Spring 2017

Are 21st Century After School Programs an Effective Academic Intervention for Elementary School Students Attending High Poverty Schools?

Jennifer Scarpati
University of New Hampshire, Durham

ARE 21st CENTURY AFTER SCHOOL PROGRAMS AN EFFECTIVE ACADEMIC INTERVENTION FOR ELEMENTARY SCHOOL STUDENTS ATTENDING HIGH POVERTY SCHOOLS?

BY

JENNIFER SCARPATI

B.S., Boston University, 1999
M.Ed., Harvard University, 2003
CAGS, Rivier College, 2006

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy

in

Education

May, 2017

This dissertation has been examined and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Education by:

Dissertation Co-Director, Suzanne E. Graham, Associate Professor; Director, Division of Educational Studies

Dissertation Co-Director, Todd A. DeMitchell, John and H. Irene Peters Professor of Education; Professor, Justice Studies Program

Leslie J. Couse, Professor & Chair, Department of Education

Loan T. Phan, Associate Professor of Education

Rick Alleva, UNH Cooperative Extension - Youth & Family Team

On April 13, 2017

Original approval signatures are on file with the University of New Hampshire Graduate School.

Acknowledgements

Completion of my doctorate has been a long journey spanning over 10 years, the birth of two children, and a job change. Reaching my goal of completing my degree would not have been possible without the help of many key people. I'd like to first acknowledge and thank my dissertation co-chairs, Professors Suzanne Graham and Todd DeMitchell for their continuous help and support. Professor DeMitchell has been my advisor since my start in the program and has provided a wealth of encouragement and good advice along my journey. I would not have been able to complete this dissertation without the technical expertise of Professor Graham who guided and encouraged me through the depths of statistical analyses. I would also like to thank Dr. Leslie Couse, Dr. Loan Phan, and Dr. Rick Alleva for serving on my dissertation committee and for their valuable insights.

No one does this work alone. I'd like to thank my friends, especially the "Baby Day" crew for the support, encouragement and companionship over the years. A special thanks goes to Dr. Cori Sheedy for serving as my accountability buddy this past year. I would also like to thank all of the dedicated and talented professionals that I have worked alongside first at the Hampstead Central School and then at Amherst Street Elementary School. Your passion for helping children is inspiring and your support of my continued education has been invaluable. Finally, the amazing children of Nashua are what first inspired me on this journey and deserve their own thanks and acknowledgement.

In saving the best for last, I would like to thank my family for their endless patience and support. My wonderful husband, Dan, has taken on more than his fair share of the household load on many occasions so that I could attend classes and do my work. My beautiful daughters, Erin and Vera, have been understanding when I needed to work instead of play. Thank you to my dog, Jovie, for serving as my second alarm clock and getting me out of bed to write on many a morning. Finally, thank you to my unbelievably supportive parents, John and Gerry Oates, for their unwavering support of my education over the past 35 years. They lit the fire that has always pushed me to reach higher.

Table of Contents

List of Tables

List of Figures

Abstract

ARE 21ST CENTURY AFTER SCHOOL PROGRAMS AN EFFECTIVE ACADEMIC INTERVENTION FOR ELEMENTARY SCHOOL STUDENTS ATTENDING HIGH POVERTY SCHOOLS?

by

Jennifer Scarpati

University of New Hampshire, May, 2017

In America, millions of children are living in poverty and/or attending high poverty schools. These students, on average, score lower on assessments of math and reading achievement than their more affluent peers. The federal government budgets money each year towards interventions targeted to raise the achievement of students living in poverty. 21st Century after school programs have received billions of dollars over the past 20 years towards this goal.

This is a secondary analysis of students attending five diverse, high poverty elementary schools in New Hampshire between 2008 and 2013. Ordinary Least Squares multiple regression analyses were used to estimate the impact of participation and dosage in the 21st Century after school program on math and reading achievement for each year of data. Multi-level modeling for change was used to estimate the longitudinal impact of participation and dosage on achievement over time.

Results were mixed on the impact of participation on achievement in math and reading. Some years, a negative impact of participation was found; while other years there was no impact.

In some years, differential positive impacts of participation were found for students identified as Hispanic, African American, or having special needs. Mixed results were also found for the impact of program dosage on achievement, with no impact seen in some years, and a positive impact found in others. A differential negative impact was found for males in one year, and a differential positive impact of dosage found in other years for students identified as English language learners or Hispanic. While no longitudinal impacts were found of program participation or dosage on math achievement, impacts were seen for reading. Students that participated in the program at high doses over the course of three years were found to score higher on tests of reading achievement than non-participants.

These results provide mixed support for 21st Century after school programs as an effective academic intervention for students attending high poverty schools. This indicates a need for further study in this area. These results should be considered in the broader arena of evaluations of other federal interventions to aid students living in poverty in order to most strategically target federal funds. Finally, results should be considered in conjunction with other 21st Century program goals, including providing affordable, supervised care for students living in poverty during the after school hours.

Introduction

> "With limited resources and the new focus on academics, we must learn what works – especially for at-risk students" (Roth, Malone, & Brooks-Gunn, 2010, p. 321).

Many children in the United States grow up in families living in poverty and attend schools with high poverty rates (Laughlin, 2014). Students living in poverty score lower, on average, than their more affluent peers on national assessments of reading and math across the grade levels. Furthermore, students attending high poverty schools score lower, on average, on assessments of reading and mathematics, than students attending schools with lower poverty rates (NCES, 2011a; NCES, 2011b; Sirin, 2005). The federal government, as well as non-profit and charity groups, spends billions of dollars annually on interventions aimed at improving the academic achievement of students living in poverty (Currie, 2010). A popular intervention over the past 30 years has been after school programming (Dynarski, Moore, James-Burdumy, Rosenberg, Deke, & Mansfield, 2004). Many theorists posit that after school programs improve outcomes for students living in poverty by reducing their environmental risks, and boosting their resilience to stressors (Egeland, Carlson, & Soufe, 1993; Hirsch, Mekinda, & Stawicki, 2010; Masten & Coatsworth, 1998).

The limited academic achievement of millions of students living in poverty is an educational, sociological, economic, and moral concern for our nation. Furthermore, there is little consensus in the research community as to which intervention programs best address this problem (Darling-Hammond, 2000). With limited federal dollars available, it is critical that policy-makers know which programs are most effective at improving the academic success of students living in poverty and/or attending high poverty schools.

This study specifically focuses on students at-risk for academic failure in reading or math in the elementary years based on their qualification for free or reduced price lunch, a common proxy for poverty for school-aged children. This study investigates the effectiveness as an academic intervention for students living in poverty of the federally sponsored 21st Century after school programs in selected New Hampshire high poverty public schools. It contributes a deeper understanding of the effectiveness of one of the federal government's primary interventions to improve the academic success of students attending high poverty schools.

Poverty: A risk factor for academic failure

In the United States, there are many young children that are considered at-risk for negative life outcomes. Use of the term *at-risk* stems from the context of research into achievement or outcome differences for individuals, groups, or communities and the subsequent reforms and interventions created to address these differential outcomes. In the broadest sense, at-risk refers to a vulnerability or increased risk faced by an individual or group. Common in the educational literature is reference to at-risk students having a higher probability of failing academically or dropping out, or to students that face circumstances that could adversely affect their performance or ability to complete school (Moore, 2006). It is a broad term, used in many different ways, and there are many that caution against its use due to the potential to stereotype or stigmatize individuals or groups (Moore, 2006).

This study considers risk as encompassing those factors associated with diminishing the opportunities and likelihood of attaining academic success. Consequently, students themselves are not considered at risk; students may be exposed to factors that increase the risk of not attaining academic success.

The concept of risk is multidimensional and can include a single trauma, many small stressors over time, chronic factors, such as poverty, or a combination of these (Condly, 2006). Often, the effects of risks and stressors on development are cumulative, with adverse effects not seen until someone faces multiple risks. In addition, these effects are typically multiplicative in nature, with much larger impacts of additional risk factors once a child has exceeded their manageable load (Condly, 2006; Rutter, 1979).

Risks that negatively impact children include growing up in poverty, unstable family lives, and a history of trauma (Masten & Coatsworth, 1998; Smokowski, 1998). According to U.S. Census statistics from 2011, 22% of children lived in families with incomes that fell below the poverty line (Laughlin, 2014). Children experience poverty at higher rates than the general population (DeNavas-Walt & Proctor, 2014). In addition, children who grow up in poverty are more likely to experience chronic health problems, be exposed to violence, live in a dangerous neighborhood, and receive a poor quality education (Miller, 2003). Families living in poverty report many stressors in their lives, such as community violence, higher levels of interpersonal conflict, and higher ratings of maternal stress. These stressors can impact parenting style, students' self-confidence, and school readiness (Duncan, Yeung, Brooks-Gunn, & Smith, 1998; Posner & Vandell, 1994). These factors can increase a child's risk for poor academic and behavioral outcomes in school.

On average, children living in poverty do not perform as well in school or on standardized achievement tests as middle and upper class peers (D'Amico, 2001; Traub, 2000). Students living in poverty score lower on the National Assessment of Educational Progress

(NAEP) in Reading and Math in grades 4, 8, and 12 (NCES, 2011a; NCES, 2011b). Moderate negative correlations are seen, on average, between students' socioeconomic status (SES) and their academic performance; however, strong negative correlations are seen between school SES and overall school academic performance (Sirin, 2005).

The U.S. Department of Education (1992) has examined a variety of factors that can increase a student's risk of low achievement on standardized tests of achievement in reading or math, or of dropping out of school. The factors considered were demographics, family characteristics, parental involvement, academic history, student behavior, teachers' perceptions of students, and school characteristics. This longitudinal study found increased risk for students with certain demographics (minorities, low socioeconomic status), from single parent households, that were transient, who were retained, who had poor attendance or behavior or who didn't complete homework, and for students attending urban schools or schools with high minority enrollments. With many years of data showing the negative impact of poverty on academic achievement, policy-makers and educators have tried different interventions aimed at improving academic performance for students living in poverty.

Interventions: Building Resilience

The federal government and private or non-profit organizations often provide funding to assist children at-risk for academic failure in the hope of improving outcomes for these children. These programs include school-based interventions designed to help children who are educationally at-risk, mentoring, after school programs, targeted tutoring, counseling, family therapies, and school restructuring, to name a few (Currie, 2000; Darling-Hammond, 2000; Schorr, 1989; Smokowski, 1998; Weissbourd, 1997). All of these interventions claim some level of effectiveness, however, with limited amounts of funding available to service this population of children, it is important for government and school officials to make informed choices on how to best use these scarce resources in order to meet the needs of children in the United States.

The concept of resilience (Egeland et al., 1993; Hirsch et al., 2010, Masten & Coatsworth, 1998) was an underlying rationale for the expansion of after school programs in the 1990's. Resilience is a strengths-based, ecological paradigm for human development. In this paradigm, the risk or deficit does not lie within the child, but within the environment. Resilience is the ability to thrive in the face of obstacles and adverse circumstances (Gordon, 1996). Resilience leads to healthy development and positive outcomes, despite exposure to risks (Zolkoski & Bullock, 2012).

A person is not resilient or non-resilient. Rather, resilience should be viewed on a continuum, with some individuals having more or less capacity for resilience than others. The level of resilience can change over time and in different contexts (Condly, 2006; Woodland, 2016). Resilience in a situation represents an interaction between the individual's abilities, motivations, and support systems and the circumstances or environments they are facing. It is an ongoing process, with many opportunities to demonstrate resilience, or not, throughout one's lifetime.

All children possess some basic characteristics that help them manage life's stressors. When confronted with stressors in their environment, some children enact these traits more than others. Research in the field has shown which traits predict higher levels of competence in the face of adversity. These traits include social competence, problem solving, autonomy, and a belief in one's future (Masten & Coatsworth, 1998; Masten & Motti-Stefanidi, 2009). Factors that have shown in the research to help students better handle environmental stressors include fostering caring relationships, having high expectations for children, and providing opportunities for meaningful contributions and participation (Egeland et al., 1993).

