goals that are very easy or very difficult relate negatively to school performance whereas “stretch” goals, which are difficult but not impossible to achieve, have a positive relationship to school performance. The analysis provided mixed results using four dummy variables as indicators of very easy and very hard goals and using a similar measure of stretch goals as a reference category. These mixed results offered no conclusive evidence of the relationship between goal difficulty and school performance.
Chapter 5 – Key Findings, Limitations, and Opportunities for Future Research

Introduction

This analysis used a quasi-experimental sampling design and data from multiple sources. The variables included in the hierarchical regression analysis offered an opportunity to gain insight into four research questions initially discussed in Chapter 1 and answered in Chapter 4. The analysis was conducted for two samples, one using imputed values and one excluding schools with missing plans. The results of each were similar and this study presents results and findings based upon the smaller sample without imputed school improvement plan data. The analysis assessed two aspects of the planning process – initial establishment of the school improvement plan and execution of the plan – and attempted to determine whether a relationship exists between what schools do as part of planning and execution and how their students perform.

Key Findings

The 2 models of school performance included in this analysis offered only modest ability to explain variance in and to predict school performance, yet both were significant overall (Model 1: \( F(16,272) = 1.751, p < 0.05 \); Model 2: \( F(16,272) = 3.254, p < 0.01 \)). The detailed analysis described in Chapter 4 revealed a number of key findings with respect to the predictor variables, and even a few of the control variables. Each of these key findings are addressed below.

Plan Focus Measure Yields Unexpected Relationship

The independent variable that measured plan focus yielded both a significant increase in variance explained \( (R^2) \) when it was introduced, as well as a significant beta-coefficient in the
final overall model (Block 6) for Model 1, thus indicating a significant relationship between the number of goals and strategies in a school improvement plan and the change in school performance in 2014-15. In Model 2, neither was true. However, the most interesting aspect of the findings for this variable is the fact that in both models the direction of the relationship was positive. It is unexpected, and therefore “key,” because it is counter-intuitive.

The literature related to the impact of the volume of goals on organizational performance, and frankly, the personal experience of the author, both suggest a negative relationship. However, some other analysts would anticipate the positive relationship that was actually found. For example, as discussed in the literature chapter, Boyne did not find that more goals led to less effectiveness. Other analysts (Blau, 1955) have concluded that the number of goals has a curvilinear relation to organizational performance, with too few leading to goal displacement (i.e., ineffectiveness caused by tunnel vision) and too many leading to lack of focus. Some analysts also have argued that a larger number of strategies will actually lead to greater agency effectiveness, but only when the agency’s workers are task immature. (Hersey and Blanchard, date) These analysts would anticipate this dissertation’s findings about goal numbers, particularly if many of the schools that had a large number of strategies also had fewer task mature teachers. However, if one equates “task immature” with “inexperienced,” one might also suspect the sample would have a relatively low mean value for the variable “tchexp,” which measures each school’s percentage of teachers with at least four years of experience. Within the sample analyzed here, the mean value for this variable is .77, meaning that on average among the schools in this sample, 77% of teachers have 4 or more years of experience. This perhaps suggests a need for further analysis of this aspect of school PM systems.
Other explanations for the findings are also possible. For example, in a school where high numbers of goals are present and where teachers and staff feel overwhelmed with expectations, perhaps they cope with those feelings by consciously or unconsciously accepting that they will never get it all done and therefore choosing to focus their time and effort on just a few things that are especially important to them or their students. If such a scenario is playing out, is it happening with individuals, or are groups of school personnel actively making such decisions together?

**Target Measure Confirms Expectations…Partly**

For Model 2, the variable measuring the presence of clear targets in school improvement plans was positively and significantly related to changes in school performance between 2013-14 and 2015-16, aligning with the expectations of Hypothesis 2 and offering support for the value of specific targets as a mechanism for providing goal clarity. The Model 2 findings also are in line with findings by Hu & Liden (2011), who also found significant, positive relationships between clear goals and performance. Moynihan, Pandey & Wright (2011) also confirmed the importance of clear goals in supporting the use of performance information. Model 1 (the single year performance measure) results indicated a negative and non-significant relationship, directly contravening the findings of Model 2 and these other analyses. This may suggest that clear targets are not particularly effective or helpful at improving school performance in the short run, but have greater impact over time, especially when the technologies for making improvements are unclear or especially difficult. It may be that once a clear target is established, time is required for the persons implementing strategies to “build up steam” for their efforts and begin to progress. It should be recognized that, as mentioned much earlier, the present analysis only reviewed one plan per school, and it assumed that each plan was representative of the school
improvement plans developed for its corresponding school over the three year time period considered by Model 2. Therefore, confirming the theory mentioned immediately above would require additional studies and analysis.

