AN ABSTRACT OF THE DISSERTATION OF

Tara Fortner for the degree of Doctor of Education in Learning, Leadership, and Community presented on May 10, 2017.

Title: Assessing the Effectiveness of New Hampshire Elementary Schools: An Effective Schools Approach

Abstract approved:

Gary Goodnough, Ph.D.

Despite consistently strong performances among NH 4th graders on the NAEP assessments, large disparities have been observed among NH elementary students on the NECAP assessments based on race and SES. The current study assessed the effectiveness of NH elementary schools, as defined by the effective schools research. Of the 209 elementary schools included in the current study, 8.6% were effective for reading, while 9% were effective for math. Multinomial logistic regression analyses revealed that student body composition, school funding practices, and teacher qualifications could each differentiate effective schools from other schools. The percentage of low-income students in a school had the strongest explanatory power on the effectiveness of schools accounting for 84% of the variability in the effectiveness of schools for reading and 86% of the variability in the effectiveness of schools for math. Student body composition was found to be highly associated with school funding practices and teacher qualifications.

Assessing the Effectiveness of New Hampshire Elementary Schools: An Effective Schools Approach

By

Tara Fortner

A DISSERTATION

Submitted to

Plymouth State University

In partial fulfillment of

the requirements for the

degree of

Doctor of Education

Defended May 10, 2017

Degree Conferred May 2017

Dissertation of Tara Fortner

Presented on May 10, 2017

APPROVED:

Gary Goodnough, Ph.D., Dissertation Committee Chair

Scott Mantie, Ph.D., Dissertation Committee

Christie Sweeney, Ph.D., Dissertation Committee

Gail Mears, PsyD., Dean of the College of Education, Health, and Human Services

I understand that my dissertation will become part of the permanent collection of Plymouth State University, Lamson Learning Commons. My signature below authorizes release of my dissertation to any reader upon request.

Tara Fortner, Author

ACKNOWLEDGEMENTS

I want to express gratitude to the members of my dissertation committee, Dr. Gary Goodnough, Dr. Scott Mantie, and Dr. Christie Sweeney, for their patience and support. Likewise, I want to acknowledge my family, most notably, Elizabeth Rose, for their patience and encouragement throughout this seemingly unending process. A special acknowledgement to Maureen Kennedy for supporting not only myself, but also my family the past few years while I completed my doctoral coursework and dissertation. Most importantly, I thank God for providing me with the resilience and perseverance to overcome the odds.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Today schools are looked upon to produce a workforce that possesses high level academic skills, critical thinking skills, and problem-solving skills requisite for the United States (US) to compete with other higher achieving countries (Benitez, Davidson, Flaxman, 2009; Darling-Hammond, 2004, 2010, 2012; Hall & Kennedy, 2006; Trilling & Fadel, 2009; Wagner, 2008). According to several researchers (Darling-Hammond, 2010, 2012; Education Trust, 2006b; Education Trust, 2008b; Education Trust, 2011; Lezotte, 2009; Trilling & Fadel, 2009; Wagner, 2008), US students are not adequately prepared to enable the US to compete in the global market place. The US is experiencing "four distinct achievement gaps - (1) between the United States and other nations; (2) between Black and Latino students and White students (3) between students of different income levels; and (4) between similar students schooled in different systems or regions" (McKinsey & Company, 2009, p. 5). These achievement gaps are having devastating economic impacts: lowering the US's 2008 gross domestic product by at least $2.435 trillion (McKinsey & Company, 2009). If significant reform does not occur, the US will not be able to compete in the global market (Darling-Hammond, 2010, 2012; Edelman & Jones, 2004; Education Trust, 2006b; Education Trust, 2008b; Education Trust, 2011; Lezotte, 2009; Trilling & Fadel, 2009; Wagner, 2008).

According to the studies by the Organization for Economic Cooperation and Development ([OECD], 2010, 2013, 2016) and other researchers (McKinsey & Company, 2009), the US's educational system is falling behind many other countries especially in math. On the 2012 and 2015 Program for International Student Assessment (PISA) reading assessment, the US performed in the middle of the OECD countries

(OECD, 2013, 2016) earning a rank of 17 out of 34 in 2012 (OECD, 2013). The US's performance on the 2012 and 2015 math PISA fell below most of the OECD countries (OECD, 2013, 2016) earning a rank of 27 out of the 34 OECD countries in 2012 (OECD, 2013). These findings are supported by a 2008 study (Education Trust) which found that more than 40 percent of US high school seniors scored in the lowest performance category, *below basic*, on the 2005 National Assessment of Educational Progress (NAEP) math assessment. McKinsey and Company (2009) analyzed the impact that this achievement gap had on the US's 2008 gross domestic product (GDP). Based on their investigation they concluded that, "If the United States had in recent years closed the gap between its educational achievement levels and those of better-performing nations such as Finland and Korea, [the US] GDP in 2008 could have been $1.3 trillion to $2.3 trillion higher" (p. 5).

Within the US there is a gap between the achievement of Black and Latino students and the achievement of White students (Benitez et al., 2009; Darling-Hammond, 2010; Domina, 2006; Edelman & Jones, 2004; Education Trust, 2008a; Education Trust, 2011; Leach & Williams, 2007; Lezotte, 2009; McKinsey & Company, 2009; Pitre, 2014; Sharma, Joyner, & Osment, 2014) which widens the longer students are in school (McKinsey & Company, 2009). McKinsey and Company (2009) examined the National Assessment of Educational Progress (NAEP) results and found that 48 percent of Black students and 43 percent of Latino students performed in the lowest achievement level, below basic, whereas only 17 percent of White students performed at that level. Evidence of this achievement gap was found in every state (McKinsey & Company, 2009). Several other writers (Darling-Hammond, 2010; Kozol, 2005; Ladson-Billings, 2006; Zhang & Cowen, 2009) have shown that the academic outcomes attained by suburban middle-class schools, comprised mostly of White students, exceeds the academic outcomes attained by urban schools that are mostly comprised of non-White students. In their 2014 study, Sharma et al. found that as the percentage of Black students in a school increases, performance on state assessments of math decline. The effect of this achievement gap was estimated as being at least $310 billion on the US's 2008 GDP (McKinsey & Company, 2009).

The third achievement gap affecting the nation is the gap between students of differing socio-economic levels (McKinsey & Company, 2009). Zhang and Cowen (2009) found that schools which did not demonstrate adequate yearly progress had larger concentrations of students from low-income households than did the better performing schools in the same school district. As with the racial achievement gap, many writers (Benitez et al., 2009: Darling-Hammond, 2010; Domina, 2006; Education Trust, 2008a, Education Trust, 2011; Kozol, 2005; Lezotte, 2009; McKinsey & Company, 2009; Sharma et al., 2014; Singh, 2015) assert that the US is not providing an appropriate education for its low-income students. Relative to low-income students in other nations, low-income students in the US are more disadvantaged (McKinsey & Company, 2009). The OECD (2010) found that "14% of the differences in student reading performance within each country are associated with differences in students' socio-economic background" (p. 14). Singh (2015) found that students' performances on the third grade Hawaiian State Assessments of math was significantly influenced by family SES. Per the

OECD (2010), the impact of poverty on academic achievement has been ameliorated in several countries wherein approximately 40-50% of students from disadvantaged socio-economic backgrounds "performed in the top quarter among students from all countries with similar socio-economic backgrounds" (p. 13). One crucial factor that was found suggests that students who come from low-income homes, yet attend schools with average socio-economic backgrounds, perform better than low-income students who attend schools wherein most students also come from low-income homes (OECD, 2010). Sharma et al. (2014) found that "as the percentage of poor students in a school increases, individual students' performance on standardized tests decrease" (p. 12). Based on their analyses, McKinsey and Company (2009) estimate that this gap deflated America's 2008 GDP by at least $400 billion.

Finally, McKinsey and Company (2009) explored regional and system-based achievement gaps. These researchers found large disparities between states, between districts within states, between schools within districts, and between classrooms within schools (p. 14). McKinsey and Company (2009) found that "there is actually more variation in student achievement *within* schools than *between* schools in the United States" (p. 14). These gaps are estimated to have lessened the US's 2008 GDP by at least $425 billion (McKinsey & Company, 2009).

The US uses the National Assessment of Educational Progress (NAEP) to assess the achievement of US students in mathematics, reading, science, writing, and other curricular areas. Each year a random selection of students across the nation, in grades 4, 8, and 12 are administered the NAEP assessments (National Center for Educational

Statistics, n.d.). These results are then analyzed to provide the country with a report on the status of education in the US. Data is available through the NCES website which allows for comparisons to be made between states. Use of this data system reveals that the mean performance of NH's fourth and eighth grade students on the reading and math assessments has exceeded the average for the nation since at least 2003. In fact, on the 2011 NAEP assessments, no state outperformed NH, and only one state (Massachusetts) performed comparably to NH's fourth grade students on the math assessment. In reading, Massachusetts' fourth grade students out- performed NH, while students in Maryland, New Jersey, and Connecticut performed comparably to NH students. Eighth grade students from two states outperformed NH eighth graders in math while three outperformed NH in reading. Nevertheless, the results of the NAEP assessments suggest that NH students, as a whole, performed well.

New Hampshire used the New England Common Assessment Program (NECAP) to assess students' reading and math abilities through the end of the 2013-2014 academic year. The NECAP reading and math assessments were taken by students in grades 3-8 and 11 in the fall of each year. In addition, students in grades five, eight, and 11 also took the writing assessments (Measured Progress, 2010, 2011, 2012).

Data obtained from the Fall 2012 NECAP assessments taken by the third, fourth, and fifth grade students in the state of NH revealed marked discrepancies between racial groups (NH DoE, 2013a, 2013b, 2013c). Data reported in Tables 1.1, and 1.2 were derived from a variety of NECAP reports obtained from the NH Department of Education (DoE) website (2013a, 2013b, 2013c). Table 1.1 shows that approximately 78% of third,

Table 1.1: Fall 2012 Reading Results % NH Students Scoring Proficient or Higher

Race/Ethnicity	3rd Graders *n*	3rd Graders %	4th Graders *n*	4th Graders %	5th Graders *n*	5th Graders %
Hispanic or Latino	675	62	630	60	616	57
Not Hispanic or Latino						
American Indian or Alaskan Native	52	61	40	68	41	76
Asian	435	82	446	79	397	80
Black or African American	236	64	252	63	272	61
Native Hawaiian or Pacific Islander	15	87	9	*	13	77
White	12,003	79	12,069	80	12,422	79
Two or more races	271	72	251	75	285	76
No Race/Ethnicity Reported	0	---[a]	0	---	1	---
Total	13,687	78	13,697	78	14,047	77

Note. n = *number of students tested; % = percentage of students tested who demonstrated proficiency*
[a] Due to limited sample size statistics for these subgroups could not be computed.

Table 1.2: Fall 2012 Math Results % NH Students Scoring Proficient or Higher

Race/Ethnicity	3rd Graders *n*	3rd Graders %	4th Graders *n*	4th Graders %	5th Graders *n*	5th Graders %
Hispanic or Latino	685	54	638	58	628	54
Not Hispanic or Latino						
American Indian or Alaskan Native	52	57	40	66	41	73
American Indian or Alaskan Native	52	57	40	66	41	73
Asian	446	81	462	78	409	81
Black or African American	241	50	253	52	274	50
Native Hawaiian or Pacific Islander	15	80	9	*	13	62
White	12,003	76	12,072	79	12,422	75
Two or more races	270	65	252	70	285	71
No Race/Ethnicity Reported	0	---[a]	0	---[a]	1	---[a]
Total	13,687	78	13,697	78	14,047	77

Note. n = *number of students tested; % = percentage of students tested who demonstrated proficiency*
[a] Due to limited sample size statistics for these subgroups could not be computed.

fourth, and fifth grade students demonstrated at least proficient abilities on the Fall 2012 reading assessments while Table 1.2 shows that approximately 75% of these students demonstrated at least proficient abilities on the Fall 2012 mathematics assessments. Upon closer analysis, though, the results reveal substantial differences between racial groups with a much larger percentage of White and Asian students demonstrating proficiency

Table 1.3: Fall 2012 Reading Results % NH Students Scoring Proficient or Better

	3rd Graders		4th Graders		5th Graders	
SES	*n*	%	*n*	%	*n*	%
Low-Income Students	3,797	63	3,743	62	3,873	63
All Other Students	9,890	84	9,954	84	10,174	83
Total	13687	78	13,697	78	14,047	77

Note. n = *number of students tested; % = percentage of students tested who demonstrated proficiency*

Table 1.4: Fall 2012 Math Results % NH Students Scoring Proficient or Higher

	3rd Graders		4th Graders		5th Graders	
SES	*n*	%	*n*	%	*n*	%
Low Income Students	3,795	58	3,741	61	3,873	57
All Other Students	9,917	84	9,985	83	10,200	81
Total	13,712	74	13,726	77	14,073	74

Note. n = *number of students tested; % = percentage of students tested who demonstrated proficiency*

than Black and Latino students. Native Hawaiian or Pacific Islander third graders also performed well on the reading and math assessments.

Analysis of the Fall 2012 NECAP results also revealed discrepancies based on socioeconomic status (NH DoE, 2013a, 2013b, 2013c). Data reported in Tables 1.3, and 1.4 were derived from reports available through the NH DoE website (2013a, 2013b, 2013c). Table 1.3 shows that while only approximately 63 percent of economically disadvantaged students demonstrated proficient reading abilities, approximately 84 percent of all other students demonstrated proficiency. Likewise, Table 1.4 shows the disparities between these two groups on the Math NECAP assessment.

For the US to compete in the global market place, US schools need to be effective for all students (Reagle, 2006; Trilling & Fadel, 2009; Wagner, 2008) regardless of race, ethnicity, and socioeconomic status (Benitez et al., 2009; Darling-Hammond, 2010). This sentiment is supported by the OECD's (2010) analysis of the 2009 PISA results in which they note that “the best performing school systems manage to provide high-quality education to all students” (p. 13). Through the development of more effective schools, the US could achieve better outcomes for its money (McKinsey & Company, 2009; Montt, 2011). More effective schools could produce more students with skills requisite for the US to compete in the global market (Darling-Hammond, 2010) thereby reducing the “underutilization of human potential in the United States [which has been] extremely costly” (McKinsey & Company, 2009, p. 5).

Purpose

The purpose of the current study was to provide information about the effectiveness of NH Elementary Schools in educating students in the areas of reading and math, and the extent to which student body composition, per pupil expenditures, and teacher qualifications have on the effectiveness of NH elementary schools. Given its focus on effectiveness for all students, regardless of race or socioeconomic status (SES), an effective schools theoretical orientation was utilized. Four questions were proposed to facilitate this research:

1. What percentage of NH elementary schools are effective in reading?
2. What percentage of NH elementary schools are effective in math?
3. To what extent do student body composition, school funding practices, and teacher qualifications explain a school’s effectiveness in reading?
4. To what extent do student body composition, school funding practices, and teacher qualifications explain a school’s effectiveness in math?

Conceptual Definitions

For the purposes of the current study the following conceptual definitions were employed:

- Race – even though the terms race and ethnicity are neither synonymous nor mutually exclusive, for the purposes of simplicity, race will be used in this study. The term race will be used to signify either race or ethnicity.
- Effective schools – according to the effective schools research, an effective school is one that is effective for all students regardless of race or SES.
- Student body composition – characteristics of the student body such as the proportion of low-income students and the proportion of non-White students.
- Non-White Students – students identified by the NH DoE as being "American Indian/Alaskan Native," "Black (not Hispanic)," or "Multiple Races."
- Low-income students – are students identified by NH DoE as being either low-income students or students who are eligible for free and reduced lunch. While there is overlap amongst these two groups of students, there are also differences. Nevertheless, for the sake of simplicity the term low-income students will be used to refer to both groups of these students.

Operational definitions are provided in the third chapter.

Limitations

The current study is designed to be a preliminary study and therefore has many limitations. It is hoped that many additional studies will be conducted in NH as well as across the US to further contribute to this line of research.

First and foremost, the current study is limited by its assessment of effectiveness based solely on the performance of students on high stakes reading and math assessments. Not only are other core curricular areas such as science, history, and writing being omitted, but there are other important developmental factors that are not being considered. For instance, a team of researchers through the University of Chicago Consortium on Chicago School Research (Farrington et al., 2012) conducted a comprehensive review of literature that revealed the substantive influence that noncognitive factors have on one's success in life. According to these researchers (Farrington et al., 2012), noncognitive factors include learning strategies (i.e., metacognitive strategies), student attitudes (i.e., mindsets), and student behaviors (i.e., perseverance). Moreover, their results suggest that educators can assist students in developing these noncognitive factors "through attention to academic mindsets and development of students' metacognitive and self-regulatory skills" (Farrington et al., 2012, p. 7). As such, the American School Counselor Association relied heavily on this research in their 2014 revisions of their standards wherein 35 non-cognitive factors that students need to possess to be successful in life were identified. Their standards include six positive mindsets, and 29 behaviors: 10 learning strategies, 10 self-management skills, and nine social skills. Other researchers (Jennings et al., 2015) note that other outcomes such as college attendance may be more important indicators of success than are standardized test scores. Farrington et al. (2012) noted that students' grade point averages in middle school and high school were better predictors of students' success in life than were group administered achievement tests.

As is consistent with the effective schools research, the current study does not consider the performance of students with disabilities, limited English proficiency, or middle school and high school students. Moreover, the current study will not utilize a sample that is representative of all US schools. Clearly the achievement of all students is important, but is outside the scope of the current research. It is hoped that additional research will be conducted in the state of NH and throughout this nation that focuses much needed attention on effective educational methods for these other groups of students.

Many readers of this research will perceive the decision to not control for SES as a limitation. Although SES, family background, and community characteristics are important variables in the education of students, their affects can be ameliorated (OECD, 2010). A truly effective school will be effective for all students regardless of their SES, race, ethnicity, family background, and community characteristics (Clewell & Campbell, 2007; Edmonds, 1979, 1982; Lezotte, 1991, 2005, 2009; Lezotte & Snyder, 2011; Rowan, Bossert, & Dwyer, 1983). Consequently, in the current study SES was not controlled for, but was used as a pivotal factor when identifying effective schools. Schools that show significant differences between subgroups of students, based on SES, were not considered effective.

Finally, the current study is limited by its conceptualization of an effective school as being one that is effective for at least 80 percent of its students: Meaning that 80 percent of its students perform proficient or better on both reading and math NECAP assessments. Ideally, an effective school will have a far higher percentage of its students achieving the minimal, benchmarks or higher.

Chapter 2: The Review of Literature

The purpose of the current study was to assess the effectiveness of NH elementary schools and to determine which variables best differentiate effective schools from those that are not effective. For the purposes of the current study, an effective school is defined as one in which at least 80 percent of all students demonstrated at least proficient abilities on the Fall 2012 NECAP assessments, did not have significant subgroup differences based on race or SES, and performed in the top third of NH Elementary Schools on the Fall of 2010 and 2011. Through this literature review, three variables of interest emerged and were used as explanatory variables: student body composition, per pupil expenditures, and teacher qualifications. The variable student body composition was broken down into two separate variables: race, as measured by the percentage of White students in a school, and percentage of low-income student, as measured by the percentage of students eligible for free and reduced lunch. It is hoped that this research will not only provide NH educators with information about the influence of these three variables on the effectiveness of NH schools, but also to set the stage for more in depth analyses of effective NH schools. Moreover, it is hoped that the current study will help to stimulate a resurgence of the effective schools research given its emphasis on effective schools being effective for all students.

Through this literature review readers will be provided with an understanding of the methods, findings, criticisms, and beliefs that drive the effective schools research. A chronological approach was used to present the history of the effective schools research as it pertains to elementary schools including information pertaining to the importance of two key variables: student body composition and teacher qualifications. Next, a myriad of criticisms of the effective schools research will be presented so that readers can better understand the rationale for the methodology proposed in the current study. Then, readers will be presented with information relevant to the status of equity of educational opportunities in the state of NH. This portion of the literature review will not only reiterate the importance of teacher qualifications, but also discuss the effects of per pupil expenditures on the education of non-White students and low-income students. In order to provide readers with an understanding of the steps taken in the US to close the achievement gaps, a discussion of the No Child Left Behind Act of 2001 (NCLB) will be included. Finally, two recent studies that relate to the effective schools research will be summarized.

The Origins of the Effective Schools Research

The effective schools research is a body of research that sets itself apart from other types of educational research through its adherence to the belief that an effective school is one which is effective for all students regardless of race, ethnicity, or socioeconomic status (SES). The effective schools research focuses on high achievement for all students with no gaps between subgroups (Lezotte & Snyder, 2011, p. 19). Through an examination of available resources, resource allocation, systemic processes, and curricular approaches, effective schools' researchers strive to identify the characteristics which set unusually effective schools apart from schools that are less effective.

Just as many movements in the US have been triggered by a desire for social justice, so too was the effective schools research. As an offshoot to President Johnson's War on Poverty, the Civil Rights Act of 1964, and the Elementary and Secondary Education Act (ESEA) of 1965, Harold Howe II, the US commissioner on education, was ordered by Congress to investigate the "lack of availability of equal educational opportunities for individuals by reason of race, color, religion, or national origin in public educational institutions" (as cited in Howe, 1966, p. iii).

Four major questions were the focus of the research conducted between 1964 and 1965:

1. How racially and ethnically segregated are the US public schools?
2. Do schools offer equitable educational opportunities?
3. How much do students learn?
4. What is the relationship between students' achievement and the schools they attend?

The findings of this investigation were published in 1966 under the title *Equality of Educational Opportunity* (Coleman et al.); this commonly referenced report is best known as "The Coleman Report" in deference to James Coleman, the lead investigator. The Coleman Report is important to the current research not only because it triggered the effective schools research, but also because it highlighted important key variables in the education of non-White students and low-income students: student body composition and teacher qualifications.

When SES was controlled for, schools appeared to be similar; however, as Coleman et al. (1966) noted, SES is highly related to academic achievement. Other researchers took this finding to mean that schools do not make a difference (Clewell & Campbell, 2007; Jansen, 1995; Lezotte & Snyder, 2011; Lezotte 2005, 2009; Marzano 2000, 2003; Wilson & Fergus, 1988). Further analysis of The Coleman Report, as summarized below, reveals otherwise.

An important finding noted in The Coleman Report (1966), but rarely mentioned in secondary sources was the discovery that ethnic minority students, other than Asian Americans, are more influenced by the quality of education they are provided than are their White peers. Consequently, Coleman et al. (1966) concluded that, "it is for the most disadvantaged children that improvements in school quality will make the most difference in achievement" (p. 22). Through their research, Coleman et al. (1966) identified two variables that were most influential for ethnic minority students: teacher quality and the composition of the student body.

As part of their analyses, Coleman et al. (1966) examined the relationship between teacher quality and student achievement. The results suggested a strong relationship between teachers' verbal ability and student achievement. Furthermore, a strong correlation was observed between student achievement and teachers' educational background, including the educational level of the teachers' parents. Coleman et al. (1966) observed that more than half of the teachers in schools came from the same city, town, or county in which they worked. As such schools that were more effective, as reflected by the quality of their graduates, were more likely to have better educated teachers, while schools that were minimally effective, as reflected by the quality of their graduates, were more likely to have less effective teachers. This pattern contributed to the perpetuation of a cycle of educational inequality.

The second major finding from their study was the strong influence that student body composition has on student achievement (Coleman et al., 1966; see also Rivkin, 2016). Coleman et al. (1966) found that the "educational backgrounds and aspirations of the other students in the school" (p. 22) are more influential than family background. Related findings (Coleman et al., 1966) suggest that Black students' locus of control is strongly influenced by the proportion of White students in their school. Moreover, students' locus of control is strongly related to their achievement: the more students believe that they have control over their lives the higher their achievement (Coleman et al., 1966). Considering that Coleman et al. (1966) also noted that US schools are overwhelmingly segregated by race, the fact that peer factors are so influential is especially noteworthy. Nevertheless, researchers rarely discuss these findings (Darling-Hammond, 2010).

Coleman and his team (1966) found that significant differences in educational opportunities are apparent regarding teacher to student ratio, and extracurricular activities. The Coleman Report asserted that White students "have more access to a more fully developed program of extracurricular activities, in particular those which might be related to academic matters" (p. 12).

Overall, the Coleman Report (1966) found that ethnic minority students have weaker verbal and nonverbal skills when they enter school due to a combination of "nonschool factors" (p. 21). Then these students are educated in segregated schools and provided with inferior educational opportunities thereby causing them to fall further and further behind their White peers.

Numerous researchers (Clewell & Campbell, 2007; Lezotte, 1991, 2005, 2009; Marzano, 2000, 2003; Witte & Walsh, 1990) note that the results of the Coleman Report (1966) have often been misinterpreted. The most prevalent misinterpretation is that the disparities in educational outcomes for White students and Black students are the result of differing life circumstances such as home, peer, and community factors. Therefore, many conclude that schools have little impact on ameliorating the disparate outcomes. Darling-Hammond (2010) noted that, "Although the [Coleman] report pointed to many inequalities that it argues should be remedied, the statement became widely viewed as a claim that school funding does not affect school achievement" (p. 101).