Protective factors foster positive adaptation amidst significant adversity (Woodland, 2016). Building protective factors in children to promote competence, confidence, character, connections, and caring can enhance resiliency in children (Hirsch et al., 2010). Like risk, resilience is multi-faceted. It can be enhanced by fostering protective factors within an individual (such as persistence), within a family (such as strong parental involvement in school), or within the broader ecological context (such as involvement in church) (Condly, 2006; Masten & Coastsworth, 1998; Miller, 2003).

Key factors in building resilience in youth include providing opportunities for mastery, and relationships with positive adults and peers (Masten, 2007). Well-designed after school programs can serve as a protective resource for urban youth. After school programs can compensate for risks and build protective factors by providing extra-curricular activities, academic support, social opportunities, and exposure to supportive peers and adults (Zolkoski & Bullock, 2012). They can reduce youth's vulnerability by providing a safe environment during a high-risk time of the day. They can also reduce vulnerability of academic failure by providing tutoring and homework assistance. After school programs can reduce stressors for students by providing a structured, caring environment during the after school hours. Finally, after school programs can promote resilience by increasing students' access to resources such as adult mentors and enrichment programs (Woodland, 2016).

After school programs can serve to minimize risks for students living in neighborhoods that are poor or unsafe (Posner & Vandell, 1994). Programs provide a safe, supervised, structured learning environment with adult supervision (Mahoney, Lord, & Carryl, 2005).

Having this environment serves to decrease student risk through limiting exposure to neighborhood risks, such as drugs or violence (Lord & Mahoney, 2007).

In addition to minimizing risk by limiting exposure to the neighborhood, after school programs provide opportunities for students to improve academically. They can provide a structured time and location for homework completion, as well as instructional support to complete it successfully (Cosden, Morrison, Gutierrez, & Brown, 2004). Beyond homework assistance, many after school programs provide individual or small group tutoring in core academic subjects (Vandell, Reisner, & Pierce, 2007).

After school programs can be a forum to promote these protective factors through exposure to competent adults who can serve as role models and provide emotional support (Pierce, Bolt, & Vandell, 2010). Providing students with structured social interactions promotes social skills, as academic support promotes competence (Auger, Pierce, & Vandell, 2013). Exposure to consistent quality interactions with adults and peers and developmentally appropriate patterns of activities (McHale, Crouter, & Tucker, 2001) can aid in building healthy growth and development (Vandell, et al., 2005).

Furthermore, students who participate in quality activities after school have been shown to be better socially adjusted and to make better choices during unstructured after school times (Posner & Vandell, 1999; Sanger, 2011). This speaks to long term gains of program participation, as students build positive self-images and supportive peer groups. These positive affiliations can help students stay on a positive developmental trajectory (Hirsch et al., 2010).

After school programs build resilience by providing enrichment opportunities in order to “level the playing field” (Posner & Vandell, 1999). After school programs can offer worthwhile activities for students as alternatives to watching TV or “hanging out” (Mahoney et al., 2005).

After school programs can provide access to extracurricular and enrichment activities that students may not otherwise have access to due to limitations of participation fees or transportation (Posner & Vandell, 1999). Students living in poverty often have an "experience differential" with more affluent peers. According to the 2011 Census, 55% of children participate in at least one extra-curricular program, such as a sport, club, or lesson (DeNavas-Walt & Proctor, 2014). Children living in poverty are less likely (by 15 to 20 percentage points) to participate in one of these activities (Laughlin, 2014). Competing household responsibilities, inability to pay fees, cultural barriers, and transportation are some reasons for this participation gap (Laughlin, 2014). This difference in the quality of experiences outside of school has a significant impact on academic outcomes (Sanger, 2011; Shernoff, 2010). Helping to close the experience differential through quality after school programs would bolster resilience for children living in poverty (Katz, 1997).

The paradigm of resilience runs in contrast to a deficit model of development, which views after school programs as keeping kids off the street and out of trouble (Shonkoff & Phillips, 2000). The resilience model aligns with the loftier goals of after school programs, such as academic and social growth. The U.S. Department of Education guidelines for 21st Century after school programs state their aim to enhance positive youth development through development of multiple domains – such as cognitive, social, and physical - in an interconnected way; acknowledging that youth have strengths and prior knowledge on which to build; and seeing youth as active agents in their own growth (U.S. Department of Education (U.S. DOE), 2003). While the 21st Century after school programs are the focus of this study, after school programs have a long, diverse history in the United States.

History of After School Programs

Beginning in the 1980's, after school programs became an increasingly popular program for school-age children. As more mothers began working outside the home, there was more demand for after school care. In addition, crime was on the rise throughout the U.S., prompting concerns about after school supervision and safety for children. Finally, as schools entered the era of accountability, more attention was given to after school programs as an intervention to improve the academic performance of children in underachieving schools (Dynarski et al., 2004).

After school programs have existed in various forms, run by various organizations such as the Boys & Girls Club of America, the YMCA, and local churches and city and town recreation departments for over 100 years (Boys & Girls Clubs of America, 2015; YMCA, 2015). However, there was a shift in the 1990's toward federal sponsorship of after school programs, and an increased linkage of these programs to the public schools. In fact, between 1987 and 1999 the percentage of public schools offering after school programs tripled, from 16 percent to 47 percent (James-Burdumy, Dynarski, & Deke, 2007).

Federal sponsorship of after school programs was formalized in 1994 when Congress sponsored the 21st Century Community Learning Centers funding which sought to open up public schools for broader use by the community during after school hours and weekends (James-Burdumy et al., 2007). These programs were funded through the Improving America's Schools Act, a federal discretionary grant, at a level of approximately one million dollars by 1996 (Dynarski, 2015). By 1998, these programs gradually became more focused on academic, enrichment, and recreational programs for at-risk students, and funding had increased to $40 million dollars (James-Burdumy et al., 2007). There have been shifts in after school program goals and features over time, particularly since the creation of the federally funded 21st Century Community Learning Centers. After school programs, in general, now serve more low income students, are more likely to be located in a school rather than a church or community center, and have a heavier focus on academics (Vandell et al., 2005).

The landscape of federally sponsored after school programs shifted again in 2001 with the passage of the No Child Left Behind legislation, which is encompassed in the Elementary and Secondary Education Act, Title IV, Part B. After school programs were specifically tied into the legislation with the following goal:

> This program supports the creation of community learning centers that provide academic enrichment opportunities during non-school hours for children, particularly students who attend high-poverty and low-performing schools. The program helps students meet state and local student standards in core academic subjects, such as reading and math; offers students a broad array of enrichment activities that can complement their regular academic programs; and offers literacy and other educational services to the families of participating children, (U.S. Department of Education, 2003).

This legislation also shifted the funding of these programs from federal discretionary grants to state block grants. With the legislation, funding for 21st Century after school programs grew to $1 billion by 2002, with a plan for funding to increase $250 million per year until it reached $2.5 billion in 2007 (James-Burdumy et al., 2007). It is estimated that the average cost per student to attend a high quality urban elementary school program is $4,320 (Grossman, Walker, & Raley, 2001).

Along with this significant level of federal investment in after school programs came questioning of the effectiveness of after school programs to improve outcomes for students. Based on poor results from some initial program evaluations, (primarily The Mathematica study (Dynarski et al, 2004)), President Bush recommended deep cuts in the program in his 2004 budget proposal, decreasing funding from nearly $1 billion annually down to $600 million in just one year. The President proposed to instead direct these funds to Title One and the Individuals with Disabilities Education Act (IDEA), two other programs with the aim of improving educational outcomes for students at-risk for academic difficulties. The proposed cuts were reviewed in the "Investment in After-School Programs Hearing" under the auspices of the Senate Labor, Health and Human Services and Education and Related Agencies Committee on Appropriations in May of 2003. These hearings ultimately resulted in partially restored funding, after receiving broad public support from many Congressmen, police associations, and youth advocates (Committee on Appropriations, 2003).

The 21st Century After school programs have served many students across the country over the past 20 years. In 2004, at the peak of funding, 1.3 million students were being served in 7,000 schools in 2,250 districts across the U.S. Appropriations have remained level between 2012 and 2015, with federal funding of just over one billion dollars per year. From this, grants to states range from just over $5.6 million to over $130 million (U.S. Department of Education, 2015). The federal government has laid out some specific goals towards which this funding is to be directed.

Program Goals of 21st Century After School Programs

After school programs are held accountable for a wide-variety of measures, including ensuring the physical and emotional safety of students, helping students develop positive relationships with staff and students, achieving high participation rates, building skills in specific areas, and involving the community (Connell & Gambone, 1999). The 21st Century Community Learning Centers grant has a target audience of students (elementary through high school) who attend high poverty and/or persistently low-achieving schools in the United States (U.S. DOE, 2003). The request for proposals for the grant asks applicants to define students' risk factors using the following measures: poverty rates, percentage of students with limited English proficiency, percentage of students receiving Title One services, drop-out rates, the achievement gap, and the education level of the community (U.S. DOE, 2003). While these factors are used to determine which schools qualify to receive these grants to operate after school programs, the programs that are offered are open to all students attending the school – not just those deemed "at-risk."

With this population defined as having a greater risk for poor educational performance, the mission of the program is to expand academic enrichment opportunities for children to help them meet local and state academic standards in reading and math. This direct and primary focus on academics represents a shift from the original goals of after school programming which focused more on student safety, delinquency prevention, and boosting youth resiliency. While academics are the current primary focus, programs may also provide youth development, arts, and recreational activities to enhance the academic program. These outcomes are expected from participation in a safe and supportive environment, with positive youth-adult interactions, academic enrichment that complements the school day, and the offering of family literacy activities. Specific activities recommended for implementation through the grant include remedial or academic enrichment activities to improve academic achievement, activities to improve the language skills and academic achievement of students with limited English proficiency, recreation activities, tutoring, and mentoring (U.S. DOE, 2003).

The U.S. Department of Education lists as its first measurable objective of the program, "Increasing percentages of students regularly participating in the program will meet or exceed state and local academic achievement standards in reading and math." According to the most recent data available on their web site, the goal is for 48.5% of participating students to show an increase in their math and English grades; in 2010, 38.4% of students met this goal for math, and 40% for English grades. The target performance indicator for improving the percentage of students scoring proficient or above on state assessments is 45%. In 2010, 26.5% of participating students met this goal in reading, and data was not available for math. Other indicators of student progress were also considered, including homework completion, class participation, and behavior. This data was gathered by surveys of teachers of participating students. The target goal was for 90% of participating students to increase their homework completion and class participation. In 2010, 74% of participating students showed progress in this area. For behavior, the goal was for 75% of students to show improvement in their classroom behavior. In 2010, 69% of participating students were rated by their teachers as showing improvement in this indicator (Elementary & Secondary Education Act (ESEA), 2011). The impact of participation on academic achievement is explored in this study for five 21st Century programs in New Hampshire.

Limitations of Prior Research

Many evaluations of after school programs have been conducted over the past 20 years. Very few of these studies have considered the impact of dosage on outcomes. In fact, a study done by the Harvard Family Research Project found that nearly 70% of program evaluations did not consider dosage in their research design (Chaput, Little, & Weiss, 2004). Researchers (e.g., Gardner, Roth, & Brooks-Gunn, 2009), have called for more research on differential effects of after school programs on students of varying socioeconomic status, emphasizing the need to study differential impacts of dosage for these groups.

Chaput et al. (2004) define three ways of considering dosage. They suggest considering program intensity, defined as the amount of time a student attends the program within one year; program duration, defined as the total amount of participation over a designated time period; and breadth of participation, defined as the variety of activities in which the student participated.