There are additional measurement difficulties connected to the timing of the plans used here. Some schools may have been planning effectively for ten years, while others may have begun to develop good plans just in 2014; yet both schools will appear the same because this dissertation takes only snapshot of plans. The school that has been effectively planning for ten years will likely have a better overall performance than the school that just recently began planning. But because the long-planning school reaped many of the benefits already, the school that just began may show more improvement over the three year period measured here. Therefore the length of time for which a school has had “good” planning may have substantial impact on the dependent variables, but it is an unmeasured difference.

**Goal Difficulty Results Are Inconclusive**

Hypothesis 3 suggested that very easy or very difficult goals relate negatively to school performance whereas “stretch” goals relate positively school performance. The results of the analysis presented here, conducted using four dummy variables as indicators of very easy and very hard goals and using a similar measure of stretch goals as a reference category, provided mixed results. This lack of conclusive evidence of the relationship between goal difficulty and school performance is not completely out of line with the literature. Boyne & Chen’s (2007) analysis of target difficulty also did not support their hypothesis; none of the “stretch,” or difficult-but-not-impossible, target models performed significantly better than their routine target-setting models. With similar findings, the goal difficulty measures in the present analysis offered no conclusive picture of the relationship between difficulty and school performance,
which is unfortunate as this issue remains especially relevant given current state level goal-setting practices for schools in North Carolina.

As mentioned previously, because Annual Measurable Objectives are set at the state level and, all schools have the same proficiency target, meaning some schools have easier-to-achieve goals than other schools. More conclusive findings of a strong positive relationship between stretch goals and performance would suggest clearly that AMO targets should not be established across the board, but instead should be customized for each school based upon their current proficiency levels. That would reflect a need to alter the current practice.

**Actual Feedback Can Make a Difference**

Hypothesis 4 suggested that useful feedback to teachers can help to improve student performance. The first model’s mixed results suggested that the presence of feedback loops is not so important yet aligned with expectations in the measurement of teacher perceptions of the value of feedback received. Model 2 demonstrated positive relationships between both feedback variables and changes in school performance, providing support for this hypothesis and suggesting that including feedback loops as part of the planning process and providing feedback focused on improving instruction each support increases in performance over time. As in Model 1, neither relationship was significant, but generally aligned with the findings by Fuchs & Fuchs (1986), who found that feedback to teachers significantly increased student performance. Hammond & Yeshanew (2007) also found positive relationships between feedback and school performance, in some cases finding significant relationships and in others finding the relationships positive but not significant, as the present analysis indicated.
Control Variables Can Be Interesting, Too

The control variables used in this study provided some unexpected and interesting results as well, especially for anyone who has worked extensively as a practitioner in K-12 education. These findings are not especially “key” with respect to the main purpose of the analysis, but they are worth considering briefly, and so they are considered below.

School size was first mentioned in the discussion of control variables in Chapter 3. It was noted that school size is a demographic for which effects on school performance remain controversial. To cite an example discussed previously, Bradley & Taylor (1998) found a positive, non-linear relationship between school size and school performance, whereas Foreman-Peck, et al. (2006) asserted that reducing school sizes, especially for certain age ranges of students, could result in significant gains in student performance.

In this analysis, school size was included as a control variable and because it is a basic demographic factor, was introduced in the first block with the other school demographics. For both Model 1 and Model 2, in every block introduced, school size produced non-significant relationships with unstandardized coefficients at or near zero. This suggests that school size is irrelevant to school performance, at least in the presence of the other variables included in the analysis.