A few years after the release of The Coleman Report, Jencks et al. (1972) published the results of their study in which they concluded that schools had little impact on ameliorating the effect of students' life circumstances (Clewell & Campbell, 2007; Lezotte, 1991, 2005, 2009; Lezotte & Snyder, 2011; Marzano, 2000). Marzano (2000) notes that Coleman et al. (1966) and Jencks et al. (1972) used the same data in their analyses. Given that verbal ability, which was used by Coleman et al. (1966) and Jencks et al. (1972) as a dependent variable, is highly related to students' background, it is not surprising that limited differences were observed between schools when SES was controlled for. Consequently, Marzano (2000) argues that the results of these two studies were not valid. Konstantopoulos and Borman (2011), through their use of contemporary statistical approaches reexamined the original Coleman et al. (1966) data. Their findings supported Marzano's (2000) claim that the original studies were not valid. The results of Konstantopoulos' and Borman's (2011) multi-level regression analysis revealed that school characteristics "explained, on average, nearly one half of the between-school variance in the minority-White achievement gap and a little over 40% of the between-school variance in the SES effect" (p. 125); Based on their findings, Konstantopoulos and Borman (2011) assert that "schools play meaningful roles in distributing equality, or inequality, of educational opportunities" (p. 125).

The Effective Schools Movement

Numerous independent researchers (Brookover et al., 1978; Edmonds, 1979, 1982; Klitgaard & Hall, 1975; Purkey & Smith, 1982, 1983; Rowan et al., 1983), who later joined forces, set out to provide evidence that all children, including non-White and low-income students in urban settings, can learn, and that schools can make a difference (Clewell & Campbell, 2007; Lezotte, 1991, 2005, 2009; Marzano, 2000). This line of research became known as the effective schools research (Lezotte, 2005; Lezotte & Snyder, 2011) or the effective schools movement (Lezotte, 1991; Lezotte & Snyder, 2011; Marzano, 2000).

Initially the focus of the effective schools movement was on identifying urban elementary schools that served high proportions of non-White students who demonstrated higher than expected levels of achievement in comparison to other demographically similar schools (Jansen, 1995; Lezotte 1991, 2005, 2009; Marzano, 2000). This was accomplished by finding schools that were similar in demographics to other schools, but produced more promising outcomes for non-White and low-income children in urban settings. These *positive outlier* schools were examined through case studies and/or contrasted groups (a.k.a. outlier studies) approaches (Clewell & Campbell, 2007; Marzano, 2000). The purpose of these studies was to identify differences between these schools, thereby enabling researchers to better understand what it was about these schools that made them unusually effective. Initially, the focus was on the differences in the resources available to the schools (Jansen, 1995). In subsequent years, effective schools' researchers shifted their focus to examining the characteristics and practices that distinguish highly effective schools from the least effective ones (Jansen, 1995).

By the mid- to late-1970s the methods employed by the effective schools' researchers became more sophisticated enabling the revelation of differing practices between highly effective schools and less effective schools. Klitgaard and Hall (1975) examined multi-year data involving different grade levels and different cities with the intention of identifying the variables that are present in effective schools. Then in the late 1970s, Brookover et al. (1978) conducted a study involving 68 elementary schools that was pivotal to the effective schools movement as it was the first study to cite the importance of school climate.

From the late 1970s through the early 1980s the work of Ron Edmonds emerged as a major force in the effective schools movement. Marzano (2000) asserts that Edmonds is the father of the effective schools movement (see also Thrupp, 2000). Through his research, Edmonds (1979, 1982) earned credit for best articulating the belief that schools can make a difference, and for coining the definition of an effective school

(Marzano, 2000). In his 1979 article, "Effective Schools for Urban Poor," Edmonds proposed six characteristics of effective schools:

> (a) They have strong administrative leadership without which the disparate elements of good schooling can neither be brought together nor kept together, (b) Schools that are instructionally effective for poor children have a climate of expectation in which no children are permitted to fall below minimum but efficacious levels of achievement, (c) The school's atmosphere is orderly without being rigid, quiet without being oppressive, and generally conducive to the instructional business at hand, (d) Effective schools get that way partly by making it clear that pupil acquisition of basic school skills takes precedence over all other school activities, (e) When necessary, school energy and resources can be diverted from other business in furtherance of the fundamental objectives, and (f) There must be some means by which pupil progress can be frequently monitored. (p. 22)

In subsequent publications, Edmonds (1982; See also Lezotte, 1991) combined the fourth and fifth characteristics into one, thereby revealing the five original correlates of effective schools which are displayed in Table 2.1. Despite his strong belief that schools could make a difference, Edmonds acknowledged the influence of the family unit on the success of students (Lezotte, 2009); however, he did not identify it as being a correlate of effective schools. It is likely though, that Edmonds' research on the importance of family involvement contributed to subsequent researchers recognizing parent involvement as a correlate of effective schools.

Table 2.1: Correlates of Effective Schools

Edmonds (1979)	Lezotte (2005)
Strong administrative leader	Instructional Leadership
High expectations for student achievement	Climate of High Expectations for Success
Orderly atmosphere conducive to learning	Safe and Orderly Environment
Emphasis on basic skill acquisition	Opportunity to Learn & Time on Task (Focus on mastery of essential skills)
Frequent monitoring of student progress	Frequent Monitoring of Student Progress
	Positive Home-School Relations (Parents support the school's mission)
	Clear and Focused Mission (Emphasizing teaching for the learning of all)

In the 1980s and 1990s, numerous researchers published the results of their effective schools' studies (Lezotte, 2009; Lezotte & Snyder, 2011; Purkey & Smith, 1982, 1983; Rowan et al., 1983; Zigarelli, 1996). Review of their work reveals great variety in the number of correlates identified as well as how these correlates were defined. Amongst these researchers was Lawrence Lezotte (1991, 2009), who identified seven correlates of effective schools. As displayed in Table 2.1, great commonality exists between Edmonds' original five correlates and the seven correlates identified by Lezotte. In response to the findings of the effective schools research, in conjunction with Edmonds' acknowledgement of the influence of the family on academic achievement, Lezotte included positive home-school relations as being a correlate of effective schools. As is consistent with Brookover et al.'s (1978) findings citing the importance of schools having a clear and focused mission, Lezotte (1991, 2009) recognized its importance by including it as one of the seven correlates of effective schools.

Criticisms of Effective Schools Research

Just as criticism sparked the conception of the effective schools movement, criticism of its methodology has fostered evolution within the field. Despite the efforts of numerous researchers (Brookover et al., 1978; Edmonds, 1979, 1982; Klitgaard & Hall, 1975; Purkey & Smith, 1982, 1983; Rowan et al, 1983) who believe that schools can make a difference, the multitude of criticisms pointed at the effective schools research has crippled its credibility (Cuban, 1998; Jansen, 1995). In the field of research methodology, it is important for research samples to be of appropriate size, to be representative of the population to which the results are to be generalized, and to be randomly selected. The research cited below suggests that the effective schools research fails in each of these three areas.

The effective schools research has frequently been criticized for its small sample sizes (Jansen, 1995; Purkey & Smith, 1982, 1983; Witte & Walsh, 1990). Some effective schools' studies, especially those utilizing case study approaches, are comprised of too few schools for the results to be generalized to other schools. Despite the limited sample sizes, the effective schools movement has relied heavily on the use of case studies which have varied in their "levels of rigor and quantification" (Witte & Walsh, 1990, p. 190).

Likewise, the effective schools research is commonly cited as using samples that are not truly representative of all US schools (Purkey & Smith, 1982, 1983). The representativeness of the schools has been limited by its focus on elementary schools

(Purkey & Smith, 1982, 1983; Witte & Walsh, 1990), and schools in urban areas with high proportions of students from low-income families (Purkey & Smith, 1982, 1983; Rowan et al., 1983). Some researchers (Rowan et al., 1983) assert that the methods employed by effective schools' researchers favored the selection of smaller schools thereby limiting the ability to generalize the results to larger schools. Effective schools' researchers do not acknowledge the limited generalizability of their findings (Purkey & Smith, 1982, 1983; Rowan et al., 1983). Moreover, Rowan et al. (1983) note that the effective schools research rarely considers the interaction effects that may occur due to some variables being differentially influential in response to variations in school characteristics.

Use of outlier schools has drawn a myriad of criticisms. Researchers note that the effective schools' researchers' use of outlier schools fails to provide the field of education with information about the full spectrum of schools (Purkey & Smith, 1982, 1983; Witte & Walsh, 1990) since they do not include typical schools (Rowan et al., 1983). Purkey and Smith (1983) assert that, "Unless schools are capable of making quantum leaps in effectiveness, it will probably not greatly profit a very poor school to compare itself with an exceptionally fine school" (p. 432). Consequently, researchers need to include schools that span the full spectrum from highly effective to ineffective (Purkey & Smith, 1982, 1983; Rowan et al., 1983).

Another major criticism of the effective schools research is in the methods employed for selecting schools. Since schools are selected that are unusually effective or unusually ineffective, a random sample is not obtained (Purkey & Smith, 1982, 1983;

Rowan et al., 1983). Numerous criticisms are cited in the literature (Purkey & Smith, 1982, 1983; Rowan et al., 1983) directed at the methods utilized to identify effective schools including, but not limited to the use of aggregated data, the subjectivity in the selection of effective schools, the poor quality assessment methods, and the lack of consistency in effective schools over time.

Effective schools research conducted at the elementary school level typically utilizes student performance on standardized achievement tests as the only outcome measure (Rowan et al., 1983; Witte & Walsh, 1990). This approach looks at achievement in limited academic areas and at a limited number of grade levels (Rowan et al., 1983). Moreover, the use of standardized achievement tests, as the only outcome measure, fails to consider the fact that schools may have different or additional goals (i.e., social emotional learning; Rowan et al., 1983). Unfortunately, there are limited options available for outcome measures at the elementary level (Witte & Walsh, 1990).

Not only are there concerns with the use of achievement tests as the lone outcome measure, but also many critics of the effective schools research cite the use of aggregated data as being a major flaw. The use of aggregated data fails to consider the variability that exists within schools (Jansen, 1995; Purkey & Smith, 1982, 1983; Rowan et al., 1983; Witte & Walsh, 1990). These same researchers (Jansen, 1995; Purkey & Smith, 1982, 1983; Rowan et al., 1983; Witte & Walsh, 1990) note that there is more variability within schools than there are between schools. Yet the effective schools research discounts the within schools' variability through its use of aggregated data. To this end, Purkey and Smith (1983) question, “Is a school effective if there is great variance between its lowest- and highest-achieving students?” (p. 448).

Rowan et al. (1983) identified four approaches utilized in the effective schools research for measuring instructional effectiveness:

- Absolute measures analyze the proportion of students scoring above or below the median national performance.
- Trend score analyses explore whether scores at particular grade levels are increasing or decreasing.
- Gain score analyses examine whether the scores for a particular group of students (i.e., class of 2020) are improving or declining relative to the normative group.
- Regression-based techniques involving the computation of residuals: researchers identify expected achievement for a group based on the school composition, then compare actual performance to expected performance to identify groups that are performing significantly above or below their expected levels. (p. 26)

Each of these approaches has its weakness (Rowan et al., 1983). Comparison of the four approaches by Frechtling in 1982 revealed inconsistencies between these approaches as well as instability over time (as cited in Rowan et al., 1983) suggesting that none of these approaches, when used in isolation, are a reliable and valid method for assessing instructional effectiveness (Rowan et al., 1983).

Moreover, the effective schools research has been criticized for using subjective criteria (Jansen, 1995; Purkey & Smith, 1982, 1983) and faulty approaches (Rowan et al., 1983) in its selection of unusually effective and unusually ineffective schools. One research study (Purkey & Smith, 1982, 1983) found that some unusually effective schools performed worse than unusually ineffective schools when contextual variables, such as SES, were controlled for. Furthermore, researchers have found inconsistency in schools that would be considered effective from one year to the next. Purkey and Smith (1983) note that it is unusual to find a school that performs significantly above or below the mean for multiple years. Meanwhile, Rowan et al. (1983) discovered that schools that fell within the highest quartile "had only a 50 percent chance of being effective the next year" (p. 26). Consequently, researchers believe that data from multiple years needs to be considered when categorizing schools as being highly effective (Purkey & Smith, 1982, 1983).

As noted previously, the myriad of effective schools' studies yielded a wide variety of effective schools' correlates (Jansen, 1995; Purkey & Smith, 1982, 1983), as well as a wide variety of descriptions thereof (Jansen, 1995). This lack of agreement, in conjunction with its heavy reliance on subjective perceptions from respondents (Purkey & Smith, 1982, 1983), has caused a lot of skepticism. Nevertheless, Purkey and Smith (1983) acknowledge that the high similarity amongst the obtained results is remarkable (p. 435).

The effective schools research has been criticized for its limited consideration of contextual factors such as students' background and the communities' culture. Criticisms offered by Dantley (1990) suggest that effective schools' researchers need to be cognizant of the culture of the community in which schools exist, and how that culture influences the school. Furthermore, Dantley (1990) asserts that effective schools' researchers' lack of multicultural considerations imposes the mainstream culture onto other cultural groups. Likewise, Thrupp (2000) claims that by not considering students' background, the effective schools research fails to recognize the influence of these factors on students' achievement outcomes. In response, Townsend (2001) cites many studies that do consider these contextual factors, then declares that effective schools' researchers recognize the influence that students' background and life circumstances have, but focus on alterable factors over which schools can have control (p. 120). Nevertheless, Rowan et al. (1983) declare that "It is important to control for the effects of [these] 'hard to change' variables" (p. 28). In contrast, Edmonds (1979) argued that "repudiation of the social science notion that family background is the principle cause of pupil acquisition of basic school skills is probably prerequisite to successful reform of public schooling for the children of the poor" (p. 23).

The effective schools' researchers' limited scope of interest has triggered other criticisms. Thrupp (2000) faults the effective schools research for its limited theoretical basis and for not considering social theory: the notion that problems should be examined through not only an empirical approach, but also through active contemplation of the impact that context has on the existence of social problems (Harrington, 2005). Townsend (2001) acknowledges the narrow research agenda as being limited to "school effects, characteristics of effective schools, context variables, teacher effects and the relationships between each of these as well as the application of the findings of these studies into school improvement, and that 'we are comfortable with this narrow agenda'" (p. 121).

Given the intent of the effective schools research to identify the correlates of effective schools and further the research on continuous school improvement, a concern offered by Dantley (1990) warrants consideration. Dantley (1990) criticizes the effective schools research for its lack of a social justice component. Dantley (1990) explains that even though the effective schools research has provided valuable information to the field of education, it does not empower others to act to address the inequities that exist. Thrupp (2000) suggests that inequalities in education, as related to family background, are socially constructed phenomena that need to be challenged. Likewise, Edmonds (1979) noted that in order to attain equitable educational opportunities, educators must empower parents to get involved in educational policy and instructional reform (p. 23).

Criticisms of the effective schools research have been directed at the limited utility of the results thereof. Rowan et al. (1983) note that the effective schools research provides no information about causation. Moreover, the effective schools research has been criticized for its inability to provide information about effect size for individual correlates since the effects of the correlates are tested one at a time (Rowan et al., 1983). To address these weaknesses, Rowan et al. (1983) demand the use of "multivariate data analysis strategies" (p. 28) such as multiple regression analyses. In addition, the effective schools research has been criticized for not providing information about how school-level factors influence the teaching/learning process (Rowan et al., 1983).

Despite the remarkable commonalities amongst the effective schools' studies, which spanned multiple decades, suggesting that there is some credibility to their findings (Purkey & Smith, 1982, 1983), "by the mid-1980s, researchers had pronounced the effective schools research seriously flawed in theory, design and methodology" (Cuban, 1998, p. 466). Then "by the end of the 1980s, policymakers spoke publicly less and less of 'effective schools' [research] and far more about … 'systemic reform'" (Cuban, 1998, pp. 462-463). By the early 1990s, the frequency of effective schools' studies had decreased dramatically (Jansen, 1995). Then in 1995 Jansen declared:

> It is fair to conclude that effective schools research has reached a definite cul-de-sac in the 1990s placing the research agenda exactly where it was left by Coleman and others 30 years earlier, i.e., that schools have limited effects on student performance. (p. 186)

Given the negative perceptions of the effective schools research it became a taboo topic in the field of educational research thereby limiting the availability of current effective schools research. Even though the effective schools research essentially ceased, the effective schools movement persisted.

Given the value that effective schools' researchers place on equity, its theoretical approach resonated with educational reformers and policy makers, and as such continued to influence educational policy and educational reform, however, it was rarely referenced by name (Cuban, 1998; see also Smith & O'Day 1990). Cuban (1998) asserts that in their 1990 article, "Systemic School Reform," Smith and O'Day "made clear that the ancestry of 'systemic reform' was in the effective schools movement" (p. 464). Even though Smith and O'Day (1990) only explicitly identified the effective school research by name a few times, they based many of their recommendations for systemic reform directly from the effective schools research. For example, Smith and O'Day (1990) cited many conditions important for systemic reform including, but not limited to: “a unifying vision and goals” (p. 246) built on a set of “collective democratic values…. [and] derived from a deeper system of shared beliefs” (p. 246); “a positive climate and atmosphere” (p. 254); “that all students have the opportunity to acquire a core body of challenging and engaging knowledge, skills, and problem-solving capacities” (p. 247); as well as, shared and strong leadership (p. 253). Cuban (1998) further notes that the lead author, Marshall Smith, was the Undersecretary of Education and later the Deputy Secretary during the Clinton administration. Through these roles, Smith was instrumental in drafting numerous educational bills that promoted the “shift toward nationalization of the effective schools movement without once using the phrase *effective schools*” (Cuban, 1998, p. 464). Smith’s approach reflects the ongoing negative perceptions of referring to the effective schools research by name, while recognizing the contributions that this body of research has had on the fields of educational research, reform, and policy.

Equality of Educational Opportunity

As is consistent with the findings of Coleman et al. (1966), current educational research suggests that schools in the US are providing less to the students who need the most (Baker, Farrie, Johnson, Luhm, & Sciarra, 2017; Hall & Ushomirsky, 2010; OECD, 2010; Peske & Haycock, 2006; Schmidt, Cogan, Houang, & McKnight, 2009; Sharma et al., 2014). One variable of particular interest is per pupil expenditures. Even though there is much debate in the field of educational research as to the relevancy of funding practices, many researchers (Baker et al., 2017; Darling-Hammond, 2010; Hall & Ushomirsky, 2010; O’Gorman, 2010; Peske & Haycock, 2006; Schmidt et al., 2009) believe that this issue cannot be ignored as an important variable in effective schools research. Darling-Hammond (2010) asserts that even though money alone cannot improve educational outcomes, educators and policy makers need to be cognizant of its influence. Research findings suggest that "school resources … covary with pupil characteristics" (Darling-Hammond, 2010, p. 118; see also, Owens, 2010); staff quality, class size, per pupil expenditures, and average teacher salary were found to be significantly related to educational outcomes (Darling-Hammond, 2010). Of particular interest to the research at hand is the question: Are ethnic minority and students from low-income families provided with equitable educational opportunities?

New Hampshire's 2011 revision of its *Equity Plan* offers insight into the status of equitable educational opportunities for NH's students. According to this report, it appears that "the poor and minority students are largely distributed among the population of a *relatively prosperous* [emphasis added] state" (p. 1). Hillsborough County, which is home to NH's two largest cities, Manchester and Nashua, has a poverty rate of 5.6%. In contrast Sullivan and Coos counties, which are the two most rural counties in NH, have poverty rates of 9.1% and 10.4% respectively. The results of Ayscue and Jau's (2014) analysis of the demographics of NH students as of 2010 suggests that NH's low-income students are not evenly distributed in the state.

Likewise, Ayscue and Jau (2014) found that NH students are not evenly distributed by race. As of 2010, 90% of NH students were White, 4% were Latino, 3% were Asian, 2% were Black, and 1% were multiracial (Ayscue & Jau, 2014). Given these demographics it is not surprising that of the 480 NH schools included in Ayscue and

Jau's (2014) analysis, 97% were predominantly White. New Hampshire's White students typically attend schools that are representative of the state demographics, whereas NH's non-White students are more likely to attend schools that have higher proportions of their same-race peers than the state average which is consistent with research conducted by Rivkin (2016) who found that US schools are highly segregated by race due to *de facto segregation.* According to Ayscue and Jau (2014), slightly less than 1% of NH schools are predominantly minority schools while slightly more than 2% of NH schools are multiracial; multiracial schools are schools in which any three races each comprise at least 10% of the school's population. Both of the predominantly minority school districts are in the Manchester-Nashua region and educated "9.1% of [NH's] Latino students and 7.8% of [NH's] Black students" (p. 36). Even though only 25.2% of NH students were low-income in 2010, "84.9% of students in [predominantly] minority schools were low income" (p. 37) which is consistent with their finding that low-income students are more likely to attend schools in which there are higher proportions of other low-income students. The prevalence of low-income students in multiracial schools, 60.1%, far exceeded the state average of 25.2% (Ayscue & Jau, 2014). Overall, Ayscue and Jau's (2014) findings reveal a pattern of "double segregation" of NH students on the basis race and SES (p. 37).

Title I was designed to level the playing field for low-income students, but the funding structure at the national level is inequitable. Title I allocations are distributed based on two factors: (a) number of low-income students in a state, and (b) the average per-pupil expenditures for the state (Liu, 2006). Consequently, states that have more wealth per pupil tend to spend more money and receive more Title I funding. According to this funding structure, California which was home to 15.4% of the nation's low-income students in 2006 received $1,280 per student, whereas Wyoming, which has less than one percent of the nation's low-income students, received $2,957 per student. Arizona, having 2.5% of the nation's low-income students, received only $881 per eligible student. New Hampshire, which possesses 0.2% of this nation's low-income students (13,140), received $2,263 dollars per student.

The Education Trust (2006) examined the funding practices within each state to determine if states were equitable with the allocation of educational funds as intended through Title I legislation. After adjusting for cost of living and related factors, the highest and lowest quartiles of schools, in regards to the percentage of students living below the poverty line, within each state were identified. Then the funding gap between these two sets of schools was compared to reveal the funding gap for each state. New Hampshire was identified as one of four states which "shortchange their highest poverty districts by more than $1,000 per student per year" (p. 6). This finding was consistent with similar comparisons conducted in 2004 (Engel, 2010). On a related note, NH demonstrated the second largest funding gap between schools with the highest proportion of ethnic minority students and the schools with the lowest proportion of ethnic minority students: more than $2,000 per student per year. Even though inequitable funding practices were observed in NH's *Equity Plan* (2011), they were quickly dismissed since "funding does not always correlate to student achievement and teacher quality" (p. 1).

New Hampshire made improvements in subsequent years, but continues to provide less funding to the highest poverty schools (Ushomirsky & Williams, 2015). Since 2011, despite ranking in the top 12 states for funding level, NH has earned poor ratings for its equitable funding practices; NH tends to provide schools that have more need with less funding than their less needy counterparts (Baker et al., 2017). When the added costs of educating students from low-income families was considered, NH provided at least 5% less funding to high-poverty schools than it did to low poverty schools (Baker et al., 2017). Since a substantial amount of revenue in NH schools is derived from local sources such as property taxes (Education Trust, 2006a; Ushomirsky & Williams, 2015; Wiener & Pristoop, 2006;) readers are cautioned to note that the inequitable funding in the state is predominantly the result of inequitable revenue generation at the local level (Education Trust, 2006a; Ushomirsky & Williams, 2015). While this pattern is present in other states as well, many states allocate more state funding to offset the inequities in local funding (Education Trust, 2006a; Ushomirsky & Williams, 2015).

Perhaps more concerning is the finding of inequitable educational opportunities between schools within the same district. Title I money received at the district level is supposed to be used to supplement local funding sources in order to provide additional monies for high-poverty schools. However, the money typically is not allocated effectively (Hall & Ushomirsky, 2010; Roza, 2006; Ushomirsky & Williams, 2015). One reason for this is that unregulated funds are used to fund the programs and services in the low-poverty schools in the district that are funded by Title I monies in the high-poverty schools in that same district (Hall & Ushomirsky, 2010; Roza, 2006; Ushomirsky & Williams, 2015). Another reason is the practice of assigning teachers with the most experience to low-poverty schools while their less experienced colleagues are placed in higher-poverty schools (Hall & Ushomirsky, 2010; Peske & Haycock, 2006; Roza, 2006; Schultz, 2014; Wiener, 2008).

Compared to low-poverty and low-ethnic minority schools, several researchers (Clotfelter, Ladd, & Vigdor, 2010; Darling-Hammond, 2010; Hall & Ushomirsky, 2010; Minor, Desimone, Phillips, & Spencer, 2015; Owens, 2010; Peske & Haycock, 2006; Schultz, 2014; Sharma et al., 2014; Wiener, 2008) found that high-poverty and high-ethnic minority schools are more likely to have less experienced teachers, and teachers who lack appropriate background and training in the areas they are teaching. Hall and Ushomirsky (2010) explain that school districts "allocate teaching positions not dollars" (p. 2; see also U.S. Department of Education [US DoE], 2011). In this model, a teacher with no experience is perceived as being equivalent to a master level teacher with 20 years of experience. Since many school district budgets fail to disaggregate personnel salaries by school, the fact that high-poverty schools are spending significantly less on salaries while their low-poverty schools are spending much more is not readily apparent (Engel, 2010; Hall & Ushomirsky, 2010; US DoE, 2011).