In addition to proper consideration of program dosage, research on after school programs has been criticized for its shortage of longitudinal studies of program effects (Roth et al., 2010). Another critique of the existing research base includes a lack of appropriate matching techniques in non-experimental studies. Researchers have too often compared participants to non-participants without proper statistical matching techniques (Scott-Little, Hamann, & Jurs, 2002). These methodological weaknesses in the research base necessitate further study of after school programs.

Overview of the Research Study

This study utilizes a large data set, and allows inferences to be made regarding after school programs that can influence policy regarding effective academic interventions for elementary students attending high poverty schools. Participation and achievement data is studied for five consecutive years, both cross-sectionally and longitudinally. This study investigates the level of usage at which after school programs might show maximum academic benefit, and for which categories of students (based on gender, race, and status as a special learner) impact is seen. It refines our understanding of the impact of a specific intervention with a 25-year history in U.S. policy.

Dosage is considered in multiple ways in this study. The intensity and duration of participation is considered. In this study, intensity of participation is measured in days attending the program per year. Duration of participation is measured as number of years attending the program within the range of 2008-2013, as well as the cumulative days of attendance during this time period. A limitation of this study is that the breadth of program dosage is not considered. The specific activities that each child participated in during their time in the program is not known in this study. Furthermore, it is recommended that intensity be further examined as the number of hours spent in the program each week (Dietel, 2009). While this is rarely done in most research, Dietel (2009) appropriately notes that a day of attendance for one student could vary from 30 minutes to 3 hours. This lack of nuance in consideration of intensity is a limitation of this research.

A limitation of this study is that it is a secondary analysis of existing data rather than an experimental or quasi-experimental study. Differences between participating and non-participating students is balanced through statistical measures to create demographically comparable groups based on known measures. However, since students were not randomly assigned to the treatment or control group, there may be factors that led some families to choose the 21st Century after school programs that also impact student achievement, that are not included in this study.

Another limitation of this study is the inability to comment on differences in quality between the five after school programs included in this study. The federal government monitors the use of 21st Century grants and quality of program implementation through the use of surveys, self-assessments, and on-site evaluations. This information is compiled in the Continuous Improvement Process for After School (CIPAS) Visitation Team Report completed every three years. While CIPAS reports were conducted for Nashua for the time period of this study, results are presented for the program as a whole. Differences in quality between sites is not included in the commentary (New Hampshire Department of Education (NH DOE), 2011).

The specific research questions addressed in this study are as follows:

<u>Research Questions</u>

1. Do students who attend 21st Century After school programs in selected elementary schools in Nashua, New Hampshire perform better academically, on average, based on

scores on the grades 3-6 Math and Reading NECAP tests, than students in the same schools who do not attend these after school programs?

2. Do the effects of attending the 21st Century After School program in selected elementary schools in Nashua, New Hampshire differ based on student characteristics such as race, gender, free/reduced lunch status, status as a student with a disability or English language learner, or grade of attendance?
3. Do the effects of attending the 21st Century After school program in selected elementary schools in Nashua, New Hampshire differ based on student attendance in the program within an academic year (# of days attended)?
4. Do attendees show greater growth in NECAP scores over time than non-attendees? Do the effects of attendance differ based on cumulative attendance over the elementary school career?

Importance of the Study

Answering these questions has real practical and policy implications for educational leaders. The federal government has invested billions of dollars over the past 20 years on after school programs. It is critical to know if this investment is yielding results for the students it aims to help, or if this money would be better targeted in a different manner to help children at risk. Furthermore, insight into the nuances of who is best helped and what dosages of participation are required to see results will help leaders and policy-makers better design programs to improve results for the children they serve. With limited government funds and millions of children living in poverty in the United States, this is a study of critical importance.

Consideration of the existing research base is a critical first step in examining the effectiveness of after school programs as an intervention for students attending high-poverty schools.

Definition of Terms

After School Program - 21st Century Community Learning Center programs or other similar programs. These public school-based programs meet regularly (typically every day after school), target students attending low-income schools, are publicly subsidized and offered at low cost to participants, and provide a combination of academic and extracurricular activities (Gardner, et al., 2009; Leos-Urbel, 2015).

Dosage (Duration) – The number of years a student attended the after school program.

Dosage (Intensity) – The number of days a student attends the after school program within an academic year or over time.

Math Achievement – Student's scaled score on the New England Common Assessment Program (NECAP) math assessment

Participant – A student that attends the after school program for 30 days or more during the academic year.

Poverty – Qualification for the federal free or reduced price lunch program

Protective factor – A condition or attribute of a person, family or community that is associated with a reduced negative impact of stress on life outcomes

Reading Achievement – Student's scaled score on the New England Common Assessment Program (NECAP) reading assessment

Resilience - The ability to thrive in the face of obstacles and adverse circumstances (Gordon, 1996), leading to healthy development and positive outcomes, despite exposure to risks (Zolkoski & Bullock, 2012).

Risk – An increased vulnerability faced by an individual or group. In this study, risk refers to students that face circumstances of living in poverty that could adversely affect their academic performance or ability to complete school (Moore, 2006).

Literature Review

As after school programming has evolved over its history from youth safe haven to key federal intervention, so has the study of its effects. Researchers have evolved from considering the effectiveness of after school programs at keeping students out of trouble to consideration of the impacts of programming on positive youth development and academic achievement. As a key initiative aimed at fostering characteristics in youth that build resilience in the face of the stressors of poverty, examining after school programs' effectiveness at promoting positive development and increasing academic performance is of key importance.

Who Attends After School Programs?

Are certain types of children or families more likely to access formal after school programs versus other care arrangements? Children whose mothers work more hours, particularly in a traditional schedule, and children of single parents are more likely to attend after school programs. These are bigger predictors of attendance than family income or level of parental education (Vandell et al., 2005). Pierce, et al. (2010) found that 23% of students (K-5) who were in non-parental care after school attended an after school program, and that they attended an average of 7.7 hours per week. Children living in poverty are less likely to participate in enrichment activities, such as sports, lessons, or clubs, than more affluent peers (Miller, 2003). This leads to an experience differential for children living in poverty.

Participation in formal after school programs has been found to be higher at the lower elementary grades, with lower participation rates starting in grades four and up (Pierce et al., 2010; Roth, et al., 2010). Participation was found to drop in half between elementary and middle school years (Roth et al., 2010). Barriers to participation for students can include transportation, program cost, language barriers, lack of interest or rapport with staff, and demands of other home responsibilities (Roth et al., 2010). These barriers are significant to consider in any study of effects of after school programming, as they could limit the representativeness of the sample.

Programmatic Goals

After school programs often have multiple goals for their students. Stated goals of the 21st Century after school program include academic growth in reading and math, increased homework completion and class participation, and decreased behavior problems (ESEA, 2011). Other programs have included improved student attendance as a program goal (After School Alliance, 2015). Some after school programs have collected measures of mediating effects on achievement and behavior, such as student motivation, work habits, and social skills (Vandell et al., 2005). The next sections consider the research base on these potential measurable outcomes of after school program participation.

Academic Impacts

Studies of the academic benefits of attending after school programs have shown mixed results. While most studies have revealed positive effects on academic growth, some have shown no effect, and others have shown a negative impact (Cosden, Morrison, Albanese, & Macias, 2001; James-Burdumy et al., 2007; Lauer, et al., 2004; Mahoney, et al., 2005; Vandell & Corasaniti, 1988; Vandell et al., 2006).

Various outcomes have been used to measure academic growth in studies of after school programs, and results can vary based on the outcome measure used. A meta-analysis of studies of after school programs found a positive and statistically significant impact on grades and achievement test scores for students who attended after school programs (Durlak, Weissberg, & Pachan, 2010). Students' report card grades are a common measure that has been used in studies of after school programs. One study by Auger, et al. (2013) showed that consistent participation in after school programs closed the math achievement gap between low income and upper income students, as measured by their course grades. A study of various after school programs (40) in the midwest showed that after school programs that employed positive staff resulted in improved grades in math and reading for participants (Pierce, et al., 2010).

Similarly, Cosden, et al. (2004) found that students who attended a structured after school program that provided ample adult contact in grade four, had higher math and reading grades in grade six than did students who did not attend the program in grade four. At the middle school level, Shernoff (2010) found that students who attended a school-based after school program for one year had higher English grades that those who didn't attend, controlling for demographics and baseline grades. Posner and Vandell (1999) found that students who participated in extracurricular and enrichment activities after school had better grades than students who spent time watching TV or "hanging out" after school, controlling for demographic factors. Similarly, Petit, Laird, Bates, & Dodge (1997) found that students who attended formal after school programs had higher grade point averages and achievement test scores than did students in self-care.

Another study by Posner and Vandell (1994) reviewed educational outcomes of 216 third graders in Milwaukee. The 34 students in this sample who attended enrichment lessons had better grades than those who did not participate. Similarly, a study by Schinke, Cole, and Poulin (2000) examined a pilot program at Boys & Girls Clubs in 5 cities. All served poor, minority middle school-aged students. Students who attended Boys & Girls Clubs that offered the pilot educational enhancement program showed modest gains in grades compared to matched peers who attended other Boys & Girls Clubs, or who did not participate in any after school program. Students in these comparison groups had grades that, on average, stayed flat or declined from baseline to the follow-up checkpoints at 18 and 30 months. Students who attended any Boys & Girls Club Center fared better than students who did not attend any program. While these studies certainly show promising findings for after school programs, the use of grades as an outcome measure is somewhat subjective in that grading philosophies can vary from school to school and teacher to teacher.

Many studies of the academic impact of attending after school programs have used standardized test scores in addition to, or in lieu of students' grades. Mahoney, et al. (2005) found that first through third grade students who attended 21st Century after school programs at three different schools in a poor, New England city had higher reading scores, as measured by the Developmental Reading Assessment (DRA), than students in other after school arrangements. In further analysis, Lord and Mahoney (2007) found this to be particularly true for students in high crime neighborhoods. In the Mathematica evaluation of the 21st Century after school programs, James-Burdumy et al. (2007) found that students with low initial reading test scores who attended the programs demonstrated greater gains in reading than those who did not attend.

Lauer et al. (2004) conducted a meta-analysis of 53 studies published between 1980 and 2004 that focused on the academic impact of attending an after school program on low achieving, low income students. Effect sizes for reading achievement scores ranged from .06-.13, with the strongest results shown for students at the early elementary level (effect size of .26). Effect sizes for math ranged from .09-.17, with the strongest results shown for students at the middle and high school levels (effect size of .44). The authors did find a significant positive effect ($p<.05$) on reading achievement for students considered at-risk that attended after school programs. Lauer's meta-analysis included only studies that had a comparison group. However, it does not reference the standards used to ensure that the comparison group was a match with the treatment group on any number of key factors.

Test scores were one outcome measure used by Cosden, et al. (2001) in their evaluation of the Gevirtz Homework Project in Santa Barbara, California. Their evaluation found that students who participated in the project regularly had higher SAT 9 scores than students that did not participate. Similarly, a study by Sheldon, Arbreton, Hopkins and Grossman (2010) found that third and fourth graders (n=381) who attended after school literacy-focused programs showed greater gains in standardized reading assessments than those students who did not attend. A study by Vandell and Ramanan (1991), found that children who were home with single, stay-at-home moms after school had lower scores on the Peabody Picture Vocabulary Test (PPVT) than children of similar income levels and ethnicities who attended after school programs.

Vandell et al. (2005, 2006, 2007) studied the intermediate and long-term outcomes for students who attended high quality after school programs at 19 elementary and 16 middle schools. The study included 1,017 elementary and 540 middle school aged students. The study looked at patterns of after school care, including students who are unsupervised, supervised at home, participating in after school activities (sports, lessons, etc.), and after school programs. Their work showed that low income minority students who regularly attended high quality after school programs for two years showed higher scores on standardized math tests compared with similar students who were unsupervised after school (12% higher scores, effect size = .50). The impact was even greater for students who attended after school programs and participated in after school sports or lessons (20% increase, effect size = .73). Students who had low supervision after school had higher self-reports of misconduct, and poorer teacher ratings of academic performance, work habits, social skills, and task persistence (effect sizes range from .17-.61) than those who attended after school programs.