Finally, three other control variables produced interesting findings, despite not being the main focus of the analysis. The variables for teacher experience (measuring the percentage of teachers with 4 or more years of experience), percentage of teachers with advanced degrees, and the percentage of National Board of Professional Teaching Standards-certified teachers were all anticipated to show a positive relationship with school performance, i.e., rises in the percentages of all of these kinds of teachers would result in increases in school performance. In Model 1, all
three, though not significant, nonetheless demonstrated relationships that were negative and therefore counter to expectations. In Model 2, only the Board-certified teacher variable had a positive relationship to school performance.

These findings are interesting because for practitioners in K-12 education, these particular topics arise periodically in the general conscience and sometimes in the policy arena and occasionally generate low-level controversy. For example, as mentioned in the discussion of the variables in Chapter 3, recent changes in teacher pay have resulted in some impacts to the teacher population such that experienced teachers are leaving North Carolina for better paying jobs elsewhere and in some cases, leaving the teaching profession entirely. Conventional wisdom is that a loss of experienced teachers is bad for both student and school performance, so it is interesting to see this contradictory, though unsubstantiated, result.

**Limitations and Potential Improvements**

**Threats to Internal Validity**

Internal validity is concerned with avoiding the introduction of unknown variables or factors into the analysis that could bias the results. There are two particular issues related to internal validity that must be addressed.

**Selection Bias as a Threat to Internal Validity**

The first relevant threat to internal validity is selection bias. As mentioned previously, this threat was mitigated through a process of randomly selecting schools. This process included discovery of the population size and definition of sample size on that basis, but more importantly the selection procedure used commonly available software with built-in functions that allow for a random selection of numbers from a list. Schools, which were individually numbered prior to the
selection of random numbers described immediately above, were then chosen by matching schools’ numbers to the list of randomly generated numbers. In preparation for the calibration process described above, schools included in the initial calibration analysis also were randomly selected from the list full sample.

Reliability as a Threat to Internal Validity

The next potential threat to internal validity that can be cited is reliability with respect to the data gathered from school improvement plans. Reliability is a necessary, but not sufficient, condition for ensuring validity (Garson 2002). The present study may have further strengthened reliability with an inter-rater reliability analysis, as originally planned. Instead, a rigorous calibration process was conducted with an initial analysis of more than 20% of the plans, randomly selected from the full sample. This analysis provided insight into how different plans within the sample were constructed and how they communicated goals, strategies, targets and other measures. The calibration process developed specific rules and a procedure for conducting school improvement plan-related counts. These rules and procedures guided the plan analysis so that the counts of goals, strategies and other plan components for each school were consistent and reliable over the long process of reviewing 273 school improvement plans. The rules that were the outcome of the calibration procedure were then applied to all of the plans, including those used to develop the rules.

While the calibration process could have been improved upon by a detailed inter-rater reliability analysis, that was not possible for this study, and the calibration rules that were developed and applied to the plan reviews were provided within the body of this analysis to provide transparency into the procedure used to conduct them.
**Threats to External Validity**

External validity is concerned with the process of generalizing conclusions from a sample to the population from which it was drawn, to other populations, or to other times and places. Garson (2002) states “The question raised is ‘To what population does the researcher wish to generalize his/her conclusions, and is there something unique about the study sample’s subjects, the place where they lived/worked, the setting in which they were involved, or the times of study, that would prevent valid generalization?’”

**Sample-related Factors Affecting External Validity**

To support generalizability, the researcher wanted to ensure that a representative sample was used to conduct the analysis. With respect to the population studied – elementary and middle schools in North Carolina – the sample size was selected so as to ensure a 95% confidence level at a 5% confidence interval. However, due to concerns about missing data, two separate samples were analyzed, one that included a larger sample of schools and imputed data for several important variables, and another that excluded the schools with missing data. The results of each analysis were similar, as pointed out in Chapter 4. However, the detailed results reported here reflect the smaller-sample analysis, slightly reducing confidence in the results (for a 95% confidence level, the confidence interval increased from 5% to about 5.5%, just above the generally accepted standard for social science research).

The sample was randomly selected from all schools in the population of North Carolina elementary and middle schools during the 2013-14 school year. The sample includes schools that range widely in size (the dataset included individual schools with as few as 48 students and as many as 1,409) and schools from urban, rural, and suburban communities across the state.
Of course, because the sample includes only elementary and middle schools, no inferences that may be made about the impact of school improvement can be extended to high schools in North Carolina or elsewhere. Regarding such potential inferences, several additional research design factors are considered below.