The finding that high-poverty schools have less qualified teachers is particularly concerning considering the research available that indicates the importance of teacher qualifications on student achievement (Boyd, Lankford, Loeb, Rockoff, & Wyckoff, 2005; Clewell & Campbell, 2007; Darling-Hammond, 2010; Konstantopoulos, 2011;

Marzano, 2003; Minor et al., 2015; Peske & Haycock, 2006; Phillips, 2010; Pitre, 2014; Schultz, 2014; Sharma et al., 2014). Darling-Hammond (2010) notes that the achievement gap "between Black and White students is almost entirely accounted for by differences in the qualifications of their teachers" (p. 106). Likewise, through their three-year study comparing highly effective schools to typical schools, Clewell and Campbell (2007) found that "teachers in highly effective schools were significantly more likely to have earned degrees beyond bachelor's degrees" than were teachers in typical schools (p. 64). Phillips (2010) found that "graduate degrees in elementary or early childhood education were positively related to students' reading achievement gains" (p. 484). These results are consistent with the results of other recent studies (Boyd et al., 2005; Minor et al., 2015; Pitre, 2014; Schultz, 2014; Sharma et al., 2014) as well as the results of the Coleman Report (1966) which found that student achievement was highly correlated with teachers' educational backgrounds. Other researchers (Chetty, Friedman, & Rockoff, 2011; Konstantopoulos, 2011; Marzano, 2003; Peske & Haycock, 2006) assert that teacher effects, in regards to quality, are cumulative: meaning that multiple years with less qualified teachers will have compounding effects on student achievement.

In Ohio's schools, the most common strategy employed to address the inequities resulting from the most experienced teachers working in low-poverty schools was to increase the number of teachers in high-poverty schools, thereby reducing class sizes (Wiener, 2008). Unfortunately, this strategy has not proven effective except when the teacher-to-student ratio is decreased substantially (Wiener, 2008). Findings from the 2009 PISA suggest that the need to increase the number of teachers is not a sufficient solution

(OECD, 2010). A more promising approach is for school principals to focus on recruiting and retaining high quality and highly effective teachers. However, building principals tend to have little ability to negotiate with educators regarding salaries. Consequently, more veteran educators continue to move to low-poverty schools in the same district when vacancies arise: as working in these schools is perceived as being less challenging (Clotfelter et al., 2010; Guarino, Brown, & Wyse, 2010; Hall & Ushomirsky, 2010; Peske & Haycock, 2006; Wiener, 2008). Sharma et al. (2014) suggest that high-poverty and high-minority schools should be provided with extra funding to hire experienced and highly qualified teachers (i.e., teachers with advanced degrees). Unfortunately, research conducted by Clotfelter et al., (2010) suggests that pay differentials may not be sufficient to attract and retain highly qualified teachers. Per the OECD (2010), recruitment of highly qualified teachers into disadvantaged schools is an international problem.

In related research, Schmidt et al. (2009) examined the relationship between SES and opportunity to learn as measured by curriculum content coverage for eighth grade math students across the US. As part of their discussion, Schmidt et al. (2009) note that curriculum decisions are typically made at the local level; consequently, there can be great variability between schools in a district, state, or region. The obtained results show that SES influences content coverage, thereby influencing academic achievement (Schmidt et al., 2009).

No Child Left Behind

In 2001, Congress passed the No Child Left Behind (NCLB) Act. This legislation focused “attention to the unacceptably low levels of achievement among low-income students, non-White students, English-language learners, and students with disabilities" (Education Trust, 2011, p. 1). Furthermore, NCLB required states address these achievement gaps. NCLB required states participating in the Title I program of the Elementary Secondary Education Act (ESEA) to have clearly defined goals for student achievement, to assess the achievement of all students, and to have a process "for identifying and assisting individual schools that need improvement" (Education Trust, 2004a, p. 7).

Per the provisions of NCLB, states were required to hold school districts accountable for demonstrating adequate yearly progress (AYP). The requirements for making AYP were multifaceted. Not only did schools, as a whole, have to demonstrate AYP, but also schools had to demonstrate AYP for each of four specified subgroups of students: socioeconomic status, race, limited English proficiency, and students with disabilities. Many states, including NH, broke down the race category further into the following subgroups: Hispanic/Latino, American Indian or Alaskan Native, Black or African American, White, and other ethnicities (Karp, 2004). In addition, schools were required to demonstrate 90 percent participation rate at the whole school level and for each of the subgroups.

Even though these requirements seemed clear, many decisions regarding AYP were left to the individual states (Balfanz, Legters, West, & Weber, 2007). For instance, states could decide how they would measure achievement, and the criteria for establishing proficiency. These freedoms led to inconsistent implementation of NCLB and AYP from one state to the next (Balfanz et al., 2007; Ushomirsky, Hall, & Haycock,

2011) as well as inconsistent expectations between states (Peterson & Hess, 2008; Ushomirsky et al., 2011). Moreover, some researchers found that after the implementation of NCLB, several states lowered their expectations (Dee & Jacob, 2010; Peterson & Hess, 2008). These inconsistencies complicated efforts to conduct research on the effectiveness of NCLB (Dee & Jacob, 2010, 2011). Nevertheless, several studies (CEP, 2008, 2011; Dee & Jacob, 2010, 2011; Hall & Kennedy, 2006; Rowan, Hall, & Haycock, 2010) attempted to measure the progress of achieving these goals at a national level.

Research conducted by Hall and Kennedy (2006) and the Center on Education Policy (CEP, 2008) suggest that most states demonstrated improved reading and math achievement at the fourth grade level. Moreover, these studies suggested that gaps were closing between low-income and high-income students in reading and math; between Black students and White students in math; and between Hispanic students and White students in reading. According to Hall and Kennedy (2006), improvement at eighth grade and high school levels was minimal. Then in 2011, the CEP reported improvement in eighth grade math. Moreover, the CEP (2011) declared that gaps between subgroups were getting smaller more often than they are getting larger, except for eighth grade math. According to the NCES (2012) the reading and math achievement gaps between White and Black students, as well as between White and Hispanic students, remained relatively unchanged between 2008 and 2012. The only achievement gap that narrowed between 2008 and 2012 was the gap between White and Hispanic 13 years olds in the area of reading (National Center for Education Statistics, 2012).

Rowan et al. (2010) found evidence that low-income fourth grade students demonstrated improvements in reading “but at far too slow a pace” (p. 3) to effectively close the achievement gap. Through their research, they discovered that some states demonstrated impressive gains in reading achievement indicating that improvements are possible, but most states are not showing improvement. According to Rowan et al. (2010), states differ greatly in the size of their achievement gaps, and in their progress to close these gaps. Nevertheless, low-income and non-White students throughout the nation are demonstrating lower levels of achievement relative to their peers (Rowan et al., 2010; See also Ushomirsky et al., 2011). New Hampshire has not demonstrated strong success with closing achievement gaps, but there is some evidence of progress: Hispanic students demonstrated improvement in reading at a faster rate than White students; Low-income students demonstrated strong growth in reading while the performance of their high-income peers was stagnant (Rowan et al., 2010).

Dee and Jacobs (2010, 2011) examined the impact of NCLB accountability on achievement in reading and math at the fourth and eighth grade levels as assessed via performance on the National Assessment of Educational Progress (NAEP). They found evidence that significant improvements in mathematics have been made since the implementation of NCLB at the fourth grade level. These researchers observed that the strongest improvements were made by Black students, Hispanic students, and low-income students. In contrast, Pitre (2014) reports that the percentage of Black students who demonstrated proficiency increased by only 2% from 2009 through 2013 (16%-18%) while the percentage of proficient Hispanic students increased by 4% (22%-26%).

Meanwhile the percentage of proficient White students increased 3% (51%-54%). Dee and Jacobs (2010, 2011) note that improvements were observed with eighth grade math scores, but not to a significant degree. At the subgroup level, Hispanic eighth graders reportedly made significant improvements in math (Dee & Jacobs, 2010, 2011). In contrast, Pitre (2014) reported an improvement from only 17% proficiency in 2009 to 21% proficiency in 2013 on eighth grade math for Hispanic students. In contrast, 45% of White eighth graders demonstrated proficiency in math in 2013. None of these researchers (Dee & Jacobs, 2010, 2011; Pitre, 2014) reported improved NAEP reading scores at the fourth or eighth grade levels. After reviewing the NAEP reading and math results for 2005 through 2013, Pitre (2014) concluded that, "Although some small gains have been made, the discrepancy in educational outcomes – the 'achievement gap' – remains persistent" (p. 212).

Numerous researchers (Benitez et al., 2009; Darling-Hammond, 2004, 2010; Hursch, 2007; Karp, 2004) cite flaws with NCLB. A major source of criticism regarding NCLB is the high proportion of schools that have been identified as not demonstrating AYP. From its inception, some educational researchers predicted that most schools in this nation would not make AYP by 2014 (Darling-Hammond, 2004, 2010; Karp, 2004). These predictions are supported by research conducted by the Center for Education Statistics (Aud et al., 2012) which found that approximately 48 percent of schools in the US did not demonstrate AYP in 2011.

As school districts tried to avoid the punitive nature of NCLB, a myriad of injustices occurred to those students NCLB intended to help (Benitez et al., 2009;

Darling-Hammond, 2004, 2010; Hursch, 2007; Karp, 2004; Wood, 2004): schools not demonstrating AYP had higher rates of teachers with less training and less experience (Education Trust, 2006a; Hursch, 2007; Peske & Haycock, 2006), higher drop-out rates as students were pushed out of schools in order to avoid their taking state assessments (Bracey, 2008; Darling-Hammond, 2004; Hursch, 2007), and lower graduation rates (Darling-Hammond, 2010; Hursch, 2007). Schools narrowed the curriculum and taught to the test (Hursch, 2007). For these reasons, Hursch (2007) argues that NCLB perpetuated educational inequalities.

Despite the lack of conclusive empirical evidence suggesting that NCLB and AYP had positive effects, and even in the face of harsh criticisms from a myriad of educational researchers, there were several positive aspects of NCLB. Most notably, NCLB brought much needed attention to the achievement gaps that are plaguing this nation (Educational Trust, 2011). With this legislation states have increased their use of data-based decision making practices, increased curriculum alignment, increased the use of differentiated instruction, and increased the belief in shared ownership of students amongst educators (Bracey, 2008).

Resurgence of Effective Schools Research

Given the ongoing failure of US schools to close the achievement gaps there is considerable attention focused on improving schools. Callendar (2007) conducted a doctoral dissertation that examined the correlates of effective schools among exemplary Texas school districts. Vermont and Washington conducted outlier studies to better understand the characteristics of unusually effective schools. Though these two states took quite varied approaches to examining outliers, commonalities were observed in their findings.

Callendar (2007) examined the relationship among the seven effective schools correlates and school district practices among exemplary Texas elementary and secondary schools. Schools identified as being “exemplary” by the Texas Education Association for multiple years were selected for possible participation in the study. To be exemplary, schools must meet multiple criteria including 90% of its students, and each of the student groups meeting the standards established by the state (Callender, 2007). Twenty-nine elementary school campuses and 12 secondary school campuses participated. Participating schools, at each level, were sorted into four groups: economically advantaged exemplary schools, economically disadvantaged exemplary schools, homogenous (>75% White) exemplary schools, and diverse (>25% non-White) exemplary schools. Surveys were used to assess the presence of the seven effective schools correlates and related school practices. The obtained results provide strong evidence that the effective schools correlates are present in these exemplary schools. However, since no comparisons were made to non-exemplary schools, it cannot be concluded that the effective schools correlates are not also present among non-exemplary schools.

In 2007, Washington State published the second edition of its report in which they examined the results of their original study and conducted a meta-analysis of current literature on school district improvement (Shannon & Bylsma, 2007, p. 8; see also Shannon & Bylsma, 2004). Criteria was established to provide the researchers with guidelines for determining whether studies could be included in their study (Shannon & Bylsma, 2004, p. 67). They then identified nine characteristics of schools that were higher performing than would be expected given their demographics: (a) clear and shared focus, (b) high standards and expectations for all students, (c) effective school leadership, (d) high levels of collaboration and communication, (e) curriculum, instruction and assessments aligned with state standards, (f) frequent monitoring of learning and teaching, (g) focused professional development, (h) supportive learning environment, and (i) high level of family and community involvement (Shannon & Bylsma, 2007). It is important to note that some of the studies included in this meta-analysis were case studies (Shannon & Bylsma, 2004). Moreover, there was no discussion of whether the data employed by the included studies was aggregated or disaggregated. As noted previously, reliance solely on aggregated data undermines the within school differences that often exist between groups of students based on race and SES.

Then in 2009, the Vermont Department of Education conducted a statewide assessment of unusually effective schools. These researchers focused on moderate to large-sized Vermont schools that had a moderate to high proportion of students from low-income families. As such, the high-performing schools contained a moderate to high proportion of students from low-income families. Their findings revealed that teachers in schools that performed better on NECAP assessments consistently rated their schools higher than teachers in lower performing schools in seven areas: (a) high expectations for all students, (b) focus on student achievement, (c) frequent assessment of student progress, (d) support for struggling students, (e) staff collaboration, (f) effective school leadership, and (g) parental involvement (VT DoE, 2009; See also Hayes, 2008). The team of researchers then conducted intensive site visits at three of the high-performing schools. Following the site visits, the team concluded that in addition to the seven characteristics that emerged through the teacher surveys, effective Vermont Schools also possessed positive school climates. The Vermont study (2009) is of particular interest to the current study given its recognition of the effective schools research, its adherence to the importance of schools being effective for all students regardless of SES, and its close proximity to the state of NH.

The commonalities between these three studies and the effective schools research is promising in suggesting a potential resurgence of effective schools research. Each study had limitations though. The study conducted in Texas (Callender, 2007), focused exclusively on exemplary schools thereby limiting the generalizability of the results to other schools. More notably though, since comparisons were not made between exemplary and non-exemplary schools, the utility of the results is limited. The work conducted in Washington (2007) included case studies, may have only used aggregated data, and may not have considered schools' effectiveness for more than one year. The Vermont study focused exclusively on schools that had moderate to high proportions of students from low-income families (Vermont DoE, 2009) warranting the need to consider subgroup differences, based on SES, less important since most students in these schools were from low-income families. The Vermont study did not consider subgroup differences based on race though, and review of its methodology fails to suggest that the results considered a school's performance for more than one year. No other statewide assessments of effective schools, based on effective schools research, could be located.

Conclusion

During the past five decades, a plethora of educational research studies have been conducted and much has been learned. The effective schools movement has been a major contributor to this line of research and is especially noteworthy given that it is driven by the belief that an effective school is one that is effective for all students regardless of race or SES. The effective schools research provided evidence that some schools are effective for all students and identified several correlates associated with these unusually effective schools (Cuban, 1998; Lezotte, 1991, 2005, 2009; Lezotte & Snyder, 2011; Rowan et al. 1983). This line of research helped schools to make changes to increase their effectiveness (Cuban, 1998; Smith & O'Day, 1990). Unfortunately, due to numerous criticisms of its methods, the effective schools research lost the respect of educational researchers and as such came to a halt in the early 1990s (Cuban, 1998; Jansen, 1995). Through Marshall Smith's work, the effective schools movement continued to affect educational policies, but it was rarely mentioned by name (Cuban, 1998; Smith & O'Day, 1990). As such the effective schools research has not yet regained prestige in the field of educational research or educational policy.

The research on variables associated with the effectiveness of teachers, school districts, funding practices, and related factors have continued, but these studies rarely focused on the school as the unit of focus or adhere to the belief that effective schools are schools that are effective for all students. Nevertheless, these studies provided valuable insight into the influence that school funding practices, student body composition, and teacher qualifications have on the achievement gaps in the US.

Even though nearly a half century has passed since the passage of ESEA and the release of the Coleman Report, and despite the efforts of Title I of the ESEA and NCLB, equality of educational opportunity has not yet been achieved (Rivkin, 2016). Within the US achievement gaps exist between White students and non-White students (Benitez et al., 2009; Darling-Hammond, 2010; Domina, 2006; Edelman & Jones, 2004; Education Trust, 2008a; Education Trust, 2011; Leach & Williams, 2007; Lezotte, 2009; McKinsey & Company, 2009; Pitre, 2014; Rivkin, 2016; Sharma et al., 2014), and between groups of students based on SES (Benitez et al, 2009; Darling-Hammond, 2010; Kozol, 2005; Lezotte, 2009; McKinsey & Company, 2009; Sharma et al., 2014; Singh, 2015; Zhang & Cowen, 2009). These achievement gaps are observed among elementary students in NH (NH DoE, 2013a, 2013b, 2013c). Research at the national level suggests that these achievement gaps are perpetuated by inequitable funding practices (Baker et al., 2017; Darling-Hammond, 2010; Hall & Ushomirsky, 2010; O'Gorman, 2010; Peske & Haycock, 2006; Schmidt et al., 2009) and teacher qualifications (Clotfelter et al., 2010; Darling-Hammond, 2010; Hall & Ushomirsky, 2010; Minor et al., 2015; Owens, 2010; Peske & Haycock, 2006; Schultz, 2014; Sharma et al., 2014; Wiener, 2008). Several studies have revealed that NH provides less funding to its highest poverty schools (Baker et al., 2017; Education Trust, 2006a; Ushomirsky & Williams, 2015), but no studies have examined the equity of teacher quality in the state of NH.

Considering the commitment of the effective schools research to providing equitable education opportunities for all students, and the influence that it has had on educational policy and reform, this writer asserts that it is an essential body of research for addressing the inequitable educational opportunities afforded to students in the US. However, it is imperative that researchers wishing to contribute to this line of research address the multitude of criticisms it has received.

Chapter 3: Research Methodology

The research design and methodology employed in this study were based directly on the myriad of criticisms of the effective schools research. This line of research has frequently been cited for problems due to too few schools being included in its studies (Jansen, 1995; Purkey & Smith, 1982, 1983; Witte & Walsh, 1990), and the use of samples that were not randomly selected or truly representative of schools in the US (Purkey & Smith, 1982, 1983; Rowan et al., 1983; Witte & Walsh, 1990). A major flaw of the effective schools research has been its tendency to focus on unusually effective and unusually ineffective schools while omitting the schools in the middle of the spectrum (Purkey & Smith, 1982, 1983; Rowan et al., 1983; Witte & Walsh, 1990). To address these concerns, the current study examined all public elementary schools that had a combined total of at least 50 third, fourth, and fifth grade students take the 2012 Fall NECAP assessment, and had data available for these three grade levels for all three years (i.e., 2010, 2011, & 2012). When two schools combined to provide continuous education for kindergarten through fifth grade students, they were considered as one school. The merging of these schools into one school decreased the number of small schools that were eliminated allowing for 210 schools to be included in this study. By examining the full population of NH public elementary schools, the current study avoided many issues related to sampling bias.

Some researchers declare that the criteria used by effective schools' researchers to determine if schools are effective or not are too subjective (Jansen, 1995; Purkey & Smith, 1982, 1983) resulting in some ineffective schools outperforming effective schools

(Purkey & Smith, 1982, 1983), as well as schools labeled effective one year being declared ineffective the following year (Purkey & Smith, 1982, 1983; Rowan et al., 1983). To remedy these weaknesses, the current study established objective criteria that were consistently applied to all schools. Moreover, the classification of schools to groups was determined based on data from three consecutive years (2010, 2011, & 2012); to be assigned into one of the four groups of schools, schools had to perform consistently for all three years. Schools that were not consistently in the top, middle, or bottom of NH schools were not included in one of the four groups of schools.

In the area of research methodology and statistical analyses, the effective schools research has been criticized for its tendency to focus exclusively on aggregated data thereby neglecting to examine the variability that exists within schools (Jansen, 1995; Purkey & Smith, 1982, 1983; Rowan et al., 1983; Witte & Walsh, 1990). To address this weakness, the current study was comprised of three parts. Aggregated data was the main focus of data analyses in the first and third parts of this study. In the second portion of this study, disaggregated data was used to test for subgroup differences. The presence or absence of subgroup differences determined whether the high-performing schools would be classified as effective or high-performing low-equity schools.

The research design of the effective schools research has also been criticized for failing to provide information related to the effect size of individual variables (Rowan et al., 1983). Consequently, in the third part of this study, the individual effect of each of the explanatory variables (school composition [PWS & PFRL], PPE, & teacher qualifications [PTMDH]) was examined.

Finally, the effective schools research has been criticized for its utilization of standardized achievement tests as the only measure of school outcomes (Purkey & Smith, 1982, 1983; Witte & Walsh, 1990). Since there were no other statewide indicators of effectiveness available at the elementary school level in the state of NH, the current study was not able to address this criticism.

All data needed for this study were pre-existing and contained within the New Hampshire Department of Education's (NH DoE) data system; consequently, this study utilized a non-experimental research design. The current study was comprised of three parts. In the first part aggregated and disaggregated data were examined through a variety of quantitative analyses to identify three distinct preliminary groups of schools for reading: high-performing schools (HPS), moderately-performing schools (MPS), and low-performing schools (LPS). This process was then repeated to identify three distinct groups for math. For the second part of this study *t*-tests and related statistical methods were employed to analyze disaggregated data from each of the HPS to further divide them into two groups: effective schools (ES) and high-performing low-equity schools (HPLES). Multinomial logistic regression analyses were employed in the third part of this study to determine the individual and combined effects of the explanatory variables on the effectiveness of schools. The three variables of interest were student body composition, school funding, and teacher qualifications. Student body composition was assessed via two explanatory variables: Percentage of White students (PWS) and percentage of students eligible for free and reduced lunch (PFRL). School funding was assessed via per pupil expenditures (PPE). Teacher qualifications was assessed via the percentage of teachers with a master's degree or higher (PTMDH).

After a brief discussion of information related to the schools included in the initial part of this study, information pertaining to the NECAP assessments is provided. Then each of the three parts of this study will be described. Since the first part of this study does not directly relate to the research questions, the outcomes thereof will be provided in this chapter. Through this discussion additional information about the schools involved in this study will be provided. The results of the second and third parts of this study, which address the research questions, will be provided in chapter four. Given the nature of the current study, additional information about the schools involved in this study will be provided in chapter four.

Operational Definitions

For the purposes of the current study, the following operational definitions were employed:

- New Hampshire school – a school in which more than 50% of students live in NH.
- Elementary school – a public school comprised of at least grades K-5 or the combination of two schools that results in at least a K-5 configuration.
- Outcome variable – in logistic regression analyses, the outcome variable is similar to the dependent variable in traditional research methodology: it is the variable that is being studied to assess the effects of the independent variables (Hosmer, 2013). For this study, the outcome variable is the effectiveness of schools as

determined by the schools' identification as being either an effective school (ES), a high-performing low-equity school (HPLES), a moderately-performing school (MPS), or a low-performing school (LPS).

- Referent group – in multinomial logistic regression analyses, one level or category of the outcome variable is used as a referent to which the other levels or categories are compared (Hosmer, 2013). Logistic regression models are built to determine the odds of an outcome which, in the case of multinomial logistic regression analyses, is the reference group (Hosmer, 2013). For the current study, effective schools (ES) was identified as the referent group. As such, for the current study all statistics report the strength of the model at explaining which schools are effective instead of being HPLES, MPS, or LPS (Hosmer, 2013).
- High-performing schools (HPS) – are schools in which at least a minimum percentage (82% for reading; 80% for math) of all third, fourth, and fifth grade students who took the Fall 2012 NECAPs demonstrated at least proficient abilities, and the school, as a whole, performed in the top third of NH Elementary Schools in the Fall of 2011 and 2010.
- Effective schools (ES) – are high-performing schools wherein non-White students performed comparably with their White peers, and students from low-income families performed comparably with their peers, on the Fall 2012 NECAP.
- High-performing low-equity schools (HPLES) – are high-performing schools in which significant subgroup differences were observed on the basis of race and/or SES on the Fall 2012 NECAP.

- Moderately-performing schools (MPS) – are schools that performed in the middle 20% of schools as assessed by the percentage of third, fourth, and fifth graders who demonstrated proficiency on the Fall 2012 NECAP assessment, and were ranked in the middle third of schools on the 2010, and 2011 NECAP assessments.
- Low-performing schools (LPS) – are schools that performed in the lowest 20% of NH schools as assessed by the percentage of third, fourth, and fifth graders who demonstrated proficiency on the fall 2012 NECAP assessment, and were ranked in the lowest third of schools on the 2010 and 2011 NECAP assessments.
- Explanatory variables – in logistic regression analyses, the explanatory variables are similar to the independent variables in traditional research methods; they are the variables that are being assessed to determine how they influence the dependent or outcome variable (Hosmer, 2013; Pampel, 2000). For the current study, the explanatory variables were the percentage of White students (PWS), the percentage of students eligible for free and reduced lunches (PFRL), per pupil expenditures (PPE), and the percentage of teachers with a master's degree or higher (PTMDH).
- Student body composition – characteristics of the student body such as the proportion of low-income students and the proportion of non-White students. In this study, percentage of White students (PWS) and percentage of students eligible for free and reduced lunch (PFRL), were used to represent the student body composition of a school. Data regarding PWS and PFRL for each school was obtained from the NH DoE data system.

- White Students – are students identified via the NH DoE under the racial category of White.
- Non-White Students (NWS) – students identified via the NH DoE as being "American Indian/Alaskan Native," "Black (not Hispanic)," or "Multiple Races." Students identified as being "Asian or Pacific Islander" were not included in any of the subgroups based on race and ethnicity.
- Low-income students - are students identified by the NH DoE as being eligible for free and reduced lunch.
- School funding practices – the per pupil expenditures (PPE) of the elementary level schools in the school district in which the school resides as reported by the NH DoE's online data system.
- Teacher qualifications – the percentage of teachers with a master's degree or higher (PTMDH) for teachers in the school district in which the school resides. This data was obtained from the NH DoE online data system.
- School size – refers to the number of students in grades three, four, and five in a school, as defined above, who took the NECAPs in the Fall of 2012.
- NH county – refers to the county in which the school is located as reported by the NH DoE.
- Multinomial logistic regression analyses – employs a regression approach to determine the ability of explanatory variables to explain the observed categorical outcomes of a phenomena. These methods involve regressing the categorical outcomes on logit models comprised of explanatory variables with adjustment

models, when necessary, to construct a logit model that can best explain the observed outcomes (Hosmer, 2013). For the current study, the outcome variable, school effectiveness, was regressed on the logit models developed through this study in order to construct a logit model that could best explain why some schools are effective while others are not.