Other studies have looked at indirect ways in which after school program participation may impact academic performance. Marsh (1992) theorized that attending after school programs prevents academic backsliding through his findings that students who attended after school programs demonstrated an improved academic self-concept and commitment to school, as well as improved motivation, study skills, and sense of personal responsibility. Similarly, Cosden, et al. (2004) found that attending after school programs resulted in improved study skills and homework completion for at-risk students. Beck (1999) also found that attending after school programs resulted in improved homework completion for students of lower socioeconomic status (SES), and showed increased confidence in their academic abilities. Providing students with the supplies, space, and support to do homework can serve to level the homework playing field for low SES students (Glazer & Williams, 2001). Finally, Cosden et al. (2001) found that attending the Gevirtz Homework Project resulted in higher ratings of self-efficacy and higher aspirations for the future. The study also found that attending the program resulted in higher teacher-ratings of student effort for students with limited English proficiency.

While the majority of studies have found at least a small positive impact for students attending after school programs, this has not been universally true. In the Mathematica evaluation of 21st Century after school programs, James-Burdumy et al. (2007) found that there was no difference in grades or standardized reading test scores for students who attended the 21st Century programs and those who did not. It should be noted that the Mathematica study has been highly criticized by scholars in the field. For example, students in the treatment and control groups had different mean reading scores at baseline (40th percentile for treatment, 50th percentile for control), but the authors did not control for this difference in their analysis. They also did not look at the changes in scores over time (growth), or test the significance of test score changes (Shernoff, 2010; Vandell et al., 2005). The study was also criticized for the fact that the authors did not properly account for the high levels of attrition and low levels of student attendance in the programs, as well as the inconsistent quality of the programs (Durlak et al., 2010). For example, students at the elementary level only attended 49 days, on average, in the first year of the study, and 32 days in the second year (Dynarski et al., 2004; James-Burdumy et al., 2007).

Finally, Vandell and Corasaniti (1988) found that students who attended center-based after school programs had lower grades and standardized test scores than students who were home alone, with a sitter, or in parental care after school. It is important to note that this study was conducted with a sample of White middle class students in a suburb of Dallas, as this is a different population than was studied in many of the above studies. The other difference worth noting was that none of the programs in this study were school-based.

While researchers have many important findings with regard to the effectiveness of after school programs in different settings, there are several gaps in the studies. Most of the studies are not longitudinal, or consider no more than two years of data (e.g., Vandell et al., 2007). Furthermore, the longitudinal studies that have been conducted still primarily use only pre and post-test measures (Cosden et al., 2001; James-Burdumy et al., 2007). These studies were primarily designed to compare students' achievement (measured by grades or achievement testing) before and after the intervention. Many of the studies did not properly account for matching control students to those in treatment through experimental design or appropriate statistical matching techniques. Furthermore, few of the studies examined the impact of attendance in the programs for multiple years, (Durlak, et al., 2010). An exception to this is Huang, Gribbons, Kim, Lee, & Baker's (2000) study of LA's Best after school programs, which does effectively model students' academic growth trajectories based on attendance in after school programs over time. With such variation in quality of research design, it is challenging to know if these programs are an effective intervention. The present study addresses these challenges by using appropriate statistical modeling techniques with a longitudinal data set in which students in the treatment and control group are statistically matched through the use of propensity scores.

Other Impacts of After School Programs

Academic benefits are a critical area of study related to after school programs, however, many researchers have studied other important outcomes such as student behavior, school attendance, and levels of prosocial behavior. For example, Auger, et al. (2013) found that attending after school programs resulted in improved attendance, behavior, work habits, and prosocial behavior for students, on average. Reduced absences has been an outcome measure noted by many studies of after school programs (After School Alliance, 2015; Huang, et al., 2000; Johnson, Zorn, Williams, & Smith, 1999; Vandell & Pierce, 1999). Durlak et al.'s 2010 meta-analysis of 75 studies on after school programs found positive and statistically significant impacts on youth in the areas of self-perception, bonding to school, positive social behaviors, and youth's personal and social well-being. These findings were noted in programs that had goals related to fostering personal and social development. Programs with a purely academic focus were not included in the meta-analysis (Durlak et al., 2010).

Many studies have examined the impact of attending after school programs on students' behavior in school. James-Burdumy, Dynarski, and Deke (2007, 2008) found that boys who attended after school programs demonstrated increased behavior problems, as did students with above average behavior problems at the start of the study.

Alternatively, Marshall, et al. (1997) found that boys from low income families in after school arrangements with greater supervision and adult interaction demonstrated fewer externalizing and internalizing behaviors than did similar students in unsupervised arrangements, or arrangements with higher levels of TV watching. Sixth graders who were primarily in unsupervised care after school had higher rates of devious behavior and lower ratings of self-esteem than did those in structured settings (Galambos & Maggs, 1991). Similarly, low-income students who were home alone after school, or home with a single mom, were rated by their mothers as having more anti-social behaviors and more peer conflicts than were children who attended after school programs (Vandell & Ramanan, 1991).

Similarly, Posner & Vandell (1999) found that students who spent more time in structured extracurricular activities had fewer teacher reports of poor behavior than those who did not. In Vandell et al.'s (2005, 2006, 2007) study of student participation in after school programs, they found that students with higher rates of program attendance saw a decline in teacher reports of aggressive behavior with peers (effect size of -.17), and gains in teachers' reports of pro-social behavior (effect size of .17) compared to students with low levels of supervision after school. At the two-year mark (Vandell et al., 2007), declines in aggressive behavior had an effect size of -.34. Scott-Little et al. (2002) found after school attendance positively associated with increases in positive socioemotional functioning, and decreases in negative student behaviors. Similarly, Johnson, et al. (1999) found a reduction in school suspensions for students attending after school programs.

Petit et al. (1997) found that elementary aged children who spent four or more hours per week in self-care demonstrated poorer behavioral adjustment in grade six. A 1994 study by Posner and Vandell found higher conduct grades for students who were supervised after school (versus “hanging out”), who were engaged in academics or enrichment activities after school, and who spent more time with adults after school. These characteristics were also associated with lower levels of anti-social behavior in this study.

Besides academic and behavioral measures, many researchers have studied the impact of attending after school programs on various student characteristics. These characteristics are typically rated by a teacher, parent, after school instructor, or by the students themselves. Vandell et al. (2005) found that high program attendance in after school programs had a positive impact on students’ work habits (effect size = .19), social skills (effect size = .19), and task persistence (effect size = .23) based on teacher reports.

Alternately, James-Burdumy et al. (2007, 2008) found that students who attended 21st Century after school programs had lower rates of homework completion, and lower ratings of class effort and preparation by their teachers. Cosden et al. (2004) found that students who attended after school programs had lower levels of parental participation in the homework process, which resulted in lower levels of students’ perception of parental support for education. On the other hand, the same authors found that participation in these programs led to higher levels of school bonding for students, and increased teacher perception of students’ effort and capabilities.

Mahoney, et al. (2005) found in their work that participation in 21st Century after school programs resulted in higher teacher ratings of students' motivation. Pierce, Bolt, and Vandell (2010) found improved social skills for boys who participated in after school programs, as rated by their classroom teachers. Similarly, Cosden et al. (2004) found that students who attended after school programs had an improved connection to school, higher self-esteem, and expanded social networks. The Mathematica study (James- Burdumy et al., 2007) reported that students who attended the 21st Century programs reported feeling safer than students who did not attend. Posner and Vandell (1994) found that students who participated in after school enrichment activities had higher ratings of peer relations, emotional adjustment, and work habits, as rated by their teachers.

While Vandell and Corasaniti (1988) found in their study of White elementary aged students in a Dallas suburb that students who attended after school care centers were rated lower socially by their peers, Shernoff (2010) found that after school participation was linked to working well with others, better psychosocial adjustment and social skills, improved social competence, and better relations with peers and adults, controlling for demographic factors. These results were demonstrated, on average, for middle school students who attended a school-based after school program for one year.

While the present study does not directly measure the impact of after school program participation on attendance, behavior, or social skills, it is important to consider the broader impacts of these programs found in the literature, beyond academics. Academic improvement is the primary goal of the federal 21st Century programs; however, all of these other factors impact a child's school performance as well. While they are beyond the scope of this study, these are impacts that could be studied in the future using a longitudinal analysis such as the one being conducted in the current study.

Variation in outcome measures: Student Characteristics

A common thread that appears throughout studies on after school programs is that program effects vary for different types of students, different types of programs, and different rates of participation. Considering the impact of student demographic factors is critical for matching programs and students for maximum impact. The most common demographic factor discussed in the literature is socioeconomic status. Many studies highlight the vulnerability of low-income, urban children during after school hours, and the possibility of formal programs for buffering these risks (Marshall et al., 1997). A survey of parents living in large cities found that families with lower incomes and lower education levels were less satisfied with their after school arrangements and options than families of higher socio-economic status (Weitzman, Mijanovich, Silver, & Brazill, 2008).

Students from families of lower socioeconomic status have a higher chance of experiencing self-care at an early age. Students from low SES families who had higher rates of self-care in grade one demonstrated more externalizing problems, on average, in grade six than similar students not in self-care (Petit et al., 1997). In addition, low income students who were primarily unsupervised after school demonstrated more externalizing and internalizing problems in school than middle or upper class students who were primarily unsupervised after school (Marshall et al., 1997). Similarly, students living in high crime neighborhoods have seen more positive outcomes from participation in after school programs (Lord & Mahoney, 2007) than students living in low crime neighborhoods.

McHale, Crouter, and Tucker (2001) found that students from lower socioeconomic families tended to participate in less structured free time activities than did wealthier families. Posner and Vandell (1999) found that low income students in after school programs spent 20% of their time on structured extracurricular and enrichment activities versus 4% of time for students not in programs. Vandell et al. (2005) found that low income students typically have fewer options available to them for constructive activities, resulting in a greater amount of unsupervised time. Vandell and Ramanan (1991) found negative impacts on behavior and academics for students living in single parent, low income families. Positive impacts of after school programs have not generally been found for students from higher socioeconomic backgrounds. These students typically have other academic or extracurricular offerings available to them (Cosden et al., 2004).

Specifically, benefits have been found for students from lower socioeconomic backgrounds who receive homework assistance in an after school program. This has been shown to positively impact academic success for these students (Cosden et al., 2004). Students whose parents can't or are unwilling to assist with schoolwork show the greatest benefits from participation. Benefits of homework completion have also been noted for other groups participating in after school programs. Structured homework help has shown greater advantage for upper elementary, middle school, and high school students (Cosden et al., 2004). In addition students with limited English proficiency who received homework help were viewed by their teachers as demonstrating greater effort (Marshall, 1997). The present study examines whether the impact of attendance in 21st Century after school programs varies for students of different socio-economic levels, and additionally, examines whether the impact of participation varies over time.

Several researchers have studied whether after school programs impact boys and girls in different ways (e.g., James-Burdumy et al., 2007; Pierce, et al., 2010). Results are mixed on the impact of after school program attendance for boys. Pierce, et al. (2010) found that boys showed evidence of improved social skills, while girls did not show any change in this area. In addition, Marshall et al. (1997) found that low income, urban, elementary aged boys in after school care demonstrated fewer externalizing behaviors than similar students who were unsupervised after school. However, James- Burdumy, et al. (2007) found that boys who attended the 21st Century after school programs showed worse behavior in school than boys who did not attend the programs. The present study also examines if participation in 21st Century after school programs impacts boys' and girls' academic achievement in different ways.