**Research Design Factors Affecting External Validity**

The external validity of the present study potentially is impacted by factors associated with the research design. First, the non-observational nature of the analysis implicitly and necessarily places much faith in the content of the school improvement plans as not simply a reflector of intentions, but also as evidence of actions. However, without the benefit of some level of first-hand observation or other insight into both the planning processes and implementation, there is no way a researcher can know how well a plan reflects what actually happened in a school and, by extension, the degree of impact on actual outcomes. Simply stated, a strong improvement plan built using good planning practices is likely to be ineffective if put on a shelf and not discussed again until the end of the school year. By the same token, a plan that looks weak on its surface but is implemented in an environment where improvement is valued and nurtured may nonetheless result in positive changes for the school community. There is a good deal of recent literature, spurred by Moynihan’s pioneering study of state management systems, that shows that many public agencies now know they “should” have performance management, and therefore they set up symbolic systems but they don’t actually use them for decision making.

The present study attempted to mitigate this limitation by incorporating a feedback variable that provided a measure of how well teachers believed that school leaders supported good instruction with feedback. Nonetheless, it was one measure among several and additional
school-level variables that offer insight into what actually happens in schools would be an improvement over the present design.

Additionally, the single-state design of the analysis also limits generalizability. The study is based upon North Carolina conditions whereas a multi-state analysis may offer better generalizability due to differences between states. For example, school improvement planning in North Carolina is a legal requirement for every school. This analysis offers no insight into whether the compulsory nature of the planning process impacts effectiveness. It may be that where such planning is voluntary, those who undertake it are more invested in its success and therefore exert greater effort during planning and implementation. Nonetheless, the present study may offer clues as to whether or not certain variables are worthy of further examination both here and in other states.

**Data Factors Affecting External Validity**

Data for the analysis comes from multiple existing sources of information about North Carolina schools, including the North Carolina Department of Public Instruction (NCDPI) as well as entities whose data collection is sponsored and paid for by the agency. Data accessed and downloaded for this study is the same data used by the agency for state and federal reporting. NCDPI and its partners have refined data collection processes over many years and routinely provide support and professional development to districts and schools that provide data. As a general rule, this should be considered some of the best available data for NC schools.

In spite of this, the usual data-related cautions apply. First, much of the data originates at the school or district level. Inconsistencies in collection are possible for some data due to local conditions and/or human error. For example, the present study identified missing data for some of the schools in the sample despite the well-established processes for collecting the data.
Along the same lines, the Teacher Working Conditions Survey, another significant data source, has been shown to be a valid, reliable instrument and has been in use for about fifteen years. Still, it cannot be considered a perfect data source due to the usual pitfalls. Aggregators of the data may err during the gathering and processing of data, and survey respondents may misunderstand questions, misremember facts, interpret events at the school level differently, or even embellish or lie in their responses.

**Other Limitations**

This study analyzed the relationship between school performance and a variety of variables suggested by the literature in the areas of performance management as well as school improvement planning. Few significant relationships were found and the final overall model, while significant, explained only about 17% of the variance in school performance. This low $R^2$ value suggests that the model is underspecified – missing one or more factors important to maximizing school performance.

For example, social science literature has long indicated that the success or failure of organizational changes, especially management changes, is heavily determined by the level of top support. When the top of an organization strongly and consistently supports a new system, and actually uses the system’s results to make important decisions, the system usually succeeds. If the top simply pays lip service to the system, the system typically fails. However, this dissertation was unable to measure top support, although the variable from the Teacher Working Conditions Survey related to feedback to teachers provided a partial and indirect measure.

Another factor unaccounted for in the present study that may show a demonstrable impact is a school’s level of financial resources. Inclusion of financial data was considered during the research design, but school-level expenditure data is hard to obtain. State funds are
allotted to school districts and districts have tremendous leeway in how they allocate funding to individual schools. Expenditure data readily available from NCDPI are reported by district and by program code, but not by school. In addition, many, but not all, NC school districts supplement state funding for teacher salaries, classroom resources, and other instructional needs with local supplemental funds provided by county commissioners and even private citizens. Schools also receive direct funding from donations that flow through parent-teacher organizations. Obtaining an accurate picture of school-level funding from this variety of sources, all either flowing through and being allocated by the district office or being provided directly to the schools had the potential to become a complicated exercise and was not pursued as part of this analysis.