- Models – in multinomial logistic regression analyses, researchers construct logit models comprised of explanatory, and if necessary potentially confounding variables with the intent of assembling a model that can explain the largest proportion of variability in the outcome variable (Hosmer, 2013). For the current study, models were constructed with the intent of explaining as much of the variability in the effectiveness of schools. In other words, models were constructed that would best explain why some schools were effective while others were not.
- Empty model - to determine if the logit models are truly influencing the outcome variables, each logit model is compared to an empty model (Pampel, 2000). These empty models attempt to explain the outcome variable by chance of alone and therefore do not contain any confounding or explanatory variables (Pampel, 2000; see also Keith, 2006). These empty models are often referred to as being the “constant model” (Pampel, 2000), “the intercept only model” (Hosmer, 2013) or the “base model” (Hosmer, 2013).
- Adjustment models – just as in other research methods, the results of logistic regression analysis can be confounded by extraneous or mediating variables. To

control for these potentially confounding variables adjustment models are constructed (Hosmer, 2013). These adjustment models, if necessary, are included in the logit models being constructed to explain the influence of the explanatory variables on the outcome variables (Hosmer, 2013).

- Full model – the goal of multinomial logistic regression analyses is to construct a full model. A model that can account for the most variability in the outcome variable (Hosmer, 2013; Pampel, 2000). Since Effective Schools was identified as the referent group, the full model for the current study would be the model that can best explain why some schools are effective while others are not. When constructing the full model though, explanatory variables that are highly correlated with each other cannot both be included in the full model (Hosmer, 2013; Pampel, 2000). In these cases, the explanatory variable that can explain the most variability in the outcome variable would be included while the other would be excluded.

Participants

To address criticisms related to insufficient sample size (Jansen, 1995; Purkey & Smith, 1982, 1983; Witte & Walsh, 1990) and selection bias (Purkey & Smith, 1982, 1983; Rowan et al., 1983; Witte & Walsh, 1990), the current study included a statewide assessment of effectiveness. Given the large number of schools in the state of NH, it is beyond the scope of the current investigation to examine the effectiveness of all schools. Since the effects of schooling are cumulative (Marzano, 2003; Peske & Haycock, 2006) this examiner believed it practical to start at the elementary level.

The current study assessed effectiveness based on the percentage of third, fourth, and fifth grade students who demonstrated proficient or better abilities on the Fall NECAP assessments for three consecutive years (2010, 2011, & 2012). The decision to use three consecutive grade levels was made to increase the number of data points within each school while decreasing cohort effects (Purkey & Smith, 1982, 1983; Rowan et al., 1983). All schools in the study met the 95 percent participation rate. Thirty-four schools, were merged into 17 schools because together they combined to educate students up through at least fifth grade. Fifty-four schools that had fewer than 50 students taking the Fall NECAP in 2012, and were not appropriate for merging with other schools, as well as one additional school, were excluded from the current study. Following these mergers and exclusions, 210 schools were included in this study. Since the research design required the establishment of four distinct groups of schools, based on schools' consistent performance in three consecutive years, about half of the schools did not meet the criterion for inclusion in one of the groups and as such were not included in the multinomial logistic regression analyses. The sorting process will be described below, and additional information about the schools involved in this study will be provided.

Instrumentation

The New England Common Assessment Program (NECAP) is the product of collaboration among four New England States: Maine, New Hampshire, Rhode Island, and Vermont. The NECAP was created to provide states with the data requisite to report on student achievement in the areas of reading/English language arts and mathematics, as well as to provide states, schools, and communities with data that can be used for program evaluation and if needed, reformation (Measured Progress, 2010, 2011, 2012). Consequently, the NECAP assessed reading and math achievement in students in grades three through eight and eleven, in the fall of each school year. Students in grades five, eight, and eleven were also assessed in writing (Measured Progress, 2010, 2011, 2012).

NECAP testing was completed in the fall of each year and designed to assess "student achievement on the prior year's grade level expectations" (Measured Progress, 2012, p. 1). Even though there were multiple forms of the NECAP tests at each grade level each year, all students were administered a common set of items. Each form contained a variety of other items which were designed to either assist in the process of equating tests from one administration year to the next, or for field testing items that were being considered for subsequent years (Measured Progress, 2010, 2011, 2012). Student performance on equating items and the field test items were not considered when computing student scores (Measured Progress, 2010, 2011, 2012). Since the current study examined student performance in the fall of third, fourth, and fifth grades, further discussion of the NECAP will be limited to content specific to these grade levels.

The items for the NECAP assessments were developed, revised, and adopted or discarded through a multi-step process to ensure good validity (Measured Progress, 2010, 2011, 2012). All items were developed by a team of individuals who had expertise in test construction as well as experienced educators (Measured Progress, 2010, 2011, 2012). More items were developed than expected to be adopted. Each test item was reviewed by content area experts as well as individuals who have special expertise in assessing bias in tests (Measured Progress, 2010, 2011, 2012). Then Measured Progress test developers thoroughly reviewed each item (Measured Progress, 2010, 2011, 2012). Next, each of the four states formed item review committees comprised of teachers, curriculum specialists, and higher education faculty that reviewed each item (Measured Progress, 2010, 2011, 2012). After additional review for potential bias, the item review committees from each state, in conjunction with test developers from Measured Progress, collaborated for the final item selection process (Measured Progress, 2010, 2011, 2012). Additional review and development examined issues such as item placement on test pages, grammar, bias, and visual appeal (Measured Progress, 2010, 2011, 2012).

The structure of the Reading NECAP is consistent from year to year. Each year the common set of reading questions included 28 multiple choice questions: two in response to long passages, two in response to short passages, and four stand-alone questions (Measured Progress, 2010, 2011, 2012). In addition, six constructed response questions, which were designed to assess higher-order thinking skills, were included (Measured Progress, 2010, 2011, 2012). Likewise, the emphasis on specific subskills, though different at different grade levels, showed little to no variability from year to year (Measured Progress, 2010, 2011, 2012).

The NECAP was designed to assess different *depths of knowledge* (Measured Progress, 2010, 2011, 2012). Per Measured Progress (2012), “depth of knowledge is not synonymous with difficulty” (p. 7). Rather depth of knowledge refers to the “complexity of the mental processing a student must use to answer the question” (Measured Progress, 2012, p. 7). *Level 1, Recall*, assesses a student’s ability to “retrieve or recite facts or to use simple skills or abilities” (p. 7). *Level 2, Skill/Concept*, assesses one’s comprehension of, and ability to make inferences based on information contained with the text. *Level 3*, *Strategic Thinking*, assesses a student's ability to not only comprehend text, but also to think abstractly, and make inferences based on the entire passage or their prior knowledge; this level is not assessed in third and fourth grade students. From year to year there is variability in the proportion of Level 1 and Level 2 questions included (Measured Progress, 2010, 2011, 2012).

Information contained in the technical manuals indicates that the Reading NECAPs have good reliability (Measured Progress, 2010, 2011, 2012). Cronbach's Alpha revealed strong ($\alpha \geq .88$) overall reliability at each grade level (Measured Progress, 2010, 2011, 2012). Moreover, analysis at the subgroup level revealed adequate levels of reliability ($\alpha \geq .83$; Measured Progress, 2010, 2011, 2012).

Analysis of the components of the Math NECAPs taken by third, fourth, and fifth grade students, reveals a similar structure. For third and fourth grade students, each test was comprised of 35 multiple choice questions, and 20 short answer questions. For fifth grade students, tests each year were comprised of 32 multiple choice questions, 12 short answer questions, and 4 constructed response questions (Measured Progress, 2010, 2011, 2012). At each grade level, the distribution of focus across the four subcategories (Numbers & Operation; Geometry & measurement; Functions & Algebra; Data, Statistics & Probability) remained consistent across testing years (Measured Progress, 2010, 2011, 2012).

As with the Reading NECAP, the Math NECAP was designed to assess differing depths of knowledge. *Level 1,* was designed to assess one's ability to "recall information and to carry out simple procedures" (Measured Progress, 2012, p. 10). *Level 2*, was designed to assess students' abilities to "make some decisions about how to approach a problem" (Measured Progress, 2012, p. 10). *Level 3*, was designed to assess a variety of higher level mathematical thinking and reasoning abilities: "strategic thinking, reasoning, planning, drawing conclusions and using concepts and evidence" (Measured Progress, 2012, p. 10). The depths of knowledge assessed via the Math NECAP remained stable at all three grade levels (Measured Progress, 2010, 2011).

The Math NECAPs have good reliability (Measured Progress, 2010, 2011, 2012). Cronbach's Alpha revealed strong ($\alpha > .90$) overall reliability at each grade level (Measured Progress, 2010, 2011, 2012). Moreover, analysis of reliability at the subgroup level revealed strong results ($\alpha > .90$); Measured Progress, 2010, 2011, 2012).

Part 1: Formation of Three Preliminary Groups

To begin, a list of all NH elementary and middle schools was obtained from the NH DoE. Middle schools were included in this review of data since the configuration of schools vary from district to district and sometimes within districts. In situations when students in a school district naturally progress between two or three consecutive schools during the course of third, fourth, and fifth grades, the schools were considered as one school. Thirty-four schools were merged into seventeen combined schools. Their obtained results were combined using weighted means when data analysis involved aggregated data. Analyses involving disaggregated data considered all third, fourth, and fifth grade students in the combined schools.

In the first part of this study, the performance of each of the 210 schools in reading and math was reviewed against the first two criteria for determining group assignments as set for the purposes of the current study. The first criterion dealt with the percentage of third, fourth, and fifth grade students within each school who demonstrated at least proficient abilities on the 2012 NECAPs. The second criterion was that schools had to perform comparably on all three years considered (i.e., 2010, 2011, & 2012).

The first criterion that was used to sort schools involved the percentage of third, fourth, and fifth graders who demonstrated at least proficient abilities on the Fall 2012 NECAPs. At the onset of this study, the first criterion was set that for a school to be considered high-performing, at least 80% of its third, fourth, and fifth grade students had to demonstrate proficient or better abilities on the Fall 2012 NECAPs. To determine if schools met this first criterion, the percentage of students who performed at each of the four levels of performance on the Fall 2012 Reading NECAP and Math NECAP assessments were analyzed. The data for every third, fourth, and fifth grade student in each school was converted such that students who performed in the *Substantially Below Proficient* or *Partially Proficient* levels were coded as not demonstrating proficiency while students who performed in the *Proficient* or *Proficient with Distinction* levels were coded as demonstrating proficiency. Then SPSS Cross Tabulations were conducted for all schools to identify their proficiency rates for reading and math. The obtained results were used to create a spreadsheet of schools based on proficiency rates in reading, and a separate spreadsheet of schools according to proficiency rates in math. Proficiency rates were rounded up such that proficiency rates at or above 79.5 were considered to meet the criterion. A total of 105 schools met this criterion for reading and 80 met the criterion for math.

Next the bottom 20% of schools were identified for both reading and math. Given that there were 210 schools at this stage of the study, the 42 schools that had the lowest proficiency rate for reading were identified as being LPS for reading. Likewise, the 42 schools that had the lowest proficiency rate for math were identified as being LPS for math. This process was repeated to identify the MPS whose proficiency rate fell in the middle 20% of NH schools. Two schools at the lower bounds of the middle 20% for reading, and two schools at the upper bounds of the middle 20% for math had the same proficiency rate. Consequently, the MPS groups each contained 43 schools.

The distinction between the middle 20% and the 80% proficiency rate for the top performing schools was definitive for math as four schools fell between these cut points. In contrast, 21 schools met the 80% proficiency rate for reading while also falling in the middle 20% of schools. Given this substantial overlap, the criterion for an effective school for reading was adjusted such that at least 82% of students in a school had to demonstrate proficiency on the 2012 Reading NECAPs for the school to be considered effective. This adjustment resulted in 84 schools meeting the proficiency rate for reading which is more consistent with the number of schools that met the proficiency rate for math (i.e., 80), and decreased the number of schools that concurrently met the criterion for HPS and the criterion for MPS.

The next criterion for determining group membership required that schools perform consistently for all three years. To meet this criterion high-performing reading schools had to have performed in the top third of NH Schools on the 2010 and 2011 Reading NECAPs. Moderately-performing reading schools had to have performed in the middle third of NH Schools on the 2010 and 2011 Reading NECAPs. Low-performing reading schools had to have performed in the bottom third of NH Schools on the 2010 and 2011 Reading NECAPs. These same criteria applied for the high-performing, moderate-performing, and low-performing math schools.

Consequently, the NECAP results for the Fall 2010 and 2011 math and reading assessments were obtained from the NH DoE website. Then the data files were sorted by overall performance in reading and in math. The top third, the middle third, and the bottom third of schools, in both reading and in math, were identified for 2010 and 2011. Then the data were reviewed to identify the schools that met the criteria for being HPS. Initially, 48 schools were identified as being high-performing in reading and 49 in math. Many schools were identified as being HPS in both reading and math.

As noted previously, the separation between the schools in which at least 82% of the students demonstrated proficiency in reading in the Fall of 2012, and the middle 20% of NH schools was not decisive. Special care was taken to consider the appropriate group placement for the six schools that demonstrated the requisite proficiency rate yet fell in the middle 20% of schools. Of these six schools, one performed in the top third of NH schools on the 2010 and 2011 Reading NECAP assessments. Since this school met both criteria for inclusion in the high-performing group, it was classified as being a HPS bringing the total number of high-performing reading schools to 49.

Table 3.1: Criteria & Preliminary Group Sizes for Reading

	2012 NECAP	2011 NECAP	2010 NECAP	Group Size
High Performing School	82% or higher proficiency rate	Top 33% of NH schools	Top 33% of NH Schools	49
Moderately Performing School	Middle 20% of NH Schools	Middle 33% of NH Schools	Middle 33% of NH Schools	19
Low Performing School	Bottom 20% of NH Schools	Bottom 33% of NH Schools	Bottom 33% of NH Schools	38
			Total	106

Note. Other schools: 104 schools did not meet the criteria for any of these three groups. These schools were considered in the computation of the percentage of NH elementary schools effective for reading, but not in subsequent analyses.

Next the LPS and MPS were identified. As noted in Table 3.1, 38 schools were identified as being LPS for reading since they performed in the lowest 20% of NH schools on the 2012 Reading NECAPs, and performed in the bottom third of NH schools on the 2011 and 2010 Reading NECAP assessments. The same process was employed to identify the LPS for math of which 36 were identified (see Table 3.2). Next, the MPS were identified. Eighteen schools were identified as being moderately-performing math

Table 3.2: Criteria & Preliminary Group Sizes for Math

	2012 NECAP	2011 NECAP	2010 NECAP	Group Size
High Performing School	80% or higher proficiency rate	Top 33% of NH schools	Top 33% of NH Schools	49
Moderately Performing School	Middle 20% of NH Schools	Middle 33% of NH Schools	Middle 33% of NH Schools	17
Low Performing School	Bottom 20% of NH Schools	Bottom 33% of NH Schools	Bottom 33% of NH Schools	36
			Total	102

Note. Other Schools: 108 schools did not meet the criteria for any of these three groups. These schools were considered in the computation of the percentage of NH elementary schools effective for math, but not in subsequent analyses.

schools as they performed in the middle 20% of schools on the 2012 Math NECAPs, and in the middle third of schools on the 2011 and 2010 Math NECAPs. The same process was repeated for reading resulting in the identification of 21 MPS. None of the five remaining schools, that met the minimum proficiency rate for reading in 2012, yet also fell in the middle 20% of NH schools, performed in the middle third of NH schools on the preceding two years' Reading NECAP assessments. Consequently, these five schools were not identified as being MPS.

Given the decision to only include schools that clearly met the pre-determined criteria for assignment to the groups, many schools were not put into one of the three preliminary groups. Of the 210 schools that were included in the initial phase of this study, only 106 were placed into groups for reading while only 102 schools were placed into groups for math. The schools that were not placed into groups either did not meet the criterion for the groups based on their Fall 2012 NECAP proficiency rate, or did not

Table 3.3: Representativeness of Math Schools by County, District, and Size

County	n	%	$\bar{x}$	Districts
Coos	2	2.0	152	2
Grafton	3	3.0	156	3
Carroll	2	2.0	122	2
Belknap	8	7.9	151	6
Sullivan	3	3.0	183	3
Merrimack	8	7.9	173	8
Strafford	9	8.9	192	6
Cheshire	5	5.0	104	4
Hillsborough	33	32.7	253	13
Rockingham	28	27.7	222	23
Total	101[a]	100	207.74	70

Note. This chart specifies the number and percentage of schools from each of the 10 NH counties, the number of school districts in each county that is represented, and the average number of participants in the schools in each county.
[a]*The decreased* n *reflects the elimination of one additional school in the third part of this study.*

Table 3.4: Representativeness of Reading Schools by County, District, and Size

	Reading Schools			
County	n	%	$\bar{x}$	Districts
Coos	1	1	---[a]	1
Grafton	4	3.8	144	4
Carroll	2	1.9	97	2
Belknap	7	6.7	136	5
Sullivan	4	3.8	119	2
Merrimack	10	9.5	197	9
Strafford	10	9.5	183	7
Cheshire	6	5.7	125	4
Hillsborough	33	31.4	236	12
Rockingham	28	26.7	223	21
Total	105[b]	100	199.99	67

This chart specifies the number and percentage of schools from each of the 10 NH counties, the number of school districts in each county that is represented, and the average number of participants in the schools in each county.
[a]*This data point was withheld to protect the identity of this school*
[b]*The decreased* n *reflects the elimination of one additional school in the third part of this study.*

perform consistently across the three assessed years, did not meet the criterion for the groups based on their Fall 2012 NECAP proficiency rate, or did not perform consistently across the three assessed years.

Each of the 10 NH counties was represented in both the reading and math portions of this study, though not equally. As noted in Tables 3.3 and 3.4, more than half of the 106 reading schools, as well as more than half of the 102 math schools, were located in Hillsborough and Rockingham counties. The average school size was varied among the 10 counties with the county with the smallest mean size being 50% smaller than the county with the largest mean size. Given this disparity in average size of schools by county, a one-way ANOVA of the average number of participants per school by county was conducted for the reading schools and a second ANOVA was conducted on the math schools. Even though results are not typically presented in the third chapter, the results of these ANOVAs are presented here to provide more detailed information about the schools involved in this study. The results of these ANOVAs are provided in Table 3.5. Both ANOVAs were statistically significant ($p < .05$) indicating that the average school size varied by county. Figures 3.1 and 3.2 provide boxplots of the distribution of school size for each county plotted against the average school size for all schools in the study ($\bar{x} = 199.99$).

Analysis of the distribution of school sizes reveals positively skewed distributions wherein the overall average for reading and math, as well as several of the average school sizes for some counties were inflated by the presence of unusually large schools. These findings indicate the necessity to consider not only the dependent effects of NH county and school size on this study's results, but also the effect of the interaction between these two variables.

Table 3.5: Analysis of Variance of Average Number of Participants by County

Reading Schools (n = 105[a])					
Source	Sum of Squares	*df*	Mean Squares	*F* ratio	*p*
Between groups	184,012.856	9	20,445.873	2.112	.036*
Within groups	919,522.134	95	9679.180		
Total	1,103,534.990	104			
Math Schools (n =101[a])					
Source	Sum of Squares	*df*	Mean Squares	*F* ratio	*p*
Between groups	195,827.809	9	21758.645	2.044	.043*
Within groups	968,717.498	91	10645.247		
Total	1164545.307	100			

[a]*The decreased* n *for reading and math reflect the additional elimination of one school in the third part of this study.*
*p < .05

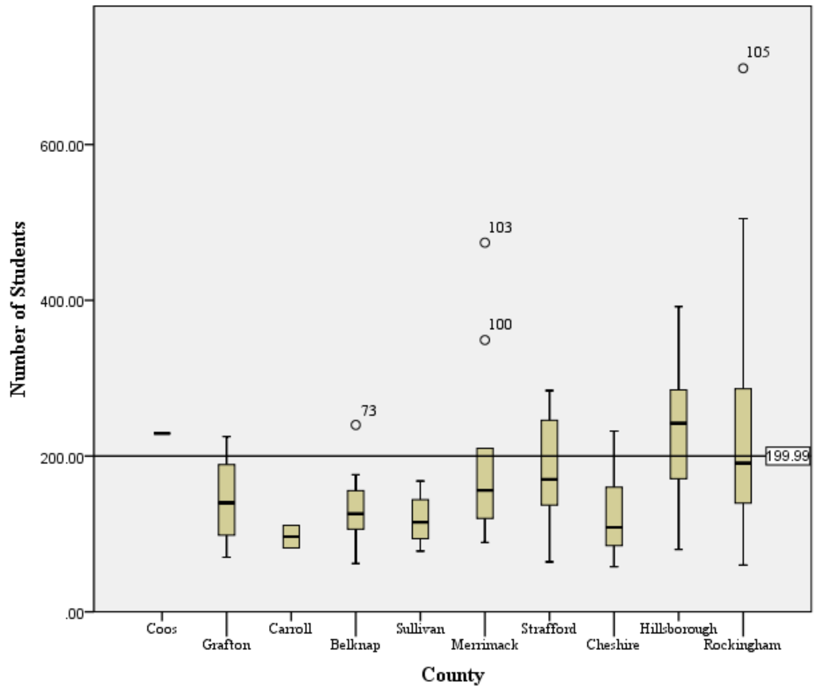

Figure 3.1: Distribution of School Size by County for Reading

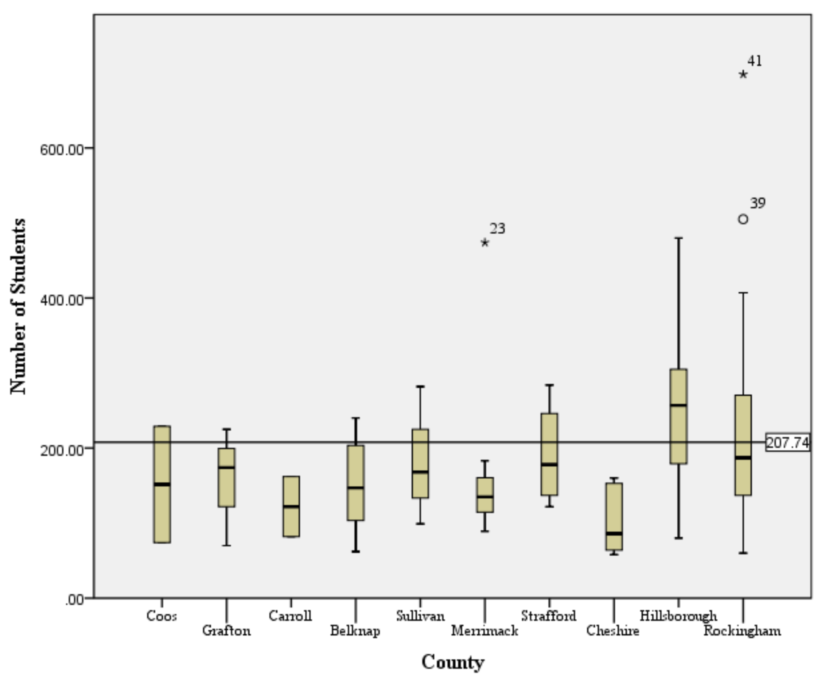

Figure 3.2: Distribution of School Size by County for Math

Part 2: Completion of the Sorting Process

The last part of the sorting process dealt with sorting of the 49 high-performing reading schools, and the 49 high-performing math schools into two groups: effective schools (ES) and high-performing low-equity schools (HPLES). The criterion for making this determination was that the 2012 performance of non-White students (NWS) within each of the HPS was comparable to the performance of the White students (WS) in that same school while the performance of students eligible for free and reduced lunch (SEFRL) was comparable to the performance of students who were not eligible for free and reduced lunch (NEFRL). This process was completed separately for reading and for math. For the purposes of simplicity, the processes that were employed will be described in depth one time, however, readers should note that the processes were conducted a total of four times: HPS for reading by eligibility for free and reduced lunch, HPS for reading by race, HPS for math schools by eligibility for free and reduced lunch, and HPS for math by race.

Independent samples *t*-tests were employed to sort the HPS into effective schools (ES) and high-performing low-equity schools (HPLES) (Howell, 1997; Keith, 2006; Minium, King, & Bear, 1993). Two comparisons were made for each of the HPS: White students (WS) versus non-White students (NWS; excluding Asian or Pacific Islander), and students eligible for free and reduced lunch (EFRL) versus students not eligible for free and reduced lunch (NEFRL). The decision to not include Asian or Pacific Islander students with other non-White students was based on the finding that these students perform much better than other non-White students (Coleman et al., 1966; NH DoE, 2012b, 2013a, 2013b, 2013c). Consequently, including them in the group for non-White students would confound the results. They were, however, included in the subgroup comparisons based on eligibility for free and reduced lunches. The subgroup comparisons based on eligibility for free and reduced lunches also included multiracial students as well as students whose race was not reported.