In Scott-Little et al.'s (2002) meta-analysis of 23 evaluations of after school programs, it was found that the most positive outcomes were found for children in the early elementary grades (one to four). In the work of Pierce, et al. (2010), they note that different program characteristics differ in their importance for different ages. For example, a diversity of activity offerings was associated with higher grades and work habits for older students, but not younger. The present study examines whether or not the grade level at which a student participates impacts any program effects on academic achievement. Consideration of program characteristics that differentially impact student learning at different ages is beyond the scope of the present study, but may be considered in follow-up studies as a result of findings from this study.

Another important consideration is students' skill level prior to beginning the after school program. James- Burdumy, et al. (2007) found that students with poor behavior who attended after school programs did worse than similar students that did not attend. This could indicate that after school programs are not beneficial for students with significant behavioral problems, or that a special type of program is necessary. On the other hand, in the same study, students with lower reading scores at the start of participation ended up with higher reading scores after program participation than did similar students who did not participate. Scott-Little et al.'s (2002) meta-evaluation found, on average, that students demonstrating the lowest academic performance levels showed the greatest gains from participation in after school programs. The same held true for English language learners (Huang et al., 2000). The present study examines whether status as an English Language Learner or student with a disability effects the impact of participation in 21st Century programs on academic achievement.

Examining the differential impact of after school programs on different students is critical for policy makers to understand in order to target the correct programming for students to maximize benefits. While the studies referenced above give some valuable information about who benefits from after school programming, none of them utilize the longitudinal approach that is used in the current study. Growth trajectories are modeled for students based on student characteristics such as socioeconomic status, gender, and status as a student with a disability or English language learner. Examining data over time also allows me to examine at what grade levels participation results in maximum academic effects, as well as whether the impact of program participation on growth in achievement differs based on student characteristics such as socioeconomic status, gender, or status as a special learner.

Variation in outcome measures: Program Characteristics

While differences in student characteristics matter, so do differences in characteristics of after school programs. Miller (2003) found that program quality explained between 27 and 47% of variance in youth outcomes. Higher quality programs had higher ratings of student engagement, and better youth outcomes. Supportive adult and peer relationships was the most important indicator of program quality. This includes staff-student interactions and peer interactions. Programs rated as providing supportive environments for students were associated with higher reading and math scores in a large-scale study of 21st Century after school programs in New York City (Leos-Urbel, 2015). Positive staff-student interactions predicted gains in math and reading across the elementary years for girls and boys, and gains in social skills for boys in a study by Pierce, et al. (2010). They also found that positive staff-child interactions resulted in positive gains in reading and math grades in grade two and reading grades in grade three. After school programs that were rated as emotionally supportive and fostering quality peer interactions predicted school adjustment for first graders (Pierce et al., 1999). Vandell et al. (2005) saw improved academic and behavioral results for students in programs with positive relationships between students and staff, as well as positive peer interactions. In addition, Cross, Gottfredson, Wilson, Rorie, and Connell (2010) found that a positive affective after school environment predicted higher youth engagement and improved outcomes.

Several staff characteristics are important to consider when looking at program quality. These include staff to student ratio, staff qualifications and training, and level of staff turnover. Lower student to staff ratios were related to improved reading achievement and student motivation (Mahoney, et al., 2005; Miller, 2003). High levels of staff turnover (1/3 turnover in program coordinators, and 2/3 turnover in staff) at 21st Century programs was associated with negative results (James-Burdumy et al., 2007). Vandell et al. (2007) and Miller (2003) found that low staff to student ratios, levels of staff training, and staff reports of job satisfaction were associated with better academic and social outcomes for students. Scott-Little et al. found in their meta-evaluation (2002) that successful programs employed well-qualified, trained staff. Cross et al. (2010) found that high levels of staff training and education level were related to levels of student engagement in after school programs, as was staff stability. Having full-time, sustainable site administration has also been found to be a significant factor in program success (Miller, 2003). Staff turnover, particularly in leadership roles, predicted less positive experiences and outcomes for youth.

The goals and structure of the after school program also impact student results. The program's focus, level of structure, and quality of curriculum deserve consideration. Programs have differing goals and focus areas, including academics, homework completion, recreation, or mentoring. Program focus impacts the outcome measures studied, as well as the level of impact in different areas. One consideration is the diversity of age-appropriate activities offered. A greater diversity of activities was associated with positive gains in math grades and work habits in grade three. Well-paced and organized programs predicted positive youth engagement (Pierce, et al., 2010), as did levels of student choice (Miller, 2003).

Communication with school staff as well as alignment with the school-day curriculum in a complementary manner are also associated with more positive outcomes for youth (James-Burdumy et al., 2007; Pierce et al., 1999). Achieving this level of communication between school and after school can be challenging. Some roadblocks to this include: not having overlapping schedules, high turnover of after school staff, and lack of understanding of program goals (Policy Studies Associates, 2001). Providing a structured setting with instructional support for homework was also associated with better outcomes (James-Burdumy et al., 2007; Pierce et al., 2010).

Lauer et al. (2004) found positive outcomes from programs with a well-defined curriculum that included academic and social goals, as well as mentoring. Fashola (1998) points to similar characteristics, but also highlights the benefits of offering one to one tutoring for student academic success. Scott-Little et al.'s 2002 meta-evaluation also references the benefits of one to one tutoring, but also found having a predictable schedule and strong links to the school-day curriculum as important indicators of program quality.

Students' perception of challenge predicted higher math and English grades, controlling for student demographics and baseline scores in a study by Shernoff (2010). Durlak et al. (2010) found in his meta-analysis that programs that offered a sequenced curriculum, active practice, focused objectives, and explicit instruction yielded significant positive effects for students.

A final element to consider is program maturity. Positive outcomes are more likely in established programs that have offered high quality activities over time (Shumow & Posner, 2005). Sheldon, et al. (2010) also found that mature, well-implemented programs were associated with positive youth outcomes.

In order to study the impact of after school programs on academic achievement and growth, the five sites that were selected for this study highlight many of the characteristics of high quality programs referenced above. They are academically oriented programs, which is important so that program goals align with the outcome measures being used. They are mature programs, having been in operation since 2000. Furthermore, there was a common director of the program throughout the five years of the study. While there was a common director, each site has its own site coordinator, and each school its own principal. Potential effects of school differences is considered in this study, although given that only five schools are studied, it is not possible to analytically examine the impact of different program characteristics on student outcomes. Results from this study provide guidance for future research in this area.

Variation in outcome measures: Dosage

Program dosage, or the amount of time students spend in the program, is another important element in looking at program outcomes. At the middle school level, program attendance has been linked with measures of program quality. However, among elementary school students, program attendance (dosage) is primarily impacted by parents' work schedules and the need for child care. This has resulted in higher attendance rates in after school programs, on average, for elementary than middle school students (Leos-Urbel, 2015; Weitzman et al., 2008).

Students with consistent attendance in after school programs (defined as attending at least four days per week) have higher academic outcomes in math test scores, higher grades, and better ratings of prosocial behavior (Auger, et al., 2013; Pierce, et al., 2010). The Mathematica study found that students with higher rates of attendance did show more positive outcomes (James-Burdumy et al., 2007), and students with higher dosages of homework help had higher SAT 9 scores (Cosden et al., 2004). Roth et al. (2010) found that middle school students who attended more hours in the program had improved school attendance.

Several recent studies have demonstrated a connection between high levels of program attendance (100+ days) and achievement scores in math (Frankel & Daley, 2007; Huang, Leon, Harven, La Torre, & Mostafavi, 2009; Welsh, Russell, Williams, Reisner, & White, 2002). The evaluation of the After-School Corporation's (TASC) after school program (Welsh et al., 2002) found that students that were highly active in the program for two years had the highest increases in math achievement. In a longitudinal study of students who attended the LA's BEST after school program over the course of four elementary school years, evaluators compared students who attended more than 100 days with students that attended between 1 and 20 days. They found statistically significant positive gains in math scores on the California state assessments. Statistically significant results were not found for Language Arts scores (Huang et al., 2009). Of these three studies, Frankel & Daley's study (2007) was the only one that found significant gains in reading achievement for students with high levels of attendance.

Some researchers (e.g., Lauer et al., 2004), have begun to look for the "sweet spot" for program attendance. Too little exposure does not impact development, but too much exposure can lead to negative results. Lauer's meta-analysis found that between 44-210 hours per year of tutoring in reading yielded maximum effect sizes (.19-.25), while 46-100 hours per year of math tutoring was ideal (effect sizes of .22-.26). Lauer's work provides a foundation for examining the impact of dosage. However, it is limited in that it examines the impact within a one year timeframe only.

While these findings are promising, Durlak et al.'s meta-analysis (2010) of after school programs noted poor attendance as a common feature of many after school programs. Average attendance rates ranged as low as 15-26% in 11 out of 75 studies. Eighty-five percent of studies showed attendance positively related to youth outcomes and program success.

Similarly, Hausner (2000) found that gains from one dosage of tutoring was not sustained over time, but that students needed "booster shots" to maintain academic gains. Vandell et al. (2006) also recognized the benefits of attendance in after school programs over the course of multiple years. They found that successive years of attendance intensified all effect sizes measured in the study. While this study tracked effects over two academic years, it is not yet known if attending longer adds additional benefit. In the study of LA's Best after school programs (Huang et al., 2000), students who attended for at least four years demonstrated improved attendance and standardized test scores on the SAT-9 Reading assessment. Greater duration of attendance was also positively linked to academic performance for elementary school students, but not middle school students in four out of eight studies included in Roth et al.'s review of the literature (2010).

Not all researchers have shown a connection between program attendance and youth outcomes (Hirsch et al., 2010; Vandell et al., 2005). Hirsch et al. (2010) found that it was necessary to look at student engagement as a mediating factor between attendance and outcomes.

Better understanding the impact of program dosage is a critical aspect of the current study. Examining the impact of dosage within a school year and over time is able to be richly studied using a longitudinal design. Treating participation as a time-varying predictor allows examination of different trends of participation over time. This information influences policy makers and programmers in working with families to ensure students receive an adequate dose of the intervention to see desired impacts.

Limitations of Prior Research

There is a rich literature studying impacts of after school programming over the past 30 years. Studies have examined many outcome measures including impacts on academics, behavior, attendance, and social skills. Various factors such as student demographics, program characteristics, and dosage have been considered. While these studies point us in the direction of what types of programs yield maximum results for which students, they fall short with their methodology.

As noted throughout this chapter, the existing research base suffers from methodological flaws. In particular, experimental design has not been rigorous. In Scott-Little et al.'s (2002) meta-evaluation of after school programs, only two program evaluations used an experimental design, and these were not traditional or school-based programs. Seven of the 15 program evaluations he reviewed only utilized pre and post-test measures, while six out of 15 used a quasi-experimental design. Researchers using quasi-experimental designs compared participants with non-participants; however, Scott-Little et al. (2002) noted that none of the authors he reviewed used appropriate matching or statistical techniques to ensure comparability of groups.

Using an experimental design to study after school program participation has many challenges. Politically, it can be controversial to offer an after school program in a school and not allow certain students to attend the program. There is also typically very high turnover in student participation in after school programs, which can impact an experimental design (Scott-Little et al., 2002). However, without this design, concerns of selection bias must be considered. Perhaps students or families that select after school programs versus care with a sitter or self-care value education more, or are more motivated to succeed (Miller, 2003). The other issue with any study of after school programs is that your control group never truly lacks "treatment." Even if they are not in a formal after school program, they are doing something after school. After school activities vary greatly both within and between children, making it very challenging to pinpoint what activities are or are not related to student growth (Miller, 2003).

Researchers have called for more longitudinal study of after school programs (Roth et al., 2010; Scott-Little et al., 2002). Longitudinal studies would allow the study of the impact of duration in after school program attendance, as well as any delayed effects that might emerge (Roth et al., 2010). Most studies consider results over a single year of treatment. Even studies that have multiple years of data have not modeled student growth trajectories over time.