The author also could not identify an alternative aggregator of school-level financial information, and collecting the data likely would have required an additional data request of the districts and schools represented in the sample.

Meier and O’Toole (2002) pointed out that class size is effectively a measure of a school’s monetary resources, and like many other analyses, this study included a class size measure. However, even if one assumes Meier and O’Toole are accurate in their assessment of the relationship between class size and money, a confounding factor for this analysis is the fact that NC state law limits maximum class size by grade-level ranges (Pre-K-2, 3-5, and so on). These class size limits are blind to school or district wealth, suggesting that class size may not be the best proxy for school fiscal resources in North Carolina.
Summary of Limitations

This study was impacted by several potentially important limitations, as clearly shown in the preceding paragraphs. However, also demonstrated is the fact that steps were taken when possible to mitigate these shortcomings and to ensure the analysis offers the best results possible.

Opportunities for Future Research

The findings of this analysis point to several interesting opportunities for future research on many of the relationships that were specified in these models. First, a further exploration of “focus” as a construct would be interesting and potentially valuable. Learning more about how best to operationalize it to assess widely-varied environments such as schools and classrooms could offer opportunities for researchers to serve up practical knowledge that principals and teachers can apply in their schools.

If future studies result in findings similar to those obtained through this study (counter-intuitive positive effects on performance), additional studies that offer insight into the mechanisms by which that occurs could be fascinating.

Second, as noted above, the results here might suggest that clear targets, while they may be a necessary but not sufficient condition for improvement, are not particularly effective or helpful at improving school performance over short time periods, but rather need more time to realize significant impacts. Again, perhaps that is especially true in widely diverse environments and especially when the technologies for making improvements are unclear or especially difficult. A comparative analysis, or even a series of analyses, might do well to identify schools with clear goals and targets and compare performance in these schools over multiple years with schools for which goals are not well-developed.
As stated in Chapter 4, the coefficient for the variable *target*, which measured the percentage of plan goals that had clear targets, changed noticeably in magnitude with the introduction of the feedback variables, suggesting a possible relationship between the two. It changed again with the introduction of the goal difficulty variables. It makes sense that clear targets, feedback, and difficulty of goals would interact to affect performance. While the benefit of feedback to achieving a goal is logical and well-established, an exploration of the interaction of feedback with goal difficulty in schools would be of particular interest. An easy-to-achieve goal may not need any feedback to be successful, whereas a stretch goal might require it, and very difficult to achieve goals may (or may not) be made more achievable with high quality feedback. One might measure the differences for every type of goal (easy, stretch, difficult) using an experimental design that provides varying degrees of feedback to schools, though a researcher would have to be careful about how to operationalize such a study so as to ensure it is conducted ethically.

Next, any further exploration focused on goal difficulty and how best to ensure “stretch” goals are set at the school level could prove to be valuable for state agencies, districts, and schools. As was noted earlier in this study, the idea of setting a single achievement or performance goal that applies equally to schools that are inherently unequal with respect to performance seems unwise in the face of decades of goal difficulty research confirming the value of “stretch” goals. A “one for all” approach is, by definition, ineffective for at least some schools. Gaining further insight into and developing techniques or approaches for determining what equals “stretch” may offer the education community some real added value. Many districts and schools – at least in North Carolina have become awash in data and analysis tools over the last several years, giving them a solid data foundation on which to base goals, if only they knew how to apply it to establishing stretch goals.
One final suggestion regarding the possibilities for future research relates to the control variables and is attributed to the author having spent ten years working in a state education agency. Over those years, numerous discussions and debates occurred between state agency, school district, and school-level personnel, as well as state legislators, regarding the value and importance to school performance of incentives for advanced degrees, National Board-certification, and significant experience in the classroom. Any additional analysis that offers a close look at the relationships between school performance and teacher experience, education, and/or certification could offer relevant, practical insight to all those involved in the debate and would be interesting and valuable. Practical knowledge that can inform decision making about the use of resources to support – or not support – these kinds of programs is needed.
References


http://etheses.dur.ac.uk/1059/.


Washington, DC.