T-tests are dependent on three assumptions: continuous data, normality, and homogeneity of variance (Miles & Banyard, 2007). To conduct independent *t*-tests researchers must work with continuous data and have samples that are not only normally distributed, but also have similar variances. The processes utilized for this stage of the study were greatly influenced by research on the impact that non-normality (DeCarlo, 1997; Wilcox, 1998), heteroscedasticity (Wilcox, 1998), and outliers can have on the power of statistical analyses (DeCarlo, 1997; Wilcox, 1998).

Given the structure of the NECAP assessments, the range of scores for the three grade levels examined as part of this study are discrete such that third graders earned scaled scores in the 300s, fourth graders in the 400s, and fifth graders in the 500s. This data is not continuous and therefore would not permit the use of *t*-tests. Consequently, the raw scores for all third, fourth, and fifth grade students who took the 2012 NECAPs in the state of NH (N = 40,302) were used to compute the population mean and standard deviation for each grade level. These statistics were then used to convert the raw scores for students in the HPS to *z* scores (Howell, 1997; Keith, 2006; Minium et al., 1993) so that the results of students in all three grade levels could be analyzed concomitantly and as a continuous data set. All *z* scores were then converted to a standard score with a mean of 100 and a standard deviation of 15. This scale was chosen so that all scores were positive numbers as required for logistic regression analyses (Hosmer, Lemeshow, & Sturdivant, 2013), and because this standard score metric is highly familiar to this researcher.

Normality refers to the normal distribution of scores. While it is highly improbable that any sample will be truly normally distributed, researchers strive for samples that approximate the normal curve. Normality is influenced by several variables including sample size. Since the data used for this study were pre-existing, the sample sizes for the schools could not be modified. Given the variability in school sizes, prior to using *t*-tests, other tests needed to be conducted to test for normality among each of the schools.

Tests of skewness and kurtosis “have been shown to have excellent properties for detecting departures from normality” (DeCarlo, 1997, p. 296) which can affect tests of means as well as tests of variance (DeCarlo, 1997). Not surprisingly then, skewness and kurtosis can limit the power of *t*-tests thereby increasing the likelihood of Type I errors (DeCarlo, 1997). Given that kurtosis examines, in part, the tails of sample distributions, tests of kurtosis can be helpful in the detection of outliers. Given that small variations from normality can lower the power of tests (Wilcox, 1998), frequency histograms and boxplots were created concurrently with checks for skewness and kurtosis for each of the HPS. This process revealed several issues of non-normality which were addressed through the handling of outliers.

Outliers have been identified as having a significant impact on the power of statistical analyses due to the impact they have on the normality of a distribution (DeCarlo, 1997; Wilcox, 1998). For each of the HPS, boxplots were created based on race and then again based on eligibility for free and reduced lunches. When potential outliers were identified, the following Interquartile Range (IQR) formula was employed to more objectively identify outliers:

$$IQR = k(Q_3 - Q_1)$$

Typically, a *k* value of 1.5 is used when computing the IQR (Bakker & Wicherts, 201; Hoaglin & Iglewicz, 1987; Wilcox 1997). The lower limit used to identify outliers is determined by subtracting the product of *k* and the obtained IQR from the first quartile, while the upper limit is determined by adding the product of *k* and the obtained IQR to the third quartile. Scores that fall below the lower limit or above the upper limit are considered outliers.

Multiple researchers (Bakker & Wicherts, 2014; Hoaglin & Iglewicz, 1987; Wilcox, 1998) cite that using $k = 1.5$ as the cut off for identifying outliers results in the over identification and removal of outliers thereby increasing the risk of Type I errors. Hoaglin and Iglewicz (1998) found that the appropriateness of *k* is dependent on the sample size, and recommended the use of a variable value for *k* (p. 1149). They examined the ideal value of *k* for sample sizes between five and 300, but they did not provide ideal values for *k* for every number. Their recommended values of *k* were derived through complex statistical computations that resulted in a truly varied value of *k.* The variability was most prevalent for sample sizes between five and twenty wherein the specified values ranged from a low of *k* = 2.1 for samples comprised of 7, 14, 18, or 19 data points, to a high of *k* = 2.3 for samples comprised of 8, 9, 12, or 17 data points. Yet for other sample sizes between five and 20, the recommended value for *k* was 2.2 (Hoaglin & Iglewicz, 1998, p. 1149).

Based on Hoaglin and Iglewicz's (1987) research, Wilcox (1998) further examined the utility of a variable *k* for identifying outliers, and concluded that a *k* value of 2.2 is appropriate for sample sizes up to 100. Hoaglin and Iglewicz (1987) recommended *k* values for sample sizes of 200 (2.4) and 300 (2.4), but not for other sample sizes greater than 100. Neither Wilcox (1998) nor other researchers have further examined ideal values of *k* for larger sample sizes (D. C. Hoaglin, personal communication, January 10, 2017). Consequently, a variable value of *k* was adopted based on the available research, but the recommended values were extrapolated to provide values of *k* when sample sizes were between 101 and 199, and greater than 300. The values of *k* employed as part of this study are provided in Table 3.6.

Scores that fell outside of the acceptable limits, using IQR with the variable *k* values noted in Table 3.6, were considered outliers. After verifying the accuracy of each of these data points, outliers were removed. Once the outliers were removed then tests for kurtosis and skewness were repeated prior to conducting the *t*-tests. All subsequent tests of kurtosis and skewness were acceptable. Since the elimination of

Table 3.6: Variable Values of *k*

Sample Size	Value of *k*
1-100	2.2
101-199	2.3
200-399	2.4
400-599	2.5
600-799	2.6
800-999	2.7

Note. Based on the work of Hoaglin & Iglewicz (1987) & Wilcox (1998) outliers can weaken the power of a study (Bakker & Wicherts, 2014; Wilcox, 1998) the better approach would have been to conduct each analysis once with the outliers, and once without the outliers (Bakker & Wicherts, 2014; Wilcox, 1998). Given the need to complete this process twice for each of the 49 HPS for reading and twice for each of the 49 HPS for math, this added precaution was not feasible.

Heteroscedasticity, or the violation of the assumption of equal variances among samples from the same population, has been recognized as an important factor to be considered in statistical analyses given its influence on power and thereby the likelihood of performing Type I errors (Wilcox, 1998). Even though *t*-tests are dependent on the assumption of homogeneity of variances it is possible to conduct *t*-tests when heteroscedasticity is detected if the *t*-tests are modified (Miles & Banyard, 2007). When two samples have similar variances (i.e., homogeneity of variance) then their variances can be pooled, but when heteroscedasticity is detected the samples' variances should not be pooled (Miles & Banyard, 2007). Pooling the variances assumes homogeneity of variance while unpooled variances does not assume of homogeneity of variance.

For this study, Levene's test of homogeneity was employed to test for heteroscedasticity for each group within most schools. When Levene's statistic was significant, suggesting that the groups did not have similar variances, then Welch's *t*-test was employed which does not pool the variance of the two samples (Miles & Banyard, 2007). Given that Levene's *t*-test is not accurate for smaller sample sizes (n <15) or when the sample size of one group is more than 1.5 times that of the smaller group (Miles & Banyard, 2007), Welch's *t*-test was used regardless of the reported Levene's statistic.

Several HPS for reading and several for math had only one non-White student, while one HPS for reading and one for math had only one student who was eligible for free and reduced lunches. Due to these exceptionally small subsample sizes, Welch's *t*-tests could not be used. In these cases, these lone students' obtained scores were converted to *z*-scores based on the mean and standard deviation of the subsample to which they were being compared. Significant subgroup differences were recognized when students' obtained scores fell more than 1.96 standard deviations away from the mean performance of their peers. This cut off was made based on the properties of the *z* distribution in which 95% of a sample distribution will fall within 1.96 standard deviations of the sample mean (Howell, 1997) and less than .05% of a population ($p < .05$) would fall more than 1.96 standard deviations from the sample mean. These computations were then compared to Levene's *t*-test to validate the obtained findings.

One school, which was high-performing for both reading and math, did not have any non-White students. Since the criterion for inclusion in both high-performing schools' groups required testing for subgroup differences, this school could not be placed in either group and therefore was eliminated from the study. The removal of this school from the study left 48 HPS for reading and 48 HPS for math.

The results of the *t*-tests were used to complete the final sorting of schools. Schools that did not have significant *t*-tests based on race or eligibility for free and reduced lunches, thereby indicating no significant difference between the subgroups, were labeled effective schools (ES). Schools in which a significant difference was observed between subgroups according to race and/or eligibility for free and reduced lunches were labeled high-performing low-equity schools (HPLES).

Part 3: Regression Analyses

In the next phase of analyses, multinomial logistic regression models were employed to determine the proportion of the variability in the categorical outcome variable of School Effectiveness (ES, HPLES, MPS, LPS) that can be explained by the four continuous explanatory variables: percentage of White students (PWS), percentage of students eligible for free and reduced lunch (PFRL), per pupil expenditures (PPE), and percentage of teachers with a master's degree or higher (PTMDH). The effective schools group was specified as being the reference group to which the other groups were compared through the regression processes. As such the multinomial logistic regression models were employed to determine the proportion of variability in the effectiveness of schools that could be explained by PWS, PFRL, PPE, and PTMDH. Adjustments were made for school size and NH county. The processes described below were completed twice: once for reading and once for math.

Socioeconomic status was not controlled for. Where a family lives is dependent on what a family can afford (Rivkin, 2016; Schmidt, et al., 2009). Since the money available to schools is dependent on local property taxes (Brimley & Garfield, 2008; Education Trust, 2006a, 2006b; Mikesell, 2011; O'Gorman, 2010; Rivkin, 2016; Wiener & Pristopp, 2006), more wealthy towns have more financial resources available (Mikesell, 2011; O'Gorman, 2010). Also, per the effective schools research methodology, effective schools are effective for all students regardless of out-of-school variables such as family SES (Edmonds, 1979).

Since the fall NECAP assessments were designed to assess academic learning from the previous year, data for each of the four explanatory variables was obtained through the NH DoE's website based on the 2011-2012 academic year. As noted previously, student body composition was assessed via PWS and PFRL, while school funding was assessed via PPE, and teacher qualifications were assessed via PTMDH. Data related to the PWS, as well as PFRL were available for each school. In contrast, data related to PPE and the PTMDH were only available at the district level. Since, as depicted in Tables 3.3 and 3.4, several school districts had more than one school included in the current study, the data points for PPE and PTMDH were not independent for schools within the same school district. This lack of independence among data points decreased the power of the current study.

The four explanatory variables and the four groups were plotted and analyzed for outliers. Given that the intended purpose behind outlier identification is to ensure that the data points used in a study are valid (Bakker & Wicherts, 2014; Hoaglin & Iglewicz, 1987; Wilcox, 1998), outliers were identified, but not automatically removed. For each outlier, this researcher returned to the data provided by the NH DoE to validate the data point. All outliers were validated and therefore were retained.

As noted previously, School Size and NH County were identified as being potential confounding variables. Given the variability among the schools included in this study, in regards to the number of students who took the NECAP assessments, as well as disproportionate number of schools coming from the ten NH counties, the effects of these variables were considered. In order to accurately determine the influence of the confounding variables on the explanatory value of each of the explanatory variables, each of the explanatory variables was first entered alone into a logistic regression model (Hosmer et al., 2013). Then logistic regressions were used to determine the individual and interaction effects of School Size and NH County, on the outcome variable. Once the individual effects of these variables (i.e., School Size, NH County, School Size * NH County) were determined, then all that were significant were forced into a full adjustment model. Then each of the individual explanatory variables was individually forced into logistic models with the full adjustment model. A comparison was then made between the models containing only the explanatory variables, and the models containing the explanatory variables and the full adjustment model to determine if the adjustment model truly strengthened the explanatory value of the models (Hosmer et al., 2013).

Then tests for collinearity were employed to determine if any of the explanatory variables were significantly correlated (Foster, Barkus, & Yacorsky, 2006; Jaccard, 2001; Menard, 2002). When correlations between any two or more variables were determined to be significant ($p < 0.05$), then the variable that had the strongest predictive power, as determined by the individual logistic regression models, was retained while the others were omitted from the final logistic regression model (Foster et al., 2006; Jaccard, 2001). The intention was to use the results of the initial logistic regression models, and the tests for collinearity, to develop a full model with the best explanatory effect on the outcome variable. The adjustment model would be forced into the model, while the explanatory variables would be entered via a forward entry process wherein the strongest explanatory variable would be entered first, then subsequent explanatory variables would be entered in the order of their strength until the addition of explanatory variables failed to strengthen the explanatory value of the full model (Foster et al., 2006).

The purpose of multinomial logistic regression analysis is to construct a model that best accounts for the outcome variable (Hosmer et al., 2013). This is accomplished by comparing the effectiveness of various models at explaining the outcome variable. For each regression analysis, several tests were employed. Likelihood ratio tests, which are based on chi square tests, were used to determine the significance of each model (Foster et al., 2006; Hosmer et al., 2013; Menard, 2002; Pampel, 2000). These tests determine the extent to which the proposed model can better predict the outcome variable than an empty model; a model that contains none of the explanatory or confounding variables.

Pseudo R-squared was used to assess the predictive value of the logistic models on the outcome value (Institute for Digital Research and Education, 2011). Unlike *R-squared*, which is commonly used in general linear models to assess the amount of variance in the outcome variable that can be attributed to the regression model, *pseudo R-squared* has more limited utility (Hosmer et al., 2013). *Pseudo R-squared* estimates the predictive value of the model relative to other models that are designed to predict the same outcome variable (Institute for Digital Research and Education, 2011). As such *pseudo R-squared* is useful for comparing various models in order to determine which model has the best explanatory value, but should not be used as an overall estimate of the effects of a predictive model (Institute for Digital Research and Education, 2011).

Whereas the likelihood ratio tests and *pseudo R-squared* assess the strength of the model as a whole, Wald tests assess the strength of individual components within the logistic regression models (Hosmer et al., 2013). More specifically, Wald tests assess the strength of explanatory variables to differentiate between different levels of the outcome variable. Typically when Wald tests are significant, it suggests that changes in the value of that specific component of the model has meaningful effects on the outcome variable (Hosmer et al., 2013). For the current study, the outcome variable was School Effectiveness, and the referent group was ES. In other words, the Wald tests would determine if differences in the values of the explanatory variables could explain group membership. As such for each logistic regression model, the Wald test assessed if changes in the values of the explanatory variables could account for differences between HPLES and ES, between MPS and ES, and between LPS and ES. Even though Wald tests signify that changes in explanatory variables influence the outcome variable, they have some limitations in their sensitivity based on the quantity and density of the sample distribution (Hosmer et al., 2013). Consequently, for the current study the interpretation of the Wald statistic will be limited to only noting if it is statistically significant or not. When one of the Wald tests for a logistic regression model is significant then each of the associated odds ratios will be interpreted to provide more insight into the influence of the explanatory variable on the outcome variable (Hosmer et al., 2013).

Odds ratios were used to quantify the impact that changes in the values of explanatory variables had on the outcome variable (Hosmer et al., 2013). As explained by Hosmer et al., (2013), “The odds ratio is widely used as a measure of association as it approximates how much more likely or unlikely (in terms of odds) it is for the outcome to be present" (p. 52). Interpretation of odds ratios provides information about the strength and nature of the influence that the explanatory variables have on the outcome variable. For each outcome group to which the referent group is compared, odds ratios are computed. Consequently, for the current study odds ratios were obtained comparing the odds of a school being a HPLES versus the odds of being an ES, the odds of being a MPS versus the odds of being an ES, and the odds of being a LPS versus the odds of being an ES. An odds ratio of 1.0 suggests an equal likelihood. Odds ratios greater than 1.0 indicate a higher likelihood, or an increased odds, of obtaining an outcome other than the referent group. For the current study, odds ratios greater than 1.0 suggest that given the changes in the explanatory variable, it is more likely that schools would belong to a group other than the effective schools group. Odds ratios less than 1.0 indicate a lower likelihood, or decreased odds, of obtaining an outcome other than the referent group; meaning the greater the odds that the school would be effective. Odds ratios less than 1.0 can also be interpreted as meaning that the relative risk of an outcome is decreased. For the current study, odds ratios below 1.0 indicate the decreased risk of a school belonging to a group other than the effective schools group. Again, meaning the greater the odds that the school would be effective.

Chapter 4: Findings and Data Analysis

A statewide assessment of the effectiveness of NH elementary schools was conducted using a combination of aggregated and disaggregated data for third, fourth, and fifth grade students who completed the Reading and Math NECAP assessments in 2010, 2011, and 2012. The purpose of the current study was to determine the percentage of NH elementary schools that were effective in reading, the percentage of elementary schools that were effective in math, and the extent to which student body composition, school funding practices, and teacher qualifications can explain the effectiveness of schools in reading and math. All analyses were conducted separately for reading and math.

In the first part of this study, 210 schools were sorted into three preliminary groups for reading and for math. The schools were first sorted based on their 2012 proficiency rate: the percentage of students who demonstrated at least proficient abilities on the 2012 NECAPs. Then schools were sorted based on the overall rank of the school, relative to other NH elementary schools, in 2010 and 2011. High-performing schools (HPS) were schools wherein at least a minimal percentage of students demonstrated at least proficient abilities in 2012 (82% for Reading; 80% for Math), and in the top third of NH elementary schools in 2010 and 2011. Moderately-performing schools (MPS) performed in the middle 20% of NH elementary schools in 2012, and in the middle third of NH elementary schools in the preceding two years. Low-performing schools (LPS) fell in the lowest 20% of schools in 2012, and in the bottom third of NH schools in 2010 and 2011. Given the strict requirements for categorizing schools, not all schools that were included in this study met the criteria for inclusion in one of these three preliminary groups. This sorting process was described in Chapter 3 and the results thereof are presented in Tables 3.1 and 3.2.

In the second part of this study, HPS were divided into two groups: effective schools (ES) and high-performing low-equity schools (HPLES). According to the effective schools research, to be considered effective, a school must be effective for all students regardless of students' race or SES. As such, for each of the HPS for reading, as well as each of the HPS for math, subgroup comparisons were made based on race (WS vs. NWS) and SES (EFRL vs. NEFRL). The subsample sizes were markedly uneven though with the larger subsample size consistently more than 1.5 times larger than the smaller subsample size. Consequently, Welch's *t*-tests were the predominant metric for assessing subgroup differences except when a subsample size was exceptionally small ($n = 1$).

Several HPS for reading and several HPS for math had only one non-White student, while one HPS for reading and one for math had only one student who was eligible for free and reduced lunches. Due to the exceptionally small sub-sample sizes, Welch's *t*-tests could not be used. In these cases, these lone students' scores were converted to *z*-scores based on the mean and standard deviation of the subsample to which they were being compared. Significant subgroup differences were recognized when students' obtained scores fell more than 1.96 standard deviations away from the mean performance of their peers. This criterion was selected based on the properties of the *z* score distribution which states that the probability of a score falling more than 1.96

standard deviations away from the mean of a sample distribution is $p < .05$ (Howell, 1997). Levene's *t*-test was then used to validate the obtained findings.

One school, which was considered high-performing for both reading and math, did not have any non-White students thereby preventing the analysis of subgroup differences based on race. Since this comparison could not be made, this school could not be categorized as being effective. However, there was no evidence to conclude that this school was not effective. Consequently, this one school was fully eliminated from the study thereby reducing the total number of schools in the study by one ($n = 209$), the number of HPS for reading by one ($n = 48$), and the number of HPS for math by one ($n = 48$).

In the final part of this study, multinomial logistic regression analyses were conducted to determine to what extent, if at all, student body composition, school funding practices, and teacher qualifications influenced the effectiveness of schools for reading and for math. Data pertaining to the percentage of White Students (PWS) and the percentage of students eligible for free and reduced lunch (PFRL) was obtained for each school from the online NH DoE data system to assess student body composition. Data pertaining to per pupil expenditures (PPE), which was used to assess school funding practices, and the percentage of teachers with a master's degree or higher (PTMDH), which were used to assess teacher qualifications, were obtained from the NH DoE online data system for the district in which each school resided. Adjustment models were built to control for NH County, School Size, and the interaction between NH County and School Size. Once the individual explanatory value of each of the explanatory variables was understood, then tests for collinearity ensued with the intent of building a more complete model that contained multiple explanatory variables. These analyses were conducted once for reading and once for math.

Percentage of NH Elementary Schools Effective in Reading

The first research question examined the percentage of NH elementary schools that were effective in reading. The first step to answering this question required sorting the 210 schools in the reading portion of this study into four groups. As depicted in Table 3.1, 49 were HPS, 19 were MPS, and 38 were LPS. The other 104 schools did not fully meet the criteria for inclusion in one of the three preliminary groups. Since these schools did not qualify as being HPS, they likewise could not be considered effective. These schools were considered in the computation of the percentage of effective schools for reading, but were not part of the regression analyses.

The next step to answer the first research question involved sorting the 49 HPS for reading into two groups: effective schools (ES) and high-performing low-equity schools (HPLES). This sorting was dependent on tests to assess for subgroup differences. To be effective, a school could not have any subgroup differences based on race or SES as measured by students' eligibility for free and reduced lunches. One of the HPS for reading did not have any non-White students and therefore was eliminated from the study leaving 209 schools, and 48 HPS for reading. Seven of the remaining 48 HPS for reading only had one non-White student, while one had only one student who was eligible for free and reduced lunches. The existence of these lone students complicated the analysis of subgroup differences since *t*-tests could not be used. To address this complication these lone students were compared to their peers' using *z*-score comparisons. Each of the non-White students in these schools performed within 0.8 standard deviations of their White peers (z = -0.68, -0.25, -0.23, 0.12, 0.13, 0.73, 0.79). The sole student eligible for free and reduced lunches in one of the HPS, performed within one standard deviation of the mean of his/her peers ($z = -0.94$). These findings indicate that these lone students performed comparably to their peers and, as such, subgroup differences did not exist in their schools. *T*-tests were used to determine the existence of subgroup differences in the remaining high-performing reading schools.

Thirty of the 48 HPS for reading exhibited significant subgroup differences ($p \leq .05$) on the basis of race, and/or SES and were therefore classified as being HPLES. Significant subgroup differences were observed based on race alone in two schools, based on SES alone in 28 schools, and based on race and SES in three schools. The significance levels for the observed subgroup differences based on race ranged from a low of $p < .001$ to a high of $p = .03$. The significance level of the observed subgroup differences based on SES ranged from a low of $p < .001$ to a high of $p = 0.049$.

The 18 HPS that did not exhibit subgroup differences based on race or SES were classified as being effective schools (ES) for reading. These results suggest that 8.6% of the 209 NH elementary schools that were included in this study were effective in reading as of the fall of 2012. As noted in Table 4.1, seven of the ten NH counties were home to at least one ES for reading. Six of the ES for reading were located in Rockingham County. Grafton, Hillsborough, and Stafford counties were each home to three effective

Table 4.1: Representativeness of Effective Reading Schools by County

County	*n*	%
Coos	0	0
Grafton	3	16.7
Carroll	1	5.5
Belknap	0	0
Sullivan	1	5.5
Merrimack	1	5.5
Strafford	3	16.7
Cheshire	0	0
Hillsborough	3	16.7
Rockingham	6	33
Total	18	100

Note. n *denotes the number of effective reading schools in each county*

schools while the three remaining effective schools were located in Carroll, Sullivan, and Merrimack counties.

Percentage of NH Elementary Schools Effective in Math

Due to the similarities in the first and second research questions, a similar approach was employed to answer the second research question: What percentage of NH elementary schools are effective in math? As depicted in Table 3.2, of the 210 schools in the math portion of this study, 49 were HPS, 17 were MPS, and 36 were LPS. The remaining 108 schools did not fully meet the criteria for inclusion in one of the three preliminary groups. These schools were considered in the computation of the percentage of effective schools for math, but were not included in the subsequent regression analyses. One of the 49 HPS for math did not have any non-White students and was therefore eliminated from the study reducing the total number of schools in the study to 209, and the number of HPS to 48.

Five HPS for math had only one student of color. Four of these students' scores fell within 1.96 standard deviations of their White peers (z = -1.52, -0.73, -0.03, 0.82). The fifth student performed more than two standard deviations below the mean of his/her White peers (z = -2.05, p = .044). One of the HPS for math had only one student who was eligible for free and reduced lunches. This student's performance fell within 0.5 standard deviations of his/her peers (z = -0.45).

Significant subgroup differences ($p \leq .05$) were observed based on race and/or SES in 29 of the 48 HPS leading to their designation as being HPLES. Nine of these HPLES demonstrated significant subgroup differences based on race alone, 27 demonstrated significant subgroup differences based on SES alone, and seven schools demonstrated significant subgroup differences based on race and SES. The significance level for the observed subgroup differences based on race, ranged from a low of p = .002 to a high of p = .044. The significance of the observed subgroup differences based on SES ranged from a low of p < .001 to a high of p = .042.

Nineteen schools did not exhibit subgroup differences based on race or SES. This finding suggests that as of the fall of 2012, 9% of the 209 elementary schools that were included in this study were effective schools (ES) for math. As noted in Table 4.2, eight of the ten NH counties were home to at least one effective math school. Eight of the effective schools were located in Rockingham County. Hillsborough was home to three of these effective math schools. Belknap County and Stafford County each had two. The remaining four schools were equally dispersed among Carroll, Cheshire, Grafton, and Sullivan counties.