Another area requiring further research is the optimal dosage of after school care to produce desired outcomes. Dosage is considered in various ways in the research. Some researchers have considered it as a continuous variable (noting total days attended), while others have considered it categorically (participates/does not participate; high attendance/ low attendance). Research is needed to find the minimal participation level necessary to get expected developmental gains. The upper end of the threshold should also be considered to ascertain what dosage may start to hinder positive growth and development (Roth et al., 2010).

This present study addresses some of the limitations of the current research base. While not experimental, the study utilizes propensity scores to ensure appropriate matching of treatment and control groups (Graham & Kurleander, 2011). In addition, this study is longitudinal – spanning five years of student participation. Finally, a primary focus of this study is examining the impact of program participation and program dosage in a variety of ways. A detailed description of the methodology for this study follows in chapter three.

Methodology

Methodologically strong research on the impact of after school programs is essential for educators and policy makers to make educated policy decisions and to attract and maintain funding sources for quality after school programs (Scott-Little et al., 2002). However, as discussed in Chapter two, the current research base on after school programs suffers from several critical methodological weaknesses addressed in this dissertation. For example, program goals are often not considered or not aligned with outcome measures. Appropriate reporting in evaluation studies has also been cited as a concern, with many studies not reporting results by grade level, or not reporting important statistical information such as sample sizes, standard deviations or proper matching of the treatment and control groups (Scott-Little et al., 2002). Research on after school programs has rarely been longitudinal (Roth et al., 2010), and has not adequately considered the impact of program dosage (Chaput et al., 2004). The remainder of this chapter details the specific methods used for this dissertation.

Site Selection

In this study, the research questions are answered using data on students in grades one through five who attended one of five elementary schools in Nashua, New Hampshire between 2008 and 2013. Expedited approval to conduct this study (#6476) was obtained on May 31, 2016 from The Institutional Review Board for the Protection of Human Subjects in Research at the University of New Hampshire (Appendix A).

While the 21st Century programs operate at a national level, this study focuses on the programs operated in the elementary schools of Nashua, New Hampshire. As of 2013, New Hampshire housed 21st Century after school programs in 67 sites in 24 districts, serving 10,764 students with federal funding of over $5.6 million. Just over half of districts that requested funding received grants from New Hampshire in the last round of allocations.

Children living in New Hampshire experience many of the same challenges as they do at the national level. With childhood poverty rates at 12.5% in 2015 (Center for American Progress, 2015), these are better than the national statistics (19.7%), but comparable. New Hampshire Department of Education officials who administer the 21st Century grants seek to ensure that these programs target youth at-risk for school challenges in order to serve those who would benefit most from the programs and support the educational and social development of students (Russell & Woods, 2012). This is evident in their statistics which show that 54% of students enrolled in the 21st Century programs qualify for free/reduced lunch, compared to 27% of students in the state. New Hampshire 21st Century programs operate in half of the state's persistently lowest achieving schools, and service a higher percentage of students scoring below proficient on state achievement tests in Reading and Math. These programs also service a higher percentage of ELL and minority students than are seen in the state on average (NH Department of Education, 2013).

Nashua has been the site of 21st Century after school programs since 2000. At its peak (2000-2003), federal funding of the program was $760,000 annually. Funding has slowly declined, and has held steady at $371,000 since 2008. In order to examine the effectiveness of 21st Century after school programs, their impact was examined in one urban area, Nashua, New Hampshire. Nashua is a small city of approximately 100,000 residents. Over 12,000 students attend school in the Nashua School District. Five of Nashua's elementary schools host 21st Century after school programs, and have had them for over ten years. These schools are all school-wide Title One schools, with poverty rates of between 60 and 85%.

While these five after school programs have a common director, there are differences among them in terms of staff consistency over time, program size, and student attendance in the programs. While there are not enough sites to consider their effect as part of the statistical model, site of attendance is included as a dummy variable in order to model any impact of school site. Differences found regarding school site are described in Chapter four, and implications for future research in this area are discussed in Chapter five.

One reason for selecting Nashua as the site for this research is a convenience consideration, given that I live and work in Nashua. To situate myself in this study, I am the principal of one of the five schools referenced above, and I taught at one of the other schools from 1999-2006. While this is a potential concern, I do plan to focus my study on data collected during the period of time when I did not work in the district (2006-2012). Furthermore, while the 21st Century programs are an offshoot of the school district, I do not serve as supervisor to any staff of this program in my current position. It is my perspective that my knowledge of the district assisted me in conducting a knowledgeable and nuanced study.

Throughout my analysis and interpretation of the data, I needed to remain aware of the concern that my position as principal of one of the five schools included in this study could impact my interpretations. As I did not work at the school during the time period of the data examined, the results are not a reflection on my work. However, I had to continuously be critical about my assumptions and any conflicts that could have arisen due to my position in the district. Toward this end, I have chosen to refer to the schools included in this study by number, rather than by name. Doing this preserves the anonymity of individuals responsible for the learning of students at each school, such as the teachers, principals, and after school site directors. Anonymity of participants is an ethical concern and preserving it helps to insure that no one suffers ill-effects from this study (Creswell, 2009).

Beyond the convenience factor, Nashua is an ideal site for this study for other reasons as well. The Nashua School District has received funding for 21st Century after school programs since 2000. This long track record of experience is a counterpoint to the infancy of the programs studied in the Mathematica study. Furthermore, the programs in Nashua have had many years of working toward the stated objective of raising academic achievement, as noted in the No Child Left Behind legislation. The five elementary schools with sites in Nashua are all urban schools with diverse populations and moderate to high poverty rates. This offers a similar demographic to the majority of the sites studied in other major research of after school program effectiveness (James-Burdumy et al., 2008; Lord & Mahoney, 2007).

Formation of the Analytic Data Set

The New Hampshire Department of Education (DOE) maintains a database called i4see which houses student-level data on all students in the state. The DOE requires all 21st Century programs to maintain extensive student data as part of the district's i4see submissions. This data has been collected over time for many years and is accessible through the 21st Century director for the Nashua School District. In addition to the data on the 21st Century students, comparable data on students from these schools that did not attend the 21st Century programs was collected from the Nashua School District Data Coordinator.

In examining the effectiveness of after school programs, there are many variables that can be considered. As described in Chapter two, students' socioeconomic status, gender, grade, and status as a learner in need of special instruction have been shown in past studies to impact the effect of after school program attendance. The research shows that program characteristics can also impact effects of after school program attendance. Although program characteristics cannot be analytically examined given that only five schools are studied, differences in program effects across the schools are studied. Finally, research has shown that the amount of time students spend in after school programs impacts the effects of attendance. Dosage is considered in several ways in this study. For clarity's sake, I outline the variables for which data was collected in the table below:

Table 1: Variables included in analytic models

Variable Type	Specific variables
Participation (*Treatment group only)*	• Treatment – participant in after school program or not (PARTIC = 1 for yes, 0 for no) • Grade level of participation – GR1PARTIC, GR2PARTIC, etc. • Site of participation (SCHOOL*PARTIC)
Dosage (*Treatment group only)*	• Years of participation in after school program (YEARS) • Attendance dosage each year and over time (DAYS, CUMDAYS, TOTDAYS) • Site dosage (SCHOOL*DAYS)
Demographic: (*Treatment & control groups)*	• Gender (MALE = 1 for boys, 0 for girls) • Race (White, Hispanic, African American (HISPANIC = 1 for Hispanic children. AFAM = 1 for African American children; the omitted category, White, served as the comparison category for model parameters) • Free/Reduced lunch status (POVERTY = 1 for students who qualify for free lunch, 0 for reduced pay or full pay students. FRL = 1 for students who qualify for free or reduced price lunch, 0 for students who do not qualify) • Grade level (GRADE) • School (1-5) • Identification as receiving ELL or Special Education services (ELL = 1 for students participating in the ELL program, 0 for students who are not in the program. SPED = 1 for students identified as part of the Special education program, 0 for students that are not identified as part of the program)

Outcome measures (*Treatment & control groups)*	• Student Achievement – NECAP scores for Reading & Math (Grades 3-6) for 2009-2013 (MATHACH, READACH)

Participation & Dosage Data

In this study, an after school participant is defined as a student who attended the 21st Century after school program at one of the five elementary schools included in this study at least 30 days within one school year. Participation is designated for each year of the study and over time. Dosage is conceptualized in several ways. First, intensity of participation is considered as the number of days a student attended any of the five 21st Century programs within one school year. Next, it is considered longitudinally as cumulative days of attendance within a span of three years. Finally, it is considered as duration of attendance, measured as the number of years a student attended the program within the span of this study (2008-2013).

What is not studied with regard to participation and dosage is the number of hours a student spent in the program each day or each school year. Some students may have stayed for one hour each day, while others may have stayed three hours. The other aspect of dosage that is not included in this study is the breadth, or diversity, of activities the student participated in each week within the program. This data is not tracked in Nashua at this time. Both of these ways of conceptualizing dosage are recommended by Chaput et al. (2004), and their omission poses a potential threat to construct validity (Shadish, Cook, & Campbell, 2002).

After school participation data was provided in Excel format with a different spreadsheet for each year requested (2009-2013). Each spreadsheet contained data including student ID numbers (SASIDS), school of attendance, and number of days they attended the program that school year. I began by creating one spreadsheet for each school, containing the attendance data for each year, linked by SASID. This resulted in five spreadsheets (one for each school), which contain one line per student, and columns for the number of days (09DAYS, etc.) attended in each of the five years.

Next, columns were added for each year to indicate the level of participation for each child ("PARTIC09," etc.). For each year, a child's participation level was coded as '0' if they did not attend the program at all, '1' if they attended between 1 and 29 days, and '2' if they attended 30 days or more. This is based on the federal guidelines for the 21st Century program which define a participant as a student who attends for 30 or more days in a school year (U.S. DOE, 2003).

For each student, I then counted how many of the five years they attended the program (0-5) and added a column titled "YEARS." Next, I used the calculation function in EXCEL to calculate the cumulative attendance for each student. This resulted in cumulative attendance for each year (CUM10, CUM11, etc.), as well as the total number of days the student attended the program between 2009 and 2013 (TOTDAYS). Finally, each student was given an overall participation code (PARTIC), based on whether they participated at all over the five years (1) or not (0). Looking at the five schools' after school data, the smallest school, School 1, had 269 unique (non-duplicated) students who attended 21st Century after school programs over the five years, while the largest school, School 3, had 681 unique students.

Demographic Data

The inclusion of demographic variables for all students included in this study is critical for two reasons. First, demographic variables are used in the propensity score models as a method for forming equivalent treatment and control groups within this secondary data set. These variables are again included in all final OLS regression models and hierarchical linear models as control variables. This provides a level of "double robustness" (Ho, Imai, King, &

Stuart, 2007) to ensure that differences between the treatment and control group are due to participation in the program and not to observed demographic differences between groups.

The other reason demographic variables are critical is to examine differential effects of program attendance for different groups of students. In this study, participation effects are considered for each site, for grade level of attendance, and for all student-level characteristics. This is done through the creation of interaction variables (e.g., MALE*PARTICIPATION, ELL*DAYS, etc.). More details on each demographic variable is included below.

I acquired demographic and achievement data for each of these five schools from the Student Data Coordinator for the Nashua School District. These were shared as Excel files containing one year of data each. The six files represented school years 2008-9 to 2014-15. Each spreadsheet contained lines consisting of student ID (SASID), school of attendance, grade, gender, race, free/reduced lunch status, ELL status, special education status, and NECAP scores (Reading and Math). I sorted these files by school to result in five spreadsheets – each containing demographic and achievement data from 2008-2015 for all of the students from one school. Within these spreadsheets, some students (SASIDS) appear multiple times, as each line represents a different year of attendance at the school. Next, data was coded from the form presented by the district into the format needed to conduct statistical analysis.

The five schools in this study have demonstrated varying degrees of student achievement on the NECAP assessments over the years, and it would be a bad assumption to think that students may not have differential academic or behavioral outcomes based on the school they attend. This would violate the independence assumption of single level regression models, thereby requiring the use of fixed effects analysis. To address school-level effects, this was treated as a fixed effect, using dummy variables to represent each school in the model.

School of attendance was coded in two ways. For the purposes of exploratory analysis, the schools were assigned number categories in alphabetical order (1 - 5). This number was assigned as the 'SCHOOL' variable for each line of data. Additional columns were added to represent four of the five schools. School three was excluded as the dummy variable because it is the largest of the five schools. Each of the remaining columns was coded as '0' or '1' based on whether or not the student attended that school that year. Coding the schools using a dummy variable allows me to include 'school' as a variable in regression equations later in the analysis. With school three assigned as the comparison category, parameter estimates associated with the dummy variables for the other four schools serve as comparisons between each school and school three.

The 21st Century after school program is only offered to students in grades 1-5. Therefore, students with data from kindergarten were removed from the data set. This eliminated approximately 1,350 lines of data across the five schools. With students appearing multiple times in the data set, I wanted to know how many unique, non-duplicated students appeared in the data set. After sorting the data first by SASID and then by Grade, I added a column called 'DUPLICATE.' I went through each student and coded the line as '0' for the lowest grade level in which that SASID appears in the data set, and as '1' for any subsequent years.

Next, I looked at grade level of participation. The demographic data provided indicated grade of attendance for that year of data. This grade tells the year the student took that NECAP assessment, what school they attended that year, if they had an IEP that year, etc. In order to use this data correctly in analysis, it was coded by year, resulting in the addition of five additional columns – '09GRADE' through '13GRADE.' This coding, however, does not tell if the student participated in the after school program or not during that grade level. To find this out, this information had to be compared with participation data. At this point, I added columns called 'GR1PARTIC, GR2PARTIC,' etc. These columns indicate students who participated in the after school program during their first grade year second grade year, etc. This allows consideration of differential impacts of participation at different grade levels. Inclusion of a dummy variable for grade one is particularly necessary as grade one participants do not have a direct outcome measure (NECAP score). This is due to the fact that students did not take the NECAP assessment until the fall of grade 3.

Following this, I began coding demographics using binary variables. The research base on after school programming informed the coding choices. Some categories of students have been shown to benefit more from participation in after school programming than others (Posner & Vandell, 1994). These categories of students were chosen as column headers for these variables. These categories include males, students living in poverty, African Americans and Hispanics, and students with ELL or Special Education status. Outlined below are the specific categories.

Student gender was captured using a variable called 'MALE,' coded as '1' for males and '0' for females. Race/ethnicity was presented as a single column with one or more of the race/ethnicity categories listed within this, based on parent report at the time of enrollment in the school district. I separated this into two columns – 'AFRICAN AMERICAN' and 'HISPANIC.' Students could be coded as '1' for either category or for both. Students identified as White, Non-Hispanic would have a '0' in both columns. There were only 189 students identified as Asian out of over 3,000 students, so I chose to remove these students from the data set to simplify analysis. Students were coded as '1' for 'ELL' if it was noted they were active in the

ELL program, and ‘0’ if they were noted as not ever enrolled or discharged from the program. Special Education status (‘SPED’) was coded similarly with a ‘1’ for having a current Individualized Education Plan (IEP), and ‘0’ if they never had one or had been discharged.

Poverty is an important factor in the theoretical framework guiding my analysis and therefore warranted a high level of consideration. Qualification for free or reduced price lunch is a common proxy for low socioeconomic status in educational research. Research on the impact of socioeconomic status on student achievement found comparable effect sizes whether they used qualification for free or reduced price lunch status, parents’ education level, parental occupation, or household income to measure socioeconomic status (Sirin, 2005). To qualify for free lunch, the student’s family must have income below 130% of the federal poverty line. Qualification for reduced priced lunch is available for families that earn between 130 and 185% of the federal poverty level (Sirin, 2005).

Concerns with this measure over or under-reporting students’ poverty status pose a potential threat to construct validity. There is the potential for families to complete fraudulent applications for the program, thereby designating them as low income when they are not. Alternatively, some families may not apply for the program due to a lack of awareness, pride, or simply because they don’t plan to utilize the school lunch program for their child.

Each student in the spreadsheet had a coding from the district as F, R, or blank, based on whether or not the student had qualified for Free or Reduced price lunch for that year based on the federal guidelines for this program. A student with no letter either hadn’t applied for the program or had not qualified. Based on this, I created two categories – ‘FRL’ and ‘POVERTY.’ In the ‘FRL’ category, all students eligible for free or reduced lunch were coded as ‘1’ and all full pay students were coded as '0.' In the 'POVERTY' category, only students who qualified for free lunch were coded as '1' and full pay and reduced pay students were coded as '0.'

Once the demographic data was properly coded for each school, I merged it with the participation data for that school, resulting in five total spreadsheets – one for each school. Students who had never attended the 21st Century program during this window of time were coded as '0' for 'PARTIC,' and had zeros filled in for all participation data. Many students appeared in the data set multiple times, indicating that they attended that school for multiple years during that time period. Participation data was copied so that it appeared with each line of student data for all duplicated students. Next, the five school data sets were merged into one full data set. Some students (SASIDs) had participation data but did not appear in any of the demographic data. These students were removed from the data set. At this point, I rechecked the coding for 'DUPLICATES.' There were many students that appeared in the data for two or more of these schools over the six year period. This resulted in additional students being coded as duplicates, as they had appeared earlier or later in the data at a different school. Along with this, I rechecked coding as a 'PARTICIPANT.' Some students who appear across multiple schools were originally incorrectly coded as non-participants because their participation data appeared for another school. Days and years of participation were corrected so that they were correct across all lines for these students.

Achievement Data

The outcome measures used in this study are the students' reading and mathematics NECAP test scores. NECAP (New England Common Assessment Program) is a standardized academic assessment created by educators from New Hampshire, Rhode Island, and Vermont in response to the federal No Child Left Behind legislation, and administered by Measured

Progress. It assesses students' achievement of established grade level expectations (GLE's) in grades three through eight, as well as 11. Students receive a scaled score for math and reading. Based on this score, students' achievement is rated as Proficient with Distinction (level 4), Proficient (level 3), Partially Proficient (level 2), or Substantially below Proficient (level 1). Students took the assessment in October, with the test assessing the previous year's teaching of the GLE's. New Hampshire used the NECAP as their accountability measure from 2005 until 2013 when they adopted the Common Core State Standards (NH DOE, 2015).

While there is some concern about using test scores as an outcome measure due to their lack of proximity from direct outcomes of program participation, such as improved homework completion, motivation, and school engagement (Miller, 2003), it is appropriate in this case due to the direct link to program goals. The primary measurable goal of the 21st Century Program is to increase the percentage of students that meet or exceed state academic achievement standards in reading and math. Furthermore, the NECAP assessment was the primary measure used for school accountability for student achievement during the time period of this study.

In this study, students' scale scores on the NECAP are used as the outcome measure and achievement levels are used to interpret the scaled scores. Scale scores are three digits long with the first digit indicating grade level, and the second and third digits indicating the student's score. Student scores can range from 00-80 and are equivalent across grade levels. For example, 40 is the cut-point for proficiency across all years, so that a score of 340 at grade 3 is equivalent to a score of 440 in grade 4 (NH DOE, 2015). Calculating the change in the last two digits of a student's score from year to year allows for a measure of growth or decline in achievement, as measured by this assessment.

In coding the achievement data, students who did not have NECAP scores for any year in the data set were removed. This removed data for 1,032 students. Most of these excluded students (937) were first or second graders in 2013 and 2014, therefore never having the opportunity to take the assessment during the window of this study. There were some students who were missing NECAP data for one year, but had data for other years (primarily first and second grade students). These lines were left in the data set, but highlighted to flag them as having missing data. Finally, NECAP data scores were reformatted for use in statistical analysis by removing the grade of testing from each score. For example, a third grader who received a perfect score on the test would score 380, a fourth grader 480 and so on. All of these scores were recoded as 80 in order to allow statistical comparison. This was done by adding two additional columns – 'READ2' and 'MATH2.'

Overview of the Data Set

With coding complete in the full data set, I was able to start dividing the data into different analytical data sets in order to conduct initial descriptive analyses. These analytical data sets included data separated by year of participation, by school, by 'no duplicates' and by 'participants' and 'non-participants.' These were then uploaded into IBM's SPSS Statistics Processor for descriptive analysis.

In the full analytic data set, there are 6,368 lines of data, representing 2,917 unique students. This difference exits because many of these students attended one or more of these five schools for multiple years between 2008 and 2014. Table 2 below shows the number of students included for each year of the study.

Table 2: Number of students included in analysis, by year

	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013
# of Students	1955	1893	1848	1764	1355

The five schools in this study vary in enrollment levels. Within this study, 28.4% of students were enrolled at School 3, 24.7% at School 4, 17.5% at School 2, 17.1% at School 5, and 12.3% at School 1. Given these participation differences by school, it was important to consider differences by school throughout my analysis.

What are the demographic characteristics of students attending these five schools between 2008 and 2014? Of the 2,917 students on which I have data, half were males and half females. Twenty-five percent of students were identified by their parents as Hispanic, and 9% were identified as African American. The remaining 66% were identified as White, Non-Hispanic by their parents. Looking at poverty levels, 71% of students qualified for the free or reduced price federal lunch program, and 64% qualified for free lunch. Some students included in this study have special learning needs, as shown by the 11% identified as English Language Learners and 12% identified as part of the Special Education program. Students' grade levels in any given year ranged from grade one to grade five, with a median of grade 3.2.

Attaining Balanced Treatment and Control Groups

The literature on after school programs has shown that certain demographics of students are more likely to access after school programs than others (Pettit et al., 1997). Demographic factors, including race and poverty status are also correlated with achievement on standardized assessments (NCES 2011a; NCES, 2011b). In this study, I estimate the treatment effect of attending the 21st Century after school program on achievement in Reading and Math. However, accurately estimating this treatment effect is challenging due to the fact that children are not randomly assigned to participate in the program. Instead, participation in the 21st Century program is voluntary – students and families can select to attend or not. If demographic factors impact selection of the program and impact achievement, then these factors could confound my estimate of the treatment effect (Ho, et al., 2007).

In fact, there are demographic differences between the groups, as I found when analyzing the data set. As shown in Table 2, the demographic characteristics of students who participated in the 21st Century Program and those who did not are significantly different, making it necessary to control for demographics between the two groups. This ensures that comparable students in the treatment and control groups are compared.

Table 3: Demographic information for participants and non-participants of the 21st Century after school program in Nashua, NH between 2008-2014 (*n* = 2917)

	Participants		Non-participants	
	Number	**Percentage**	**Number**	**Percentage**
Male	805	49%	664	52%
Hispanics****	462	28%	280	22%
African Americans*	166	10%	95	7%
Qualify for free or reduced lunch****	1230	75%	851	66%
Qualify for free lunch****	1102	67%	768	60%
Enrolled in ELL program	195	12%	133	10%
Enrolled in Special Education program	210	13%	144	11%

* Difference in means between treatment and control group is significant at the $p < .05$ level .
** Difference in means between treatment and control group is significant at the $p < .01$ level .
*** Difference in means between treatment and control group is significant at the $p < .001$ level .

In this study, I used propensity score methods to ensure that students in the control group and treatment group are matched on key demographic variables. I used available demographic variables to estimate propensity scores for all students in the sample, and then used these estimates to calculate inverse propensity weights for the control group (Thoemmes, 2012). The propensity score shows how likely a person (student) is to select the treatment (21st Century after school program), given observed covariates (demographic characteristics) (Thoemmes, 2012). Student demographic variables were used as predictors in a logistic regression model that models the conditional probability of a student being assigned to the treatment group (i.e., after-school program). In this way, propensity scores can be used to approximate the results from a randomized experiment, using observational data sets, such as this one (Shadish, Clark, Steiner, & Hill, 2008).