Table 4.2: Representativeness of Effective Math Schools by County

County	*n*	%
Coos	0	0
Grafton	1	5.3
Carroll	1	5.3
Belknap	2	10.5
Sullivan	1	5.3
Merrimack	0	0
Strafford	2	10.5
Cheshire	1	5.3
Hillsborough	3	15.8
Rockingham	8	42.1
Total	29	100

Note. n *denotes the number of effective reading schools in each county*

Explanatory Variables on Outcome Variable for Reading

The third research question examined the extent to which student body composition, per pupil expenditures, and teacher qualifications explain a school's effectiveness in reading. As noted previously, student body composition was assessed via two explanatory variables: percentage of White Students (PWS) and percentage of students eligible for free and reduced lunch (PFRL); and teacher qualifications were assessed by the percentage of teachers with a master's degree or higher (PTMDH). To answer this question a series of logistic regression models were constructed to test for the most complete model; the model that could account for the greatest proportion of the variability in the effectiveness of schools for reading that could be explained by the explanatory variables.

Given the potential confounding influence of NH County and School Size on the logistic regression analyses, an adjustment model was created for reading. The adjustment model for reading was constructed by performing four multinomial logistic regression analyses which are summarized in Table 4.3. In the first model, the outcome

Table 4.3: Adjustment Model Reading

Model	Likelihood ratio χ^2	*df*	*p* value	Cox & Snell Pseudo R^2
1. NH County	46.82	27	.010	.36
2. School Size	3.97	3	.265	.04
3. NH County * School Size	47.54	30	.022	.36
4. NH County x (NH County * School Size)	87.63	54	.003	.57
Notes. Outcome Variable: School Effectiveness; Referent Group: Effective Schools				

variable, School Effectiveness, was regressed on NH County revealing a statistically significant likelihood ratio (p = .010, df = 27, *pseudo* R^2 = .36). The obtained *pseudo R squared* of .36 indicates that 36% of the variability in the effectiveness of schools could be explained by the county in which the school resides. For the second model, School Effectiveness was regressed on School Size yielding a non-significant likelihood ratio (p = .265, df = 3, *pseudo* R^2 = .04) suggesting that School Size was no better at explaining the effectiveness of schools than was an empty model: a model that contained none of the confounding or explanatory variables. In other words, the size of the school had no significant effect on explaining the effectiveness of the schools in this study. The third model, which regressed School Effectiveness on the interaction effect between NH County and School Size (NH County * School Size), was significant ($p < .022$, df = 30, *pseudo* R^2 = .36). The obtained *pseudo R squared* of .36 is commensurate with the obtained *pseudo R squared* for the first model and suggests that 36% of the variability in the effectiveness of schools can be explained by the interaction between NH County and School Size. A full adjustment model was then constructed for reading. Given that School Size did not have a significant effect on School Effectiveness for reading, it was not including in the adjustment model. The full adjustment model for reading controlled for the effects of NH County, as well as the interaction between NH County and School

Size. This model was significant (*p* = .003, *df* = 54, *pseudo* R^2 = .57). These results indicate that more than half (57%) of the variability in the effectiveness of schools can be explained by the county in which the school was located as well as the interaction between the NH County where the school was located, and the school's size.

To test for the effects of each of the explanatory variables, each was entered into a logistic regression model alone, then with the adjustment model in order to ensure that the adjustment models truly were influencing the effects of the explanatory variables. Each of the likelihood ratios for the explanatory variables was statistically significant ($p < .001$) suggesting that each was better at explaining the effectiveness of schools for reading than was an empty model: a model that contains neither explanatory nor confounding variables. The strength of each of the explanatory variables was enhanced by the inclusion of the adjustment model meaning that each of the explanatory variables better explained the effectiveness of schools when NH County and the interaction between NH County and School Size were considered. Since the strength of the models was enhanced by the inclusion of the adjustment model, the models that included the individual explanatory variables with the adjustment model, were the focus of the subsequent discussion; the results of the models containing only the explanatory variables are provided in Table 4.4, but will not be discussed below.

First, School Effectiveness was regressed on PWS, with the adjustment model, revealing a statistically significant likelihood ratio ($p < .001$, *df* = 57, *pseudo* R^2 = .69). These results indicate that the PWS was more effective than an empty model: a model containing none of the explanatory variables or confounding variables, at explaining the

Table 4.4: Multinomial Logistic Regression Models for Reading

Model	Likelihood ratio χ^2	*df*	*p* value	Cox & Snell Pseudo R^2
1. PWS	36.28	3	< .001	.29
2. PWS with Adjustment Model	124.28	57	< .001	.69
3. PFRL	112.14	3	< .001	.66
4. PFRL with Adjustment Model	194.53	57	< .001	.84
5. PPE	22.22	3	< .001	.19
6. PPE with Adjustment Model	111.47	57	< .001	.65
7. PTMDH	36.03	3	< .001	.29
8. PTMDH with Adjustment Model	109.62	57	< .001	.65

*Notes. Outcome Variable: School Effectiveness; Referent Group: Effective Schools; Adjustment Model: NH County x (NH County * School Size)*

variability in the effectiveness of schools. The obtained *pseudo R squared* suggests that this model explains 69% of the variability in the effectiveness of schools. As noted in Table 4.5, the Wald statistic, which tests for the significance of explanatory variables in a statistical model, was significant for LPS. The corresponding odds ratio of 0.71 suggests that as the PWS increased by 1%, the relative risk of a school being low-performing decreased by 28.8%. The Wald statistic was not significant for MPS or HPLES, however, interpretation of those odds ratios provides more insight into the explanatory effect of the PWS on the outcome variable. As the PWS increased by 1%, the relative risk of being a MPS decreased by 13.5%, while the relative risk of being a HPLES decreased by 6.6%. These results indicate that as the percentage of White students (PWS) in a school increases by 1%, schools are significantly more likely to be effective in reading.

Table 4.5: Effective Reading Schools on % White Students with Adjustment Model

	Wald	*p*	*OR*
High-Performing Low Equity Schools	0.68	.41	0.93
Moderately Performing Schools	2.52	.11	0.87
Low Performing Schools	11.02	.001**	0.71

Notes. Referent Group: Effective Schools
** $p < .05$; ** $p < .01$*

Table 4.6: Effective Reading Schools on % Eligible for Free Reduced Lunches with Adjustment Model

	Wald	*p*	*OR*
High-Performing Low Equity Schools	6.17	.013*	1.78
Moderately Performing Schools	8.03	.005**	1.95
Low Performing Schools	14.74	<.001**	3.38

Notes. Referent Group: Effective Schools
$* p < .05$; $** p < .01$

When the outcome variable, School Effectiveness, was regressed on the PFRL, with the adjustment model, another significant likelihood ratio was observed ($p < .001$, $df = 57$, *pseudo* $R^2 = .84$) indicating that PFRL with the adjustment model was significantly better at explaining the effectiveness of a school than was an empty model. The obtained *pseudo R squared* of .84 indicates that 84% of the variability in the effectiveness of schools could be explained by this model. The Wald statistics as reported in Table 4.6 were significant for LPS, MPS, and HPLES. The odds ratios suggest that as the PFRL increased by 1%, schools were significantly less likely to be effective. Further interpretation reveals that as PFRL increased by 1% schools were 3.38 times more likely to be LPS than ES, 1.95 times more likely to be MPS than ES, and 1.78 times more likely to be HPLES than ES. These findings indicate that as the percentage of low-income students increases in a school, the less likely the school will be effective.

Regression of School Effectiveness on PPE, with the adjustment model, revealed a statistically significant finding ($p < .001$, $df =57$, *pseudo* $R^2 = .65$) suggesting that PPE significantly influenced the effectiveness of schools for reading. The obtained *pseudo R squared* suggests that this model explains 65% of the variability in the effectiveness of schools. The Wald statistic was significant for LPS ($W = 6.26$, $p < .001$, $OR = 0.99$). The corresponding odds ratio of 0.99 indicates that as PPE increased by $1, the relative risk of being a LPS decreased by 1%. To get a more realistic understanding of these impacts, the unit of measurement for PPE was converted from $1 to $100. The resulting likelihood ratio and corresponding results were consistent with the original model (p < .001, df =57, *pseudo* R^2 = .65; W = 6.26, p < .001, OR = 0.93), while the odds ratio decreased to 0.93. This converted odds ratio suggests that for every $100 increase in PPE, the relative risk of a school being a LPS decreased by 7%. The Wald statistics were not significant for MPS or for HPLES (see Table 4.7). The odds ratios for MPS and HPLES suggest that increases in PPE by $100 had no discernible impact on decreasing the relative risk of a school being a moderately-performing school or a high-performing low-equity school These results suggest that while increasing the PPE by $100 decreased the likelihood that a school would be a LPS, it did not decrease the odds that schools would be a moderately-performing or high-performing low-equity school. The pattern of these odds ratios suggests that there might be a ceiling point for the effects of PPE on the effectiveness of schools for reading. Up to this ceiling, increases in PPE will decrease the relative risk of a school being a LPS, but beyond that ceiling, increases in PPE will not decrease the risk of a school being either a MPS or a HPLES.

Table 4.7: Effective Reading Schools on Per Pupil Expenditures w/ Adjustment Model

	Wald	*p*	*OR*
High-Performing Low Equity Schools	.268	.605	1.01
Moderately Performing Schools	1.64	.201	1.03
Low Performing Schools	6.26	.012*	0.93

Notes. Referent Group: Effective Schools
** p < .05; ** p < .01*

Table 4.8: Effective Reading Schools on % Teachers w/ Master's Degree or Higher with Adjustment Model

	Wald	*p*	*OR*
High-Performing Low Equity Schools	1.06	.304	0.97
Moderately Performing Schools	2.57	.109	0.94
Low Performing Schools	13.38	<.001**	0.84

Notes. Referent Group: Effective Schools
** p < .05; ** p < .01*

Regression of School Effectiveness on PTMDH, with the adjustment model, revealed a statistically significant finding ($p < .001$, $df = 57$, *pseudo* $R^2 = .65$) suggesting that this model was significantly better at explaining the effectiveness of schools than was an empty model. The obtained *Pseudo R squared* indicates that 65% of the variability in the effectiveness of schools could be explained by the PTMDH. The Wald statistic was significant for LPS (see Table 4.8). The corresponding odds ratio indicates that as the PTMDH increased by 1%, the relative risk of a school being a LPS decreased by 16%. The Wald statistics were not significant for MPS or HPLES. Nevertheless, as the PTMDH increased by 1%, the relative risk of a school being a MPS decreased by 6.3% while the relative risk of a school being a HPLES decreased by 3.5%. These results suggest that the more qualified the teachers are in a school, the better the odds that the school will be effective.

The purpose of the multinomial logistic regression analysis is to build the regression model that best explains the variability in the outcome variable (Hosmer, 2013; Pampel, 2000). For the current study the purpose of the multinomial logistic regression analysis is to develop a model that best explains the variability in School Effectiveness. The construction of this model should include as many of the explanatory variables as possible, however, when explanatory variables are correlated they cannot both be included (Foster et al., 2006; Jaccard, 2001). As such, it is important to first identify the individual explanatory power of each of the models, then to test for collinearity. Given that each of the explanatory models was significant ($p < .001$), *pseudo R squared* was employed to identify which of the models had the best explanatory effect on the outcome variable. As noted above, the model containing PPE with the adjustment model, and PTMDH with the adjustment model could each explain 65% of the variability in the effectiveness of schools. The model containing PWS with the adjustment model was slightly stronger as it could explain 69% of the variability in the effectiveness of schools. The model containing PFRL with the adjustment model was found to account for 84% of the variability in the effectiveness of schools which exceeds the explanatory effects of the other explanatory variables revealing that PFRL better explains the effectiveness of schools in reading than do the PWS, PPE, and PTMDH. Therefore, the full model should include at least PFRL since it had the best explanatory effect on School Effectiveness. The inclusion of other explanatory variables is dependent on tests to ensure that they do not covary with PFRL.

Table 4.9: Correlational Matrix for Reading Explanatory Variables

Explanatory Variables	1	2	3	4
1 PWS	---	-.70**	.47**	.17
2 PFRL	-.70**	---	-.24*	-.51**
3 PPE	.47**	-.24*	---	.31**
4 PTMDH	.17	-.51**	.31**	---

Notes. n = *105*
** p < .05, two-tailed.   ** p < .01, two-tailed.*

Collinearity among the explanatory variables was assessed using tests of correlation. The correlational coefficient for each pair of explanatory variables was assessed using 2-tail tests of significance and are reported in Table 4.9. The obtained results, indicate that the PFRL was significantly correlated with each of the three other explanatory variables. The strongest correlation was observed among the two student body composition variables: PWS and PFRL (r = -.70, p < .001). Figure 4.1 provides a scatterplot of the correlation between these two variables. The obtained correlational coefficient and scatterplot suggest that as the percentage of White students (PWS) in a school increases, the percentage of students who are eligible for free and reduced lunches (PFRL) decreases. The scatterplot reveals that the lower right quadrant, depicting the schools with the lowest percentage of White Students and the highest percentage of low-income students, contains only low-performing schools. Though small, the correlation between the PFRL and PPE as depicted in Figure 4.2, was significant (r = -.24, p = .01) suggesting that as the percentage of students who are eligible for free and reduced lunches (PFRL) in a school increases, the per pupil expenditures (PPE) decrease. A moderately negative correlation was observed between PFRL and PTMDH (r = -.51, p < .001) suggesting that as the percentage of students who are eligible for free and reduced lunches (PFRL) increases, the percentage of teachers with a master’s degree or higher (PTMDH) in the district decreases (see Figure 4.3). Review of Figures 4.2 and 4.3 reveals the existence of horizontal lines in the data points which are the results of the reliance on district level data for the reporting of per pupil expenditures the percentage of teachers with a master’s degree or higher. Given that PFRL is a measure of low-income students,

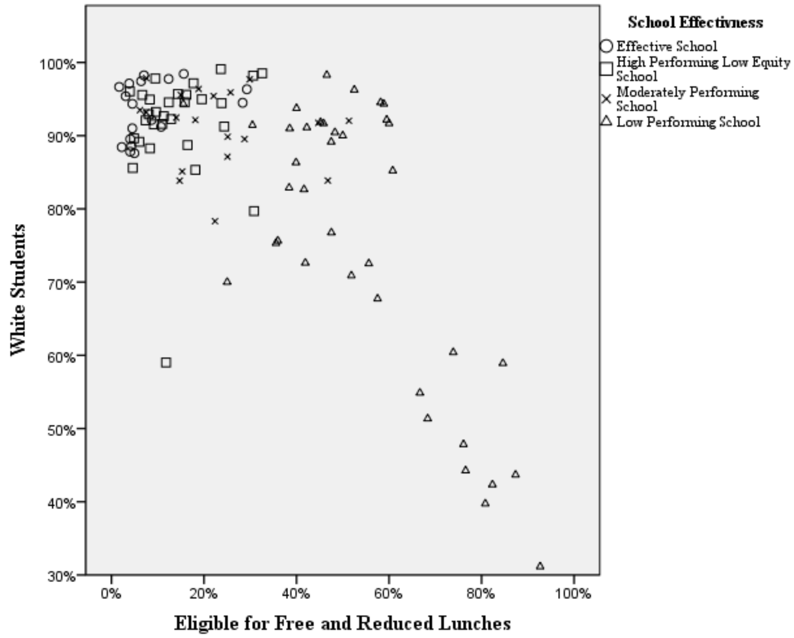

Figure 4.1: Scatterplot Percentage of White Students vs. Percentage of Eligibility for Free and Reduced Lunches

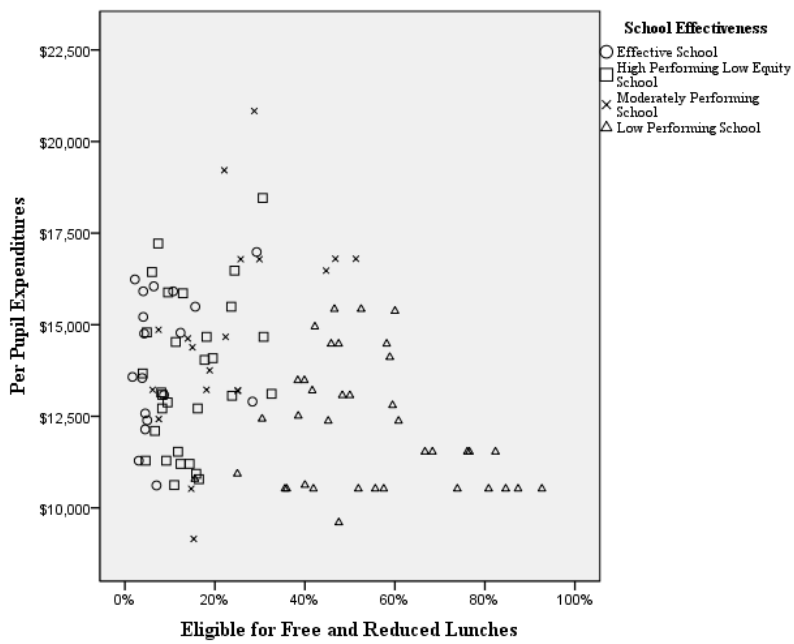

Figure 4.2: Scatterplot Per Pupil Expenditures vs. Percent Eligible for Free and Reduced Lunches

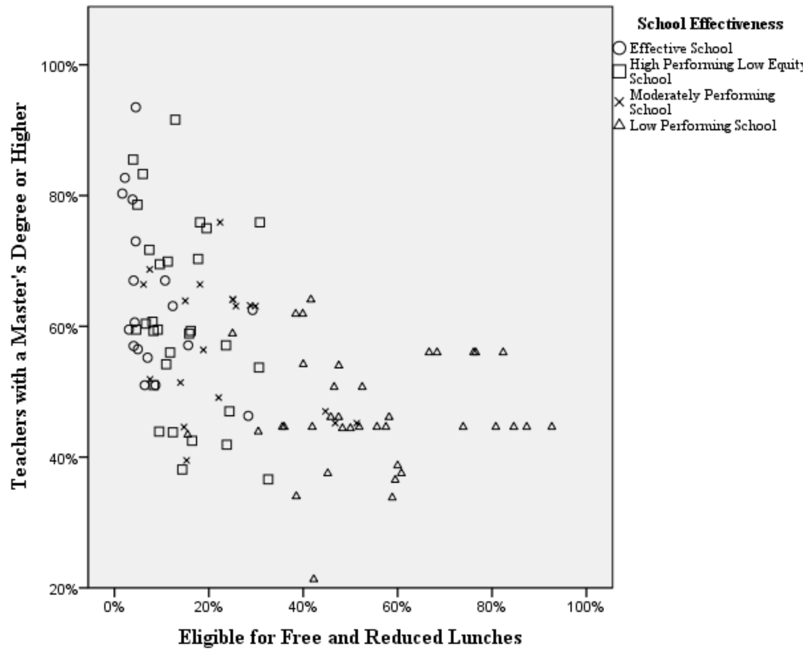

Figure 4.3: Scatterplot Percentage of Teachers with a Master's Degree or Higher vs. Percent Eligible for Free and Reduced Lunches

the obtained correlations suggest that as the percentage of low-income students in a school increases, the percentage of White students, the per pupil expenditures, and the percentage of teachers with a master’s degree or higher, all decrease.

Since PFRL was identified as having the strongest explanatory effect on the outcome variable, and because it was significantly correlated with each of the other explanatory variables, a more complete model that assessed the combined influence of PFRL with the other explanatory variables could not be assembled for reading. As such, the best model for reading included only the PFRL with the adjustment model.

Explanatory Variables on Outcome Variable for Math

The fourth research question examined the extent to which study body composition, per pupil expenditures, and teacher qualifications explain a school’s effectiveness in math. As with the reading portion of this study, an adjustment model was created to control for the effects of NH County, School Size, and the interaction between these two variables. Prior to including the explanatory variables with the adjustment model, each was entered into its own logistic regression model to identify the individual influence of each of the explanatory variables with and without the adjustment model (Hosmer et al., 2013). The collinearity of the explanatory variables was examined. Then the best model was identified.

The adjustment model for math was constructed in a similar manner as was the adjustment model for reading. Four models were constructed. As noted in Table 4.10, the models containing NH County, School Size, and the interaction between NH County and

Table 4.10: Adjustment Model Math

Model	Likelihood ratio χ^2	*df*	*p* value	Cox & Snell Pseudo R^2
1. NH County	49.44	27	.005	.39
2. School Size	15.93	3	.001	.15
3. NH County * School Size	71.66	30	< .001	.51
4. NH County x School Size x (NH County * School Size)	113.96	57	< .001	.68

Notes. Outcome Variable: Group; Referent Group: Effective Schools

Table 4.11: Multinomial Logistic Regression Models for Math

Model	Likelihood ratio χ^2	*df*	*p* value	Cox & Snell Pseudo R^2
1. PWS	31.52	3	< .001	.27
2. PWS with Adjustment Model	131.18	60	< .001	.73
3. PFRL	101.17	3	< .001	.63
4. PFRL with Adjustment Model	200.63	60	< .001	.86
5. PPE	18.68	3	< .001	.17
6. PPE with Adjustment Model	125.03	60	< .001	.71
7. PTMDH	33.80	3	< .001	.28
8. PTMDH with Adjustment Model	127.49	60	< .001	.72

*Notes. Outcome Variable: Group; Referent Group: Effective Schools; Adjustment Model: NH County x School Size x (NH County * School Size)*

School Size, alone, were each clinically significant suggesting that they each were better at explaining the variability in the effectiveness of schools than was an empty model. Consequently, each was entered into the full adjustment model. The obtained *pseudo R squared* of .68, as reported in Table 4.10, suggests that 68% of the variability in School Effectiveness could be explained by the full adjustment model for math.

As noted in Table 4.11, regression of the outcome variable on the PWS, with the adjustment model, revealed a statistically significant likelihood ratio ($p < .001$, $df = 60$, *pseudo* $R^2 = .73$). These results indicate that this model explained 73% of the variability in the effectiveness of schools for math. The Wald statistic was significant for LPS (see Table 4.12) and for HPLES, but not for MPS. The obtained odds ratios suggest that for every 1% increase in the PWS, the relative risk of being a LPS decreased by 22%, the relative risk of being a MPS decreased by 11%, and the relative risk of being a HPLES decreased by 12%. These findings suggest that as the PWS in a school increases so do the odds that a school will be effective.

When School Effectiveness, was regressed on the PFRL with the adjustment model, a significant likelihood ratio was observed ($p < .001$, $df = 60$, *pseudo* $R^2 = .86$)

Table 4.12: Effective Math Schools on % White Students with Adjustment Model

	Wald	*p*	*OR*
High-Performing Low Equity Schools	4.43	.035*	0.88
Moderately Performing Schools	3.46	.063	0.89
Low Performing Schools	13.74	<.001**	0.78

Notes. Referent Group: Effective Schools
** p < .05; ** p < .01*

Table 4.13: Effective Math Schools on % Eligible for Free Reduced Lunches with Adjustment Model

	Wald	*p*	*OR*
High-Performing Low Equity Schools	6.2	.013*	7.48
Moderately Performing Schools	6.97	.008**	8.46
Low Performing Schools	8.04	.005**	9.92

Notes. Referent Group: Effective Schools
** $p < .05$; ** $p < .01$*

This model accounted for 86% of the variability in the effectiveness of schools. As noted in Table 4.13, the Wald statistics were significant for LPS, MPS, and HPLES. The odds ratios suggest that as the PFRL increased by 1%, schools were less likely to be effective. Further interpretation reveals that as PFRL increased by 1%, schools were 9.92 times more likely to be LPS than ES, 8.46 times more likely to be MPS than ES, and 7.48 times more likely to be HPLES than ES. These results suggest that as the PFRL increases, the odds that a school will be effective decreases.

Regression of School Effectiveness on the PPE, with the adjustment model, revealed a statistically significant finding ($p < .001$, $df = 60$, *pseudo* $R^2 = .71$). These findings indicate that PPE significantly influenced the effectiveness of schools in mathematics and could account for 71% of the variability in the effectiveness of schools. The Wald statistic was significant for LPS ($W = 12.39$, $p < .001$, $OR = 0.99$). The odds ratio indicates that as PPE increased by $1, the relative risk of a school being a LPS decreased by 1%. The resulting odds ratio ($OR = 0.22$), after converting the metric of measurement for PPE from $1 to $100, suggests that for every $100 increase in PPE, the relative risk of being a LPS decreased by 7.9%. Neither the Wald statistic for MPS (see Table 4.14), nor the Wald statistic for HPLES were significant; the corresponding odds

Table 4.14: Effective Math Schools on Per Pupil Expenditures w/ Adjustment Model

	Wald	*p*	*OR*
High-Performing Low Equity Schools	.52	.470	0.98
Moderately Performing Schools	2.92	.088	0.96
Low Performing Schools	7.52	.006**	0.92

Notes. Referent Group: Effective Schools
** p < .05; ** p < .01*

ratios suggest that increasing PPE by $100 would decrease the odds that a school would be MPS by 4.2%, and that a school would be a HPLES by 1.7%.

Regression of group on the PTMDH, with the adjustment model, yielded significant findings ($p < .001$, $df = 60$, *pseudo* $R^2 = .72$). These results indicate that the PTMDH significantly influenced the effectiveness of schools in math and could explain 72% of the variability in the effectiveness of schools. As reported in Table 4.15, the Wald statistics were significant for LPS, MPS, and HPLES. The odds ratios suggest that as the PTMDH increased by 1%, the likelihood of a school being effective increased. More specifically, a 1% increase in the PTMDH decreased the relative risk that a school would be a LPS by 12.7%, a MPS by 11.2%, and a HPLES by 5.7%.

Review of the obtained results suggests that the model containing PFRL with the adjustment model had the strongest explanatory effect for math ($p < .001$; $df = 60$; *pseudo*

Table 4.15: Effective Math Schools on % Teachers w/ Master's Degree or Higher with Adjustment Model

	Wald	*p*	*OR*
High-Performing Low Equity Schools	3.99	.046*	0.94
Moderately Performing Schools	10.57	.001**	0.89
Low Performing Schools	14.20	<.001**	0.87

Notes. Referent Group: Effective Schools
** p < .05; ** p < .01*

Table 4.16: Correlational Matrix for Math Predictor Variables

Explanatory Variables	1	2	3	4
1. PWS	---	-.72**	.44**	.12
2. PFRL	-.72**	---	-.29**	-.52**
3. PPE	.44**	-.29**	---	.28**
4. PTMDH	.12	-.52**	.28**	---

Notes. n = *101*
** p < .05, two-tailed. ** p < .01, two-tailed.*

R^2 = .86). The obtained *pseudo R squared* suggests that 86% of the variability in the outcome variable can be explained by this model leaving less than 14% of the variation unexplained. Since this model had the strongest explanatory effect for School Effectiveness in math, it was identified as being the essential explanatory variable to be included in the full model.

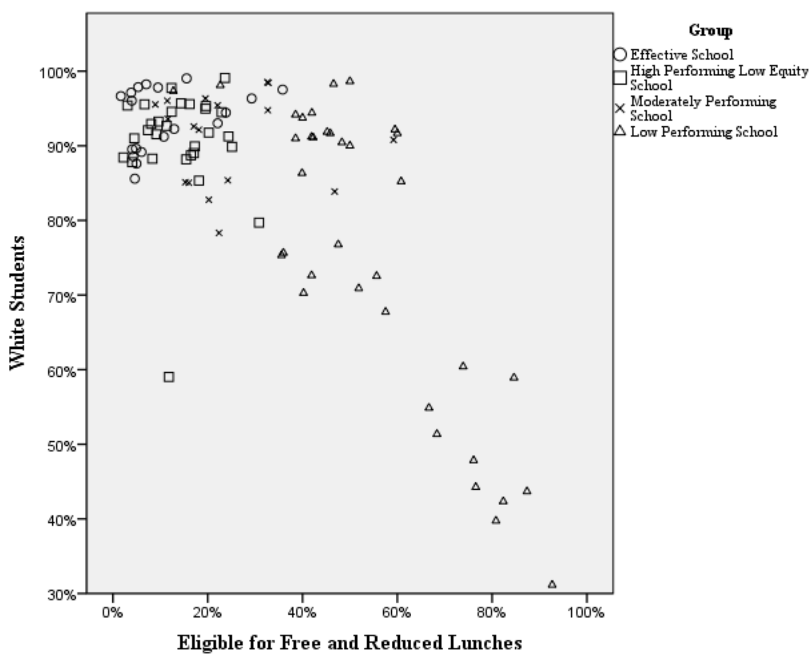

Figure 4.4: Scatterplot Percent White Students vs. Percent Eligible for Free and Reduced Lunch

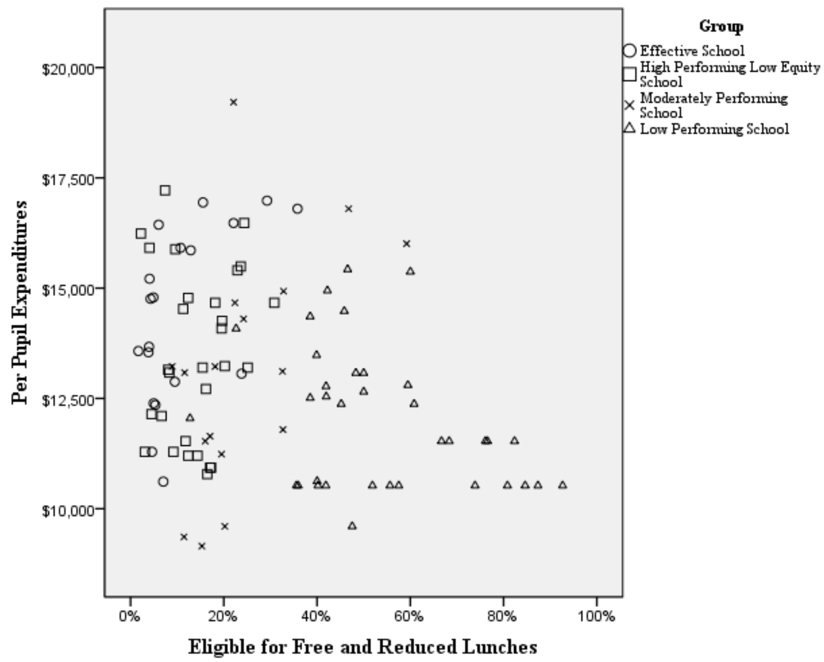

Figure 4.5: Scatterplot Per Pupil Expenditures vs. Percent Eligible for Free and Reduced Lunches

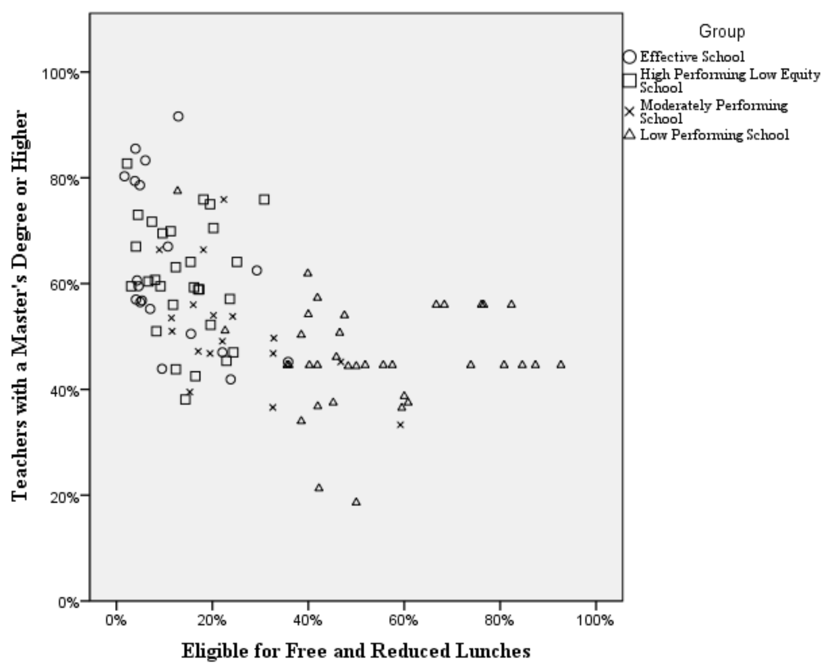

Figure 4.6: Scatterplot Percent Teachers with a Master's Degree or Higher vs. Percent Eligible for Free and Reduced Lunches

As was done in response to the third research question, tests of collinearity among the explanatory variables were assessed using 2-tail tests of correlation to determine if other explanatory variables could be included in the full model. The same process was employed and resulted in highly similar results (see Table 4.16): PFRL was found to be significantly correlated with PWS, PPE, and PTMDH. Scatterplots depicting these correlations are provided in Figures 4.4, 4.5, and 4.6. These results suggest that as the PFRL increases, the PWS, the PTMDH, and the PPE decrease. Given that data pertaining to PPE and PTMDH were not available at the individual school level, these variables were not truly independent from each other.

Since the model containing PFRL with the adjustment model had the strongest explanatory effect on the effectiveness of schools for math, and PFRL is significantly correlated with each of the other explanatory variables, a more complete model could not be constructed for math. As such, the best logistic regression model for math included only the PFRL with the adjustment model.

Conclusion

The current study examined the effectiveness of NH elementary schools in reading and math based on the performance of third, fourth, and fifth graders who took the reading and math NECAPs in 2010, 2011, and 2012. Data from the NH DoE were used to sort schools into groups to identify the percentage of NH elementary schools that were effective in reading, the percentage of NH elementary schools that were effective in math, and the extent to which four explanatory variables (PWS, PFRL, PPE, & PTMDH) could explain the variability in the effectiveness of schools in reading and in math.

The first two research questions pertained to the percentage of NH elementary schools that were effective for reading and math. Of the 209 NH elementary schools that were included in this study, 18 (8.6%) were effective for reading and 19 (9%) were effective for math.

The third and fourth research questions examined how well the explanatory variables (PWS, PFRL, PPE, & PTMDH) explain School Effectiveness (ES, HPLES, MPS, & LPS) for reading and math given ES as the referent group when NH County and School Size are controlled. Since each of the explanatory variables was better at explaining the outcome variable than were the empty models, the null hypotheses for the third and fourth research questions should be rejected. For both reading and math, PFRL, with the adjustment model, accounted for the most amount of variability in the effectiveness of schools explaining 84% of the variability for reading and 86% of the variability for math. Since PFRL was significantly correlated with each of the other three explanatory variables, for both reading and math, a more complete model could not be constructed.

Chapter 5: Conclusions and Suggestions for Further Research

The performance of US students on international, national, and statewide assessments consistently reveals achievement gaps based on race and SES (Benitez et al., 2009: Darling-Hammond, 2010; Domina, 2006; Education Trust, 2008a, Education Trust, 2011; Kozol, 2005; Lezotte, 2009; McKinsey & Company, 2009; Sharma et al., 2014; Singh, 2015). The existence of achievement gaps between White students and non-White students, as well as between low-income students and their peers is having an adverse effect on the US's ability to compete in the global marketplace (Darling-Hammond, 2010, 2012; Education Trust, 2006b; Education Trust, 2008b; Education Trust, 2011; Lezotte, 2009; Trilling & Fadel, 2009; Wagner, 2008) thereby crippling its fiscal well-being (McKinsey & Company, 2009).

Since the mid-1900s, numerous politicians and researchers have questioned the extent to which US public schools could be effective for all students regardless of out-of-school factors such as race and SES. The controversial interpretation of The Coleman Report (1966) led many to believe that schools had limited ability to ameliorate the impact of these out-of-school factors. Further analysis of The Coleman Report, along with analysis of a subsequent study that re-examined the original data set (Konstantopoulos & Borman, 2011) suggests that improvements in the effectiveness of schools would have the strongest effect on non-White and low-income students. Coleman et al. (1966) cited student body composition and teacher quality as being the two variables that had the strongest impact on the educational outcomes of these students.

In the decades since the publication of The Coleman Report, the effectiveness of schools has been thoroughly examined. In the first two decades following The Coleman Report, the effective schools research was prevalent; however, due to a myriad of criticisms regarding its effectiveness, perceptions of this body of research among educational researchers quickly became negative (Cuban, 1998). As such, effective schools research ceased by the early 1990s. Through the policy work of Marshall Smith, the effective schools movement continued to influence educational policies and reform movements (Cuban, 1998).

A multitude of educational research has been conducted in recent years revealing inequitable educational opportunities as a result of inequitable funding practices (Baker et al., 2017; Hall & Ushomirsky, 2010; Organization for Economic Cooperation and Development [OECD], 2010; Peske & Haycock, 2006; Schmidt et al., 2009; Sharma et al., 2014) and teacher qualifications (Clotfelter et al., 2010; Darling-Hammond, 2010; Hall & Ushomirsky, 2010; Minor et al., 2015; Owens, 2010; Peske & Haycock, 2006; Schultz, 2014; Sharma et al., 2014; Wiener, 2008), both of which tend to vary based on student body composition (Darling-Hammond, 2010; Owens, 2010).

These inequitable educational opportunities translate to inequitable life opportunities which are having significant social and economic effects on the vitality of the US (McKinsey & Company, 2009). The limited effectiveness and low-equity of US schools leaves many young adults unprepared to enter college or the workforce thereby crippling the US's ability to participate in the global marketplace (Darling-Hammond, 2010; Education Trust, 2006b; Education Trust, 2008b; Education Trust, 2011; Lezotte, 2009; Trilling & Fadel, 2009; Wagner, 2008). As noted by Linda Darling-Hammond (2012), "No society can thrive in a technological, knowledge-based economy by depriving large segments of its population of learning" (p. 24).

To address this problem, researchers must do a better job of examining and quantifying the explanatory effects of key school variables such as student body composition, teacher quality, and funding practices on the effectiveness of schools for all students. Given its focus on effective schools being effective for all students, the theoretical orientation of the effective schools research was used as the basis for the current research. According to this line of research, an effective school is one that is effective for all students regardless of students' race or SES.

As with other bodies of research, the effective schools research was heavily criticized for faulty research methods such as overly small (Jansen, 1995; Purkey & Smith, 1982, 1983; Witte & Walsh, 1990) and non-representative sample sizes (Purkey & Smith, 1982, 1983), the lack of stability of results from one year to the next (Purkey & Smith, 1982, 1983), and its over reliance on aggregated data (Jansen, 1995; Purkey & Smith, 1982, 1983; Rowan et al., 1983; Witte & Walsh, 1990). Moreover, effective schools' researchers' tendency to focus on outlier schools without regard to typical schools, as well as the limited utility of the results in identifying the relative influence of various school variables on increasing the effectiveness of schools, left many questioning the value of the effective schools research (Purkey & Smith, 1982, 1983; Rowan et al., 1983; Witte & Walsh, 1990).

The current research design was formulated in light of the extensive criticisms of the effective schools research. The consideration of a school's performance in three consecutive years enabled a higher degree of confidence in determining a school's effectiveness than do other studies in which determinations are made solely on a school's performance in one year. The use of disaggregated data permitted the consideration of within-school differences. Finally, to increase the utility of the current study, the majority of NH elementary schools were included in this study including low-performing, moderately-performing, and high-performing schools; Only 54 small schools and two other schools, were omitted from the current study.

Based on an extensive review of the effective schools research and other more current research, three variables were of particular interest to this study: student body composition, school funding practices, and teacher qualifications. Student body composition refers to not only the racial composition of students, but also to the proportion of students who come from low-income families. Considering both aspects of student body composition enabled the separate consideration of their effects while also providing more insight into how they covary.

The current study was comprised of three parts. First, NH elementary schools were sorted into three preliminary groups: HPS, MPS, and LPS. Schools that did not meet the criteria for these groups were included in the total number of schools (n = 209), but were not considered when examining the effects of the explanatory variables. Then the HPS were split into ES and HPLES. Finally, multinomial logistic regression models were constructed to determine how well the explanatory variables could account for the effectiveness of schools for reading and for math in response to the third and four research questions.

Discussion of Findings

The first two research questions examined the percentage of NH elementary schools that were effective for reading and the percentage of NH elementary schools that were effective for math. Of the 209 schools that were included in this study, only 18 were effective for reading and 19 were effective for math. This translates to an effectiveness rate of 8.6% for reading and 9% for math. Readers are cautioned to recall that the methods by which schools were classified as being effective were far more extensive than is typically employed in research of this nature and therefore the reported statistics should be considered simply as a baseline for the effectiveness of NH elementary schools as of the fall of 2012. Given the wealth of research at the national level that suggests the inequity of education in the US, this researcher was cautiously concerned at the onset of the current study that the percentage of effective schools would be lower. It is hoped, however, that the effectiveness rate of NH elementary schools will increase dramatically in upcoming decades as the impact of these explanatory variables is better translated into equitable educational practices.

Analysis of the inter-relationships amongst the four explanatory variables reveals significant correlations. The strongest correlations were observed between the two student body composition variables: percentage of White students and the percentage of students eligible for free and reduced lunches (reading: $r = -.70$, $p < .001$; math: $r = -.72$, $p < .001$) suggesting that as the percentage of White students in a school increases, the percentage of students eligible for free and reduced lunches decreases. Moreover, the strength of this correlation supports the assertion that these two variables both assess student body composition.

The percentage of students eligible for free and reduced lunches was negatively correlated with per pupil expenditures for reading (r = -.24, p = .01) and for math (r = -.29, p = .004). Meanwhile, the percentage of White students in a school was found to be positively correlated with per pupil expenditures for reading (r = .47, p < .001) and for math (r = .44, p < .001). These results suggest that the schools with higher proportions of low-income students, as measured by the percentage of students eligible for free and reduced lunch, and the percentage of non-White students, provide less funding per pupil than do schools with fewer low-income students and fewer non-White students.

The percentage of students eligible for free and reduced lunches was moderately correlated with the percentage of teachers with a master's degree or higher for reading (r = -.51, p < .001) and for math (r = -.52, p < .001) suggesting that schools that have higher proportions of low-income students tend to have less qualified teachers than do schools that have lower proportions of low-income students. The percentage of White students in a school was not significantly correlated with the percentage of teachers with a master's degree or higher for reading (r = .17, p = .09) or for math (r = .21, p = .23) suggesting a weak association between race and teacher qualifications.

Each of the explanatory variables included in this study was found to be significantly better at accounting for the variation in school effectiveness for reading and for math than were empty models. The obtained results suggest that student body composition, school funding practices, and teacher qualifications are each able to explain a substantial amount of the variability in the effectiveness of schools for both reading and for math. The explanatory effects of these variables were enhanced when combined with an adjustment model which controlled for NH County and School Size suggesting that the influence of these explanatory variables is stronger when the NH County in which the school is located, and the size of the school are considered.

Per pupil expenditures explained 65% of the variability in school effectiveness for reading, and 71% of the variability in school effectiveness for math suggesting that the higher the rate of funding per pupil, the more likely a school would be effective. More specifically, the analysis suggests that as per pupil expenditures increased by $100, the relative risk of a school being a low-performing school decreased by 7% for reading, and 7.9% for math.

The percentage of teachers with a master's degree or higher explained 65% of the variability in school effectiveness for reading, and 72% of the variability in school effectiveness for math. Suggesting that the higher the proportion of teachers in a school with at least a master's degree, the higher the likelihood that a school would be effective. Further analysis suggests that as the percentage of teachers with a master's degree or higher increased by 1%, the odds of a school being a low-performing school in reading decreased by 16%. Likewise, a 1% increase in the percentage of teachers with a master's degree or higher decreased the odds that a school would be a low-performing math school by 12.7%, a moderately-performing math school by 11.2%, and a high-performing low-equity school be 5.7%. This pattern of results suggest that increases in teacher qualifications appears to have the most pronounced effect on decreasing the likelihood that a school would be a low-performing reading school, but had the most consistent effect on improving the effectiveness of schools in math. These findings are very intriguing and if replicated would be worthy of further investigation.

The percentage of White students in a school could explain 69% of the variability in school effectiveness for reading, and 73% of the variability in school effectiveness in math. These results suggest that schools with higher proportions of White students are more likely to be effective than are schools with lower proportions of White students. Moreover, the results of this regression revealed that as the percentage of White students in a school increased by 1%, the likelihood of a school being a low-performing school in reading decreased by 28.8%. Similarly, a 1% increase in the percentage of White students decreased the likelihood of that school being a low-performing school in math by 22%. Even though a 1% increase in the percentage of White students was not significant at decreasing the risk of schools being moderate-performers in reading or math, a significant effect was observed with decreasing the relative risk of being a high-performing low-equity school for math. A 1% increase in the percentage of White students yielded a decrease in the relative risk of a school being a high-performing low-equity school in math instead of an effective school by 12%. Hence, among the high-performing low-equity schools, a 1% increase in the percentage of White students appears to have a pronounced impact on increasing the likelihood that a school would be effective in math, but had a minimal impact on increasing the likelihood that an otherwise high-performing school would be effective in reading. While this finding is intriguing, it is not substantiated in research and as such should be interpreted with caution as it may be an anomaly within the current study.

The effectiveness of NH elementary schools for reading and math was best explained by the percentage of students eligible for free and reduced lunches in a school. This model accounted for 84% of the variation in the effectiveness of schools for reading and 86% of the variability in the effectiveness of schools for math. These findings suggest that schools that have lower proportions of low-income students are more likely to be effective. For reading, as the percentage of students eligible for free and reduced lunches in a school increased by 1%, schools were 3.38 times more likely to be low-performing schools than effective schools, 1.95 times more likely to be moderately-performing schools than effectives schools, and 1.78 times more likely to be high-performing low-equity schools than effective schools. For math, a 1% increase in the percentage of low-income students appeared to have a more pronounced effect. Such a change increased the odds that a school would be a low-performing school by 9.92, a moderate-performing school by 8.46, and a high-performing low-equity school by 7.48. Even though the effect of the percentage of low-income students in a school was significant for both reading and math, this explanatory variable appears to have a stronger effect on math than reading.

Overall, the results of the current study are consistent with research available at the national level in suggesting the inter-relationship among student body composition, funding practices, and teacher qualifications. Moreover, the current study suggests that student body composition, school funding practices, and teacher qualifications can each account for observed variations in the effectiveness of schools, for both reading and math, especially when the schools' locations and sizes are controlled for. Slight observed changes in the explanatory variables consistently effected the relative risk of schools being low-performing schools. Some of the explanatory variables were observed to have a stronger influence on the effectiveness of schools in math than in reading.

As is consistent with literature on student body composition, the results of the current study suggest that schools with the highest proportions of non-White students and schools with the highest proportion of low-income students, are at high risk for being low-performing schools (Baker et al., 2017; Darling-Hammond, 2010; Hall & Ushomirsky, 2010; O'Gorman, 2010; Peske & Haycock, 2006; Schmidt et al., 2009). Moreover, the results of the current study are consistent with literature on equity in educational opportunities in suggesting that the schools with higher concentrations of low-income students are provided with less funding per student and less qualified teachers (Boyd et al., 2005; Clewell & Campbell, 2007; Darling-Hammond, 2010; Konstantopoulos, 2011; Marzano, 2003; Minor et al., 2015; Peske & Haycock, 2006; Phillips, 2010; Pitre, 2014; Schultz, 2014; Sharma et al., 2014). Finally, the results of the current study suggest that student body composition, school funding practices, and teacher qualifications can explain the effectiveness of schools; most notably that changes in the observed values of the explanatory variables can increase the odds of schools being effective in reading and math. Given that the findings of this study have not yet been replicated in the state of NH some caution needs to be employed in the consideration of these results. Nevertheless, given the commonalities of the current findings and the research available at the national level, this writer hopes that NH educational administrators and policy makers will consider efforts to increase the per pupil expenditures and the proportion of teachers with a master's degree or higher in NH elementary schools that serve the highest proportion of low-income students.

Delimitations and Recommendations for Further Research

Despite attempts to create the most methodically sound study, the current study could not address all relevant issues. Perhaps the most significant delimitation is the use of high stakes reading and math tests as the sole metric for assessing the effectiveness of schools. First, there are other important academic skills such as science, writing, humanities, and the arts (Farrington et al., 2012) that were not considered in the current study. Secondly, these high stakes tests often have limited authenticity with real life tasks designed to assess higher level thinking and reasoning skills. Moreover, the desired outcomes of schools should far exceed how well students perform on tests of academic abilities, such as students' abilities to collaboratively engage with others, to think critically, and to be productive members of their community (American School Counselor Association, 2014; Farrington et al., 2012).

To truly assess the effectiveness of schools, research should not be limited to the elementary level. Just as the developmental and functional needs of students change as they get older, so does the context in which they are educated. Unlike elementary students who are typically taught by one classroom teacher in the context of the same peer group for the duration of an academic year, middle and high school students are in multiple classrooms and with a wider array of peers every day. Unfortunately, such an examination exceeded the capacity of the current study.

An important element of the effective schools research is the correlates of effective schools. Research spanning multiple decades cites the importance of school climate, strong leadership, high expectations for success, positive home-school relationships, data-based decision making, and time on-task, in the effectiveness of schools (Edmonds, 1982; Lezotte, 1991, 2009; Lezotte & Snyder, 2011). Might it be possible that one or more of these correlates could mitigate the effects of the explanatory variables that were the focus of the current study? Consideration of this question was not feasible as part of the current study. Subsequent research should consider the use of mixed methods so that the influence of these correlates can be considered in conjunction with the effects of student body composition, school funding practices, and teacher quality. More specifically, future research might wish to combine the quantitative methods employed with this study with the qualitative methods employed in Vermont's study (2009).

Finally, the current study was not able to consider the effectiveness of NH elementary schools for other groups of at-risk students. In addition to racial minority and low-income students, students who have limited English proficiency, and learning disabilities should also be provided with equitable educational opportunities. Consideration of these groups of students was beyond the scope of the current study, but should be examined via similar research studies.

Limitations and Additional Recommendations for Further Research

A major limitation of the current study is the unusually small subsample sizes within schools. Several of the high-performing schools only had one student of color which reflects the fact that 90% of NH students are White (Ayscue & Jau, 2014). Given these demographics, the utility of a state such as NH to adequately examine the effects of race on school effectiveness is limited. In order to better examine this construct, subsequent research should consider examining the effects of race at a regional level.

The demographics of the state, in regards to the vast differences between the rural nature of the northernmost counties and the more populated and racially diverse southern counties, warrants consideration. Even though NH County and School Size were controlled for, the disproportionate number of effective schools located in the southern part of the state is concerning. The current research failed to provide insight as to whether this occurred because those counties have more schools or because of the characteristics of those counties themselves. To better account for these contrasting demographics, it may be beneficial for subsequent research to be conducted at a regional level so that rural schools can be compared to other rural schools while more urban and suburban schools can be compared. Expanding this research to the regional level should improve its utility.

Unfortunately, data related to per pupil expenditures and the percentage of teachers with a master's degree or higher were not available at the individual school level. The lack of independence of this data between schools in this study reduced the power of the study. Given the wealth of research that suggests inequitable funding practices and teacher qualifications from one school to the next within school districts

(Engel, 2010; Hall & Ushomirsky, 2010; Peske & Haycock, 2006; Roza, 2006; Schultz, 2014; Wiener, 2008), subsequent research would benefit from opportunities to obtain this data at the individual school level. In order to improve the utility of subsequent research, it is imperative that policy changes occur which require more accurate data accounting methods in relation to school funding practices and teacher qualifications at the individual school level. The resolve of state departments of education to permit this data to be reported at the district level downplays the impact that inequitable funding practices have on the effectiveness of schools within districts thereby perpetuating the inequity that occurs within many school districts across this nation. Improved data on these variables will increase the power of subsequent research so that researchers can better understand the complex interplay of these variables on school effectiveness.