When estimating propensity scores for the purpose of achieving balance, it is important that all confounding covariates are included. Leaving out important variables that may impact selection and outcomes results in unobserved variable bias (Shadish, et al., 2008). The validity threat of this hidden bias can be attenuated by including relevant covariates that represent a typical group for what you are measuring, allowing for relevant conclusions and generalizations to be drawn (Rubin, 1974). Demographic factors included in this study, such as race, gender, grade, and poverty status have been shown in the literature (Pettit et al., 1997; Posner & Vandell, 1999; Vandell & Ramanan, 1991), to be highly correlated with participation in after school programs.

The limitation of this process is that the data set does not contain all possible factors that may impact the choice of families to have their student attend the 21st Century after school program or not, and thus may not completely address selection issues. For example, the data set does not contain any parental information, such as marital or employment status, which have been found to be factors impacting after school attendance (Pettit et al., 1997). The lack of this data poses a threat to internal validity (Shadish, et al., 2002). While I don’t have access to this information, I can ensure that the groups are appropriately matched based on available demographic factors. This process has been shown to attenuate selection bias (Graham & Kurleander, 2011).

The 21st Century program in Nashua has the capacity to enroll about 20% of the school population. Due to this fact, the data set contains far more students who did not attend the program than those who did. One way to achieve balance in the data set would be to match each participant to a non-participant with a similar propensity score. This however, would eliminate a large number of non-participants from the study. Keeping a large sample size has some statistical benefits, such as smaller standard errors and reduced variance (Thoemmes, 2012). In order to achieve balance between the groups and maintain statistical power, weighting students based on their propensity scores is a good option.

Once the propensity scores were estimated, inverse propensity weights were calculated for the control group. When estimating treatment effects, I estimated the impact of attending on those likely to attend, not on the general population of potential attendees, referred to as the Average Treatment Effect on the Treated, or ATT. Since I estimated the ATT, all students who attended the 21st Century program were assigned a weight of one, and students in the control group receive a weight dependent on their propensity score (P/1-P) (Freedman & Berk, 2008). Students who would have been likely to select the program, but didn't (high propensity scores), would have a high weight, and students with low propensities to select the program based on demographic factors, would receive low weights. In this way, every student who did not participate in the program can still be included in the analyses without having imbalance between the groups.

Achieving balance between the treatment and control group as shown above results in a similarity of empirical distributions of the full set of covariates in both groups (Thoemmes, 2012). Table 4 illustrates this balance for the first year of data in this study. Tables demonstrating balance in the subsequent four years of data are included in Chapter four.

Table 4: Mean demographic characteristics for treatment and control groups for 2008-2009 school year before and after balancing using Inverse Propensity Weights (*n* = 1955)

	N	%MALE	%HISP	%AFAM	%FRL	%POV	%ELL	%SPED	SCHL	GRADE	PROP SCORE
Control	1516	50	17	8**	64***	57***	5	14	3.13*	2.92	.2171***
Treat	439	51	19	12	81	72	7	14	2.99	3.05	.2502
After weighting on Inverse Propensity Scores											
Control	439	51	19	12	81	72	7	14	3.00	3.04	.2484

Treat	439	51	19	12	81	72	7	14	2.99	3.05	.2502

* Difference in means between treatment and control group is significant at the $p < .05$ level .
** Difference in means between treatment and control group is significant at the $p < .01$ level .
*** Difference in means between treatment and control group is significant at the $p < .001$ level .

Even though the variables are balanced between the two groups, it is recommended (Ho et al., 2007) that the variables also be included in subsequent regression models in order to achieve “Double Robustness” in estimations. Including key variables in both stages of data analysis reduces bias when making causal inferences (Ho et al., 2007).

Data Analysis: Exploratory

First, I analyzed and described the data that is relevant to my questions. I address questions such as how many students are in the sample for each year of the data, including how many students attend each of the five schools. Next, I examined the number of participants versus non-participants overall, and by school, for each year of the study. Given that the outcome measure is performance on the Math and Reading NECAP, it was important to obtain descriptive analyses on this data as well. For the overall data for Math and Reading, as well as for each year of the data, I examined the range of scores, as well as the median, mean, and standard deviation. Outcome data was also broken down by school and by grade level (3-6). Finally, I compared descriptive data on NECAP scores for participants versus non-participants for each year of the study.

For exploratory analysis of demographic data, I calculated the percentages of overall students based on gender, race, poverty status, ELL status, and special education status. I then calculated these percentages for participants versus non-participants. Next, I considered if participation varies by grade level by running histograms for participants versus non-participants by grade level, and determined the mean grade level for participants and non-participants, as well as by school.

After conducting these initial analyses of the demographic factors in this data set, I began looking for relationships between these factors and student achievement. To do this, I conducted t-tests for equality of means, and estimated correlation coefficients between each of the following and NECAP scores for overall data, as well as by year – gender, race, poverty status, ELL status, special education status, school of attendance and grade. These relational analyses provided me with an indication of which demographic variables might be most significant to my later analyses. For exploratory analyses and all analyses that follow, an alpha level of .1 was used for null hypothesis significance tests. This level is commonly used in the social sciences, and allows for a 10% chance that results are attributable to chance (Shadish, et al., 2002). Having the alpha level set at .1 increases the possibility of Type 1 error and is a possible threat to statistical conclusion validity. Given the small effect sizes found in other studies of after school programs (e.g., Lauer et al., 2004), the higher alpha level is appropriate for detecting small effects. The large sample size used in this study yields higher levels of statistical power, improving the validity of statistical inferences (Shadish, et al., 2002).

Next, I examined the distribution of how many days each year and total years students attended, by school. I examined histograms, and calculated the range, median, mean, and standard deviations for each. I looked at the relationship between grade level of participants and the number of days students attended that year, through scatter plots, t-tests for equality of means, and correlation coefficients. Finally, I conducted the same analyses looking at the relationship between number of days and years attended and NECAP scores for both reading and math. This provides an indication of if there is a relationship (positive or negative) in this sample between cumulative days or years of attendance and NECAP scores.

To conduct exploratory data analysis of change in achievement over time, I graphed observed empirical growth records of achievement versus time for a randomly selected group of 60 participants and non-participants across schools. Further exploratory work involved fitting ordinary least squares (OLS) regression models to the observed data for each student and graphing the resulting estimated regression lines. I further explored distributions of estimated initial achievement and rate of change from OLS regression models fit to the observed reading and math achievement data over time for all students in the sample. These exploratory analyses allowed me to preliminarily identify trends in growth in reading and math for participants and non-participants, thus providing initial answers to my research questions.

Data Analysis: Causal

Question 1**: Do students who attend 21st Century After school programs perform better academically, on average, based on scores on the grades 3-6 Math and Reading NECAP tests, than students in the same schools who do not attend these after school programs?**

As noted previously, one of the unique features of this study is the nuance with which student participation is being considered. Different research questions within this analysis require that we consider participation in various ways. For question one, each student is coded as a participant or non-participant for each year of the study (2008-2013). A student is considered a participant if they attended the after school program at least 30 days during that school year, and a non-participant if they participated between 0 and 29 days that year. In this way, a student's status as "participant" may very well change from year to year of the study.

Question one required examination of the math and reading achievement of participants versus non-participants for each year. In order to do this, I fit a series of simple and multiple regression models to math and reading achievement data for each year of the study, resulting in 10 sets of analyses. First, for each outcome, I fit a model including participation in the after school program as the only predictor. Next, I controlled for the students' grade of enrollment in the model, as students in first through fifth grade are all included. I then fit models controlling for school of attendance. Finally, I progressively added demographic factors such as poverty status, gender, race, and special learning status to the model in order to control for any variation caused by these factors. A sample model is:

$$MATHACH_i = \beta_0 + \beta_1(PARTIC_i) + \beta_2(GRADE_i) + \beta_3\,(S1_i) + \beta_4\,(S2_i) + \beta_5\,(S4_i) + \beta_6\,(S5_i) + \varepsilon_i$$

Interpreting this model, β_0 would represent the average math achievement for students at School 3 (the comparison school) who did not attend the after school program at least 30 days in that school year. Next, $\beta_1$would represent the difference (positive or negative) in math achievement, on average, for students who attended the after school program at least 30 days in that school year. Referring back to the first research question, if β_1 were positive, this would indicate that students who attend 21st Century After school programs do perform better academically (in Reading and/or Math), on average, than students in the same schools who do not attend these after school programs. Looking across the five years of data in this study would indicate if any differences in performance were consistent year to year, or if results varied. The value and significance of β_2 would tell us if the student's grade in school is related to higher or lower math achievement scores, and if so if students in higher grades have higher or lower scores, on average.

In the model, the values of β_3 through β_6 are connected to the dummy variables representing school of attendance for the student. There are four dummy variables in order to represent the five schools in the study. For example, if the student attended School 1, then a one would be inserted for the variable connected with β_3 . The other three dummy variables would have values of 0 and would drop from the equation. The value of $\beta_3$would indicate the average difference in math achievement between School 1 and School 3.

As I proceeded with model building, I examined the R^2statistics and residuals from the fitted models to consider how much explained and unexplained variability there is in NECAP scores. Results of fitted achievement models of math and reading achievement for each year of the study are displayed for participants and non-participants, as well as by other key variables.

To begin to dig deeper into the data and explain greater variation in NECAP scores amongst students, I next added a progression of interactions to the models in order to answer the second research question:

Question 2: **Do the effects of attending the 21st Century After School program differ based on student characteristics such as race, gender, free/reduced lunch status, status as a student with a disability or English language learner, or grade of attendance?**

As evidenced in the literature review, attendance in after school programs does not have the same effects on all students (Lord & Mahoney, 2007; Marshall et al., 1997; Pierce, et al., 2010). Therefore, it is important to consider the interaction of demographic factors with program participation, and any potential effects this might have on student achievement.

To address my second research question, I examined differences in the effects of program participation on test scores based on student characteristics. Similar to question one, I fit a series of multiple regression models to math and reading achievement data. Participation was still considered a dichotomous variable for each year, with each student either attending 30 days or more, or not.

First, I considered interaction effects between key demographic variables and participation in the after school program. I fit a taxonomy of models and evaluated goodness of fit for each. I used the initial descriptive analysis to guide the order in which I added interaction effects to the models. First, interactions between school of attendance and participation were added. Next, interactions between participation and poverty status, race, gender, and special learning status were added. Finally, interaction variables between participation and grade level were added. An example of a model is:

$$MATHACH_i = \beta_0 + \beta_1(PARTIC_i) + \beta_2(GRADE_i)$$
$$+ \beta_3(FRL_i) + \beta_4(SPED_i) + \beta_5(PARTIC_i * FRL_i) + \beta_6(S1_i) + \beta_7(S2_i)$$
$$+ \beta_8(S4_i) + \beta_9(S5_i) + \varepsilon_i$$

In interpreting this model, β_0 would represent the average math score for a first grade student attending School 3 who did not participate in the after school program that school year, and is not receiving free or reduced lunch and is not identified as receiving special education services. Going from there, β_1 represents the average difference in math achievement scores for a student who did participate in the program, controlling for grade, qualification for free or reduced price lunch, special education status, and school. Next, β_2 represents the average difference in math achievement score based on the student's grade. β_3 shows the impact of poverty status on math achievement for students who did not participate in the after school program, controlling for grade and special education status. Then, β_4 represents the average difference in math achievement scores for a student who receives special education services, controlling for poverty and participation status. Finally, β_5 represents an interaction effect between participation and poverty status. If the parameter estimate were positive, it would show that participation in the after school program has a greater benefit on math achievement for students who qualify for free or reduced priced lunch than for those who do not. The remaining parameter estimates control for school of attendance.

As I proceeded with model building, I examined the R^2 statistics and residuals from the fitted models to consider how much explained and unexplained variability there is in NECAP scores. Once a final model was fit for each subject