Conclusions

The results of the current study indicate that as of the fall of 2012, only 8.6% of the 209 NH elementary schools included in this study were effective in reading and only 9% were effective in math. Student body composition, school funding practices, and teacher qualifications each explained a significant amount of the variability in the effectiveness of schools. The percentage of low-income students in a school had the strongest explanatory effect, accounting for more than 80% of the variability in the effectiveness of schools.

This study contributes to the literature citing the inter-relationships of student body composition, school funding practices, and teacher qualifications. More specifically, the current study revealed that among NH elementary schools the percentage of low-income students in a school is negatively correlated with the percentage of White students, the percentage of teachers with at least a master's degree, and the per pupil expenditures. In other words, schools that have the highest concentrations of low-income students are more likely to have more non-White students, less qualified teachers, and less funding per student.

Given the importance for schools to be effective for all students, regardless of race or SES, the decision to not control SES should be considered an asset to the current study. Recent research conducted by the OECD (2010) suggests that the effects of SES can be ameliorated. The findings of the current study, the OECD (2010) study, as well as the existent literature on the relationship between student body composition, school funding practices, and teacher qualifications, will hopefully trigger researchers to consider SES as a variable to be addressed as opposed to a variable that is beyond our control.

Even though much has been learned about the effectiveness of schools in the past few decades, the words written by Ron Edmonds in 1979 continue to capture the essence of the truth:

> We can whenever we choose, successfully teach all children whose schooling is of interest to us. We already know more than we need to do that. Whether or not we do it must finally depend on how we feel about the fact that we haven't so far. (p. 23)

Through these words, Ron Edmonds challenges educators and researchers to act to provide equitable educational opportunities for all students. As such, educators must abandon the belief that schools cannot mitigate the effects of out-of-school variables. Improved methods for the accurate assessment and monitoring of school effectiveness must be central to this movement. Given its focus on equitable educational opportunities for all students regardless of race and SES, a true resurgence of the effective schools research, which addresses its criticisms, is strongly encouraged.

References

American School Counselor Association. (2014). *ASCA mindsets and behaviors for student success: K-12 college- and career readiness standards for every student.* Retrieved from the ASCA website https://www.schoolcounselor.org/ school-counselors-members/about-asca/mindsets-behaviors

Aud, S., Hussar, W., Johnson, F., Kena, G., Roth, E., Manning, E., Wang, X., & Zhang, J. (2012). The condition of education 2012 (NCES 2012-045). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from www.nces.ed.gov/pubs2012/2012045.pdf

Ayscue, J. B., & Jau, S. (with Flaxman, G., Kucsera, J., & Siegal-Hawley, G), (2014). Diversity in the distance: The onset of racial change in Northern New England schools. Retrieved from The Civil Rights Project website: https://www.civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/integration-and-diversity/atct_topic_view?b_start:int=7&-C=

Baker, B., Farrie, D., Johnson, M., Luhm, T., & Sciarra, D. G. (2017). *Is school funding fair? A national report card* (6th ed.). Retrieved from the Education Law Center website: http://www.schoolfundingfairness.org/is-school-funding-fair/reports

Bakker, M., & Wicherts, J. M. (2014). Outlier removal, sum scores, and the inflation of the type I error rate in independent samples *t* tests: The power of alternatives and recommendations. *Psychological Methods, 19*(3), 409-427.

Balfanz, R., Legters, N., West, T. C., & Weber, L. M. (2007). Are NCLB's measure, incentives and improvement strategies the right ones for the nation's low-performing high schools? *American Educational Research Journal, 44*(3), 559-593.

Benitez, M., Davidson, J., & Flaxman, L. (2009). *Small schools big ideas: The essential guide to successful school transformation.* San Francisco, CA: Jossey-Bass.

Bracey, G. W. (2008). NCLB: Achievement Robin Hood? *Principal Leadership, 9*(3), 47-49.

Brimley, V., & Garfield, R. R. (2008). *Financing education in a climate of change (10th ed.).* Boston, MA: Pearson.

Brookover, W. B., Schweitzer, J. H., Schneider, J. M., Beady, C. H., Flood, P. K., & Wisenbaker, J. M. (1978). Elementary school social climate and school achievement. *American Educational Research Association, 15*(2), 301-318.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management, 27,* 793-818.

Callender, B. (2007). *The relationship among effective school correlates, school and district practices, and exemplary student performance in Texas* (Doctoral Dissertation). Retrieved from Dissertations & Theses: A&I.

Center on Education Policy. (2011). *Is achievement improving and are gaps narrowing for Title I students?* Retrieved from http://www.cep-dc.org/

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood* [NBER Working Paper

#17699]. Retrieved from www.equality-of-opportunity.org/assets/documents/teachers_wp.pdf

Clewell, B. C., & Campbell, P. B. (with Perlman, L.) (2007). *Good schools in poor neighborhoods: Defying demographics, achieving success.* Washington, D.C.: The Urban Institute Press.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher mobility, school segregation, and pay-based policies to level the playing field. *Education, Finance, and Policy, 6*(3), 399-438.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Government Printing Office.

Dantley, M. E. (1990). The ineffectiveness of effective schools leadership: An analysis of the effective schools movement from a critical perspective. *Journal of Negro Education, 59*(4), 585-598.

Darling-Hammond, L. (2004). From “separate but equal” to “No Child Left Behind”: The collision of new standards and old inequalities. In J. A. Banks (Series Ed.), *Multicultural Education Series: Many children left behind* (pp. 3-32). New York, NY: Teachers College Press.

Darling-Hammond, L. (2010). The flat world and education: How America’s commitment to equity will determine our future. In J. A. Banks (Series Ed.), *Multicultural Education Series: Many children left behind* (pp. 3-32). New York, NY: Teachers College Press.

Darling-Hammond, L. (2012). Soaring systems. *Educational Review*, *24*(1), 24-33.

DeCarlo, D. T. (1997). On the meaning and use of kurtosis. *Psychological methods, 2*(3), 292-307.

Dee, T. S., & Jacob, B. (2010). Evaluating NCLB. *Education Next, 10*(3), 54-61.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418-446.

Domina, T. (2006). Brain drain and brain gain: Rising educational segregation in the United States, 1940-2000. *American Sociological Association*, *5*(4), 387-407.

Edelman, M. W., & Jones, J. M. (2004). Separate and unequal: America's children, race, and poverty. *Growing up American*, *14*(2), 134-136.

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15-24.

Edmonds, R. R. (1982). Programs of school improvement: An overview. *Educational Leadership, 40*(3), 4-11.

Education Trust. (2004a). The ABCs of "AYP": Raising achievement for all students. Retrieved from the Education Trust website: http://www.edtrust.org/sites/ edtrust.org/files/publications/files/ABCAYP.PDF

Education Trust. (2004b). *Measured progress*. Retrieved from The Education Trust website: http://www.edtrust.org/sites/edtrust.org/files/publications/files/ MeasuredProgressOct2004.pdf

Education Trust. (2006a). *Funding Gaps 2006.* Retrieved from The Education Trust website: http://www.edtrust.org/sites/edtrust.org/files/publications/files/

FundingGap2006.pdf

Education Trust. (2006b). *Yes we can: Telling truths and dispelling myths about race and education in America.* Retrieved from The Education Trust website: http://www.edtrust.org/sites/edtrust.org/files/ publications/files/YesWeCan.pdf

Education Trust. (2008a). Core problems: Out of field teaching persists in key academic courses and high-poverty schools. Retrieved from The Education Trust website: http://www.edtrust.org/sites/edtrust.org/files/publications/files/ SASSreportCoreProblem.pdf

Education Trust. (2008b). *Their fair share: How Texas sized gaps in teacher quality shortchange low-income and minority students*. Retrieved from The Education Trust website: http://www.edtrust.org/sites/edtrust.org/files/publications/files/ TXTheirFairShare.pdf

Education Trust. (2011). Executive summary. In N. Ushomirsky, D. Hall, & K. Haycock's, *Getting it right: Crafting federal accountability for higher student performance and a stronger America.* Retrieved from the Education Trust website: http://www.edtrust.org/sites/edtrust.org/files/publications/files/ Getting_It_Right.pdf

Engel, M. (2010). School finance: Inequality persists. In R. Heiner (Ed.) *Conflicting interests: Readings in social problems and inequality*, pp. 125-130. New York, NY: Oxford University Press.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners. The* *role of noncognitive factors in shaping school performance: A critical literature review.* Retrieved from The University of Chicago Consortium on Chicago School Research website https://consortium.uchicago.edu /publications/ teaching-adolescents-become-learners-role-noncognitive-factors-shaping-school

Foster, J., Barkus, E., & Yavorsky, C. (2006). *Understanding and using advanced statistics*. Thousand Oaks, CA: Sage.

Guarino, C. M., Brown, A. B., & Wyse, A. E. (2010). Can districts keep good teachers in the schools that need them most? *Economics of Education Review, 30,* 962-979.

Hall, D., & Kennedy, S. (2006). Primary progress, secondary challenge: A state-by-state look at student achievement patterns. Retrieved from The Education Trust website: http://www.edtrust.org/sites/edtrust.org/files/publications/files/ PPSCreport.pdf

Hall, D., & Ushomirsky, N. (2010). *Close the hidden funding gaps in our schools.* Retrieved from The Education Trust website: http://www.edtrust.org/sites/ edtrust.org/files/publications/files/Hidden%20Funding%20Gaps_0.pdf

Harrington, A. (2005). Introduction: What is social theory? In A. Harrington (Ed.), *Modern social theory: An introduction* (pp. 1-15). New York, NY: Oxford University Press.

Hayes, S. (2008). *Characteristics of successful high-poverty schools: What the research tells us* [Literature Review]. Retrieved from Vermont Department of Education Roots to Success website: http://rootsofsuccess.files.wordpress.com/ 2009/09/successful_schools_lit_review1.pdf

Hoaglin, D. C., & Iglewicz, B. (1987). Fine tuning some resistant rules for outlier labeling. *Journal of American Statistical Association, 82*(400), 1147-1149.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). New York, NY: Wiley.

Howe, H., II (1966). [Letter]. In J. S. Coleman, E. Q. Campbell, C. J. Hobson, J. McPartland, A. M. Mood, F. D. Weinfeld, & R. L. York's, *Equality of Educational Opportunity.* Washington, DC: U.S. Government Printing Office.

Howell, D. C. (1997). *Statistical methods for psychology (4th ed.).* Belmont, CA: Duxbury Press.

Hursch, D. (2007). Exacerbating inequality: The failed promise of the No Child Left Behind Act. *Race Ethnicity and Education, 10*(3), 295-308. doi: 10.1080/13613320701503264

Institute for Digital Research and Education. (2011). *FAQ: What are pseudo* R-*squareds?* Retrieved from http://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/

Jaccard, J. (2001). Interaction effects in logistic regression: A primer. In M. S. Lewis-Beck (Series Ed.), *Quantitative Applications in the Social Sciences.* Thousand Oaks, CA: Sage.

Jansen, J. D. (1995). Effective schools? *Comparative Education*, *31*(2), 181–200.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B. Y., & Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America.* New York, NY: Basic Books.

Jennings, J. L., Deming, D., Jencks, C., Lopuch, M., & Schueler, B. E. (2015). Do differences in school quality matter more than we thought? New evidence on educational opportunity in the twenty-first century. *Sociology of Education, 88*(1), 56-82.

Karp, S. (2004). NCLB's selective vision of equality: Some gaps count more than others. In J. A. Banks (Series Ed.), *Multicultural Education Series: Many children left behind* (pp. 53-65). New York, NY: Teachers College Press.

Keith, T. Z. (2006). *Multiple regression and beyond.* Boston, MA: Pearson.

Klitgaard, R. E., & Hall, G. R. (1975). Are there unusually effective schools? *Journal of Human Resource, 10,* 90-106.

Konstantopoulos, S. (2011). Teacher effects in early grades: Evidence from a randomized study. *Teachers College Record, 113*(7), 1541-1565.

Konstantopoulos, S., & Borman, G. D. (2011). Family background and school effects on student achievement: A multilevel analysis of the Coleman data. *Teachers College Record, 113*(1), 97-132.

Kozol, J. (2005). *The shame of the nation: The restoration of apartheid schooling in America.* New York, NY: Three Rivers Press.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher, 35*(7), 3-12.

Leach, M. T., & Williams, S. A. (2007). The impact of the academic achievement gap on the African American family: A social inequality perspective. *Journal of Human Behavior in the Social Environment, 15*(3), 39-59. doi: 10.1300/J137v15n02_04

Lezotte, L. W. (1991). *Correlates of effective schools: The first and second generation.* Retrieved from http://www.effectiveschools.com/images/ stories/escorrelates.pdf

Lezotte, L. W. (2005). More effective schools: Professional learning communities in action. In R. [Richard] DuFour, R. Eaker, & R. [Rebecca] DuFour (Eds.) *On common ground: The power of professional learning communities* (pp. 177-191). Bloomington, IN: Solution Tree.

Lezotte, L. W. (April, 2009). Effective schools: Past, present, and future. In K. Brock (Chair), *2009 Brock Symposium for Excellence in Education*. Symposium conducted at the University of Tulsa, OK. Retrieved from http://www.effectiveschools.com/images/stories/brockpaper.pdf

Lezotte, L. W., & Snyder, K. M. (2011). *What effective schools do: Re-envisioning the correlates.* Bloomington, IN: Solution Tree.

Liu, G. (2006). How the federal government makes rich states richer. In The Education Trust's *Funding Gaps 2006,* pp. 2-4. Retrieved from The Education Trust website: http://www.edtrust.org/sites/edtrust.org/files/publications/files/ FundingGap2006.pdf

Marzano, R. J. (2000). *A new era of school reform: Going where the research takes us* (REL Publication No. RJ96006101). Retrieved from Mid-continent Research for Education and Learning website: http://www.mcrel.org

Marzano, R. J. (2003). *What works in schools: Translating research into action.* Alexandria, VA: Association for Supervision and Curriculum Development.

Measured Progress. (2010). *New England Common Assessment Program 2009-2010 Technical Report.* Retrieved from http://www.education.nh.gov/instruction/assessment/necap/documents/techrpt_sept1510.pdf

Measured Progress. (2011). *New England Common Assessment Program 2010-2011 Technical Report.* Retrieved from http://www.education.nh.gov/instruction/assessment/necap/documents/techrpt_july2011.pdf

Measured Progress. (2012). *New England Common Assessment Program 2011-2012 Technical Report.* Retrieved from http://www.ride.ri.gov/Assessment/DOCS/NECAP/Reports_Results/2011-12_NECAP_Technical_Report-w_Appendices.pdf

McKinsey & Company. (2009). *The economic impact of the achievement gap in America's schools.* Retrieved from www.mckinseyonsociety.com/the-economic-impact-of -the-achievement-gap-in-americas-schools/

Menard, S. (2002). Applied logistic regression analysis (2nd ed.). In M. S. Lewis-Beck (Series Ed.), *Quantitative Applications in the Social Sciences.* Thousand Oaks, CA: Sage.

Mikesell, J. L. (2011). *Fiscal administration: Analysis and applications for the public sector (8th ed.).* Boston, MA: Wadsworth.

Miles, J., & Banyard, P. (2007). *Understanding and using statistics in psychology: A practical introduction.* Los Angeles, CA: Sage.

Minium, E. W., King, B. M., & Bear, G. (1993). *Statistical reasoning in psychology and education (3rd ed.).* New York, NY: Wiley.

Minor, E. C., Desimone, L. M., Phillips, K. J. R., & Spencer, K. (2015). A new look at opportunity-to-learn gap across race and income. *American Journal of Education, 121*, 241-269. doi: 0195-6744/2015/12102-0004$10.00

Montt, G. (2011). Cross-national difference in educational achievement inequality. *Sociology of Education, 84*(1), 49-68. doi:10.177/0038040710392717

National Center for Education Statistics. (n.d.). *NAEP data explorer.* Retrieved from https://nces.ed.gov/nationsreportcard/naepdata/

National Center for Education Statistics. (2012). *NAEP 2012: Trends in academic progress.*

New Hampshire Department of Education. (2011). *New Hampshire's equity plan for ensuring that poor or minority children are not taught by inexperienced, unqualified or out-of-field teachers at higher rates than are other children.* Retrieved from http://www.education.nh.gov/nclb/documents/equity_plan.pdf

New Hampshire Department of Education. (2012a). *2012 NH Adequate Yearly Progress – Status. State summary report: Based on May 2012 NH ALPS and October 2011 NECAP Assessment Data* [Report]. Retrieved from http://www.education.nh.gov/instruction/accountability/ayp/documents/2012_state_grades3-8.pdf_grades3-8.pdf

New Hampshire Department of Education. (2013a). *Fall 2012 NECAP tests. State summary report: 2012-2013 students* [Report]. Retrieved from http://reporting.measuredprogress.org/NHProfile/dlrpt.aspx?lid=161030

New Hampshire Department of Education. (2013b). *Fall 2012. Beginning of Grade 3*

NECAP tests. State results [Report]. Retrieved from http://reporting.measuredprogress.org/NHProfile/dlrpt.aspx?lid=161020

New Hampshire Department of Education. (2013c). *Fall 2012. Beginning of Grade 4 NECAP tests. State results* [Report]. Retrieved from http://reporting.measuredprogress.org/NHProfile/dlrpt.aspx?lid=161021

New Hampshire Department of Education. (2013d). *Fall 2012. Beginning of Grade 5 NECAP tests. State results* [Report]. Retrieved from http://reporting.measuredprogress.org/NHProfile/dlrpt.aspx?lid=161023

O'Gorman, M. (2010). Educational disparity and the persistence of the Black-White wage gap in the U.S. *Economics of Education Review*, *29*(4), 526-542. doi: 10.1016/j.econedurev.2009.11.003

Organization for Economic Cooperation and Development. (2010). *PISA 2009 results: Overcoming social background - equity in learning opportunities and outcomes* (Volume II). Retrieved from www.oecd.org/pisa/pisaproducts/ 48852584.pdf

Organization for Economic Cooperation and Development. (2013). *Programme for international student assessment (PISA): Results from PISA 2012* [Country Note – United States]. Retrieved from https://www.oecd.org/unitedstates/ PISA-2012-results-US.pdf

Organization for Economic Cooperation and Development. (2016). *Programme for international student assessment (PISA): Results from PISA 2015* [Country Note – United States]. Retrieved from https://www.oecd.org/pisa/PISA-2015-United-States.pdf

Owens, A. (2010). Neighborhoods and schools as competing and reinforcing contexts for educational attainment. *American Sociological Association, 83*(4), 287-311. doi: 10.1177/0038040710383519

Pampel, F. C. (2000). Logistic regression: A primer. In M. S. Lewis-Beck (Series Ed.), *Quantitative Applications in the Social Sciences.* Thousand Oaks, CA: Sage.

Peske, H. G., & Haycock, K. (2006). *Teaching inequality: How poor and minority students are shortchanged on teacher quality.* Retrieved from The Education Trust website: http://www.edtrust.org/sites/edtrust.org/files/publications/files/TQReportJune2006.pdf

Peterson, P. E., & Hess, F. (2008). Few states set world-class standards: In fact, most render the notion of proficiency meaningless. *Education Next, 8*(3), 70-73.

Phillips, K. J. R. (2010). What does "Highly Qualified" mean for student achievement? Evaluating the relationships between teacher quality indicators and at-risk students' mathematics and reading achievement gains in first grade. *The Elementary School Journal, 110*(4), 464-493.

Pitre, C. C. (2014). Improving African American student outcomes: Understanding educational achievement and strategies to close opportunity gaps. *The Western Journal of Black Studies, 38*(4), 209-217.

Purkey, S. C., & Smith, M. S. (1982). Too soon to cheer? Synthesis of research on effective schools. *Educational Leadership, 40*(3), 64-69.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal, 83*(4), 426-452.

Reagle, C. (2006). Creating effective schools where all students can learn. *The Rural Educator, 27*(3), 24-33.

Rivkin, S. (2016). Desegregation since the Coleman report: Racial composition of schools and student learning. *Education Next, 2,* 28-37.

Rowan, A. H., Hall, D., & Haycock, K. (2010). *Gauging the gaps: A deeper look at student achievement.* Retrieved from The Education Trust website: http://www.edtrust.org/sites/edtrust.org/files/publications/files/NAEP%20Gap_0.pdf

Rowan, B., Bossert, S. T., & Dwyer, D. C. (1983). Research on effective schools: A cautionary note. *Educational Researcher*, *12*(4), 24-31.

Roza, M. (2006). How districts shortchange low-income and minority students. In The Education Trust's *Funding Gaps 2006.* Retrieved from The Education Trust website: http://www.edtrust.org/sites/edtrust.org/ files/publications/files/FundingGap2006.pdf

Schmidt, W. H., Cogan, L. S., Houang, R. T., & McKnight, C. (2009). *Equality of educational opportunity: A myth or reality in U.S. schooling.* Retrieved from The Education Policy Center at Michigan State University website: http://education.msu.edu/epc/forms/Schmidt _et_al_2009_Equality_ of_ educational_opportunity.pdf

Schultz, L. M. (2014). Inequitable dispersion: Mapping the distribution of highly qualified teachers in St. Louis metropolitan elementary schools. *Education Policy Analysis Archives, 22*(90), 1-22. doi: 10.14507/epaa.v22n90.2014

Shannon, G. S., & Bylsma, P. (2004). *Characteristics of improved school districts: Themes from research.* Retrieved from Office of Superintendent of Public Instruction, Olympia, WA website: http://www.k12.wa.us/publications

Shannon, G. S., & Bylsma, P. (2007). *Nine characteristics of high-performing schools: A research-based resource for schools and districts to assist with improving student learning* (2nd ed.) (Publication No. 07-0015). Retrieved from Office of Superintendent of Public Instruction, Olympia, WA website: http://www.k12.wa.us/publications

Sharma, A., Joyner, A. M., & Osment, A. (2014). Adverse impact of racial isolation on student performance: A study of North Carolina. *Education Policy Analysis Archives, 22*(14), 1-18. doi: 10.14507/epaa.v22n14.2014

Singh, M. (2015). Influence of socioeconomic disadvantages on mathematics achievement: A multilevel cohort analysis. *The Journal of Educational Research, 108*, 347-357. doi: 10.1080/00220671.2014.899956

Smith, M .S., & O'Day, J. (1990). Systemic school reform. *Journal of Education Policy, 5*(5), 233-267.

Thrupp, M. (2000). Sociological and political concerns about school effectiveness research: Time for a new research agenda. *School effectiveness and school improvement, 12*(1), 7-40.

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times.* San Francisco, CA: Jossey-Bass.

Townsend, T. (2001). Satan or savior? An analysis of two decades of school effectiveness research. *School Effectiveness and School Improvement, 12, 115-129.*

U.S. Department of Education. (2011). *Comparability of state and local expenditures among schools within districts: A report from the study of school-level expenditures.* Washington, DC. Author.

Ushomirsky, N., Hall, D., & Haycock, K. (2011). *Getting it right: Crafting federal accountability for higher student performance and a stronger America.* Retrieved from The Education Trust website: http://www.edtrust.org/sites/ edtrust.org/files/publications/files/Getting_It_Right.pdf

Ushomirsky, N., & Williams, D. (2015). *Funding gaps 2015: Too many states still spend less on educating students who need the most.* Retrieved from the Education Trust website: https://edtrust.org/wp.../FundingGaps2015_ TheEducationTrust1.pdf

Vermont Department of Education. (2009). *Roots of success: Effective practices in Vermont schools.* Retrieved from Vermont Department of Education Roots of Success website: http://rootsofsuccess.files.wordpress.com/2009/09/ rootsofsuccess_full2.pdf

Wagner, T. (2008). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need - and what we can do about it.* New York, NY: Basic Books.

Wiener, R. (2008). *No accounting for fairness: Equitable education funding remains elusive in Ohio.* Retrieved from The Education Trust's website:

http://www.edtrust.org/sites/edtrust.org/files/publications/files/NoAcctgforFairnessOH_0.pdf

Wiener, R., & Pristoop, E. (2006). How states shortchange the districts that need the most help. In The Education Trust's *Funding Gaps 2006.* Retrieved from The Education Trust website: http://www.edtrust.org/sites/edtrust.org/files/publications/files/FundingGap2006.pdf

Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist, 53*(3), 300-314.

Wilson, C. P., & Fergus, E. O. (1988). Combining effective schools and school improvement research traditions for achieving equity-based education. *Equity & Excellence, 24*(1), 54-65.

Witte, J. F., & Walsh, D. J. (1990). A systematic test of the effective schools model. *Educational Evaluation and Policy Analysis, 12*(2), 188-212.

Wood, G. (2004). A view from the field: NCLB's effects on classrooms and schools. In J. A. Banks (Series Ed.), *Multicultural Education Series: Many children left behind* (pp. 33-50). New York, NY: Teachers College Press.

Zhang, H. C., & Cowen, D. J. (2009). Mapping academic achievement and public school choice under the No Child Left Behind legislation. *Southeastern Geographer, 49*(1) 24-40.

Zigarelli, M. A. (1996). An empirical test of conclusions from effective schools research. *Journal of Educational Research, 90,* 103-110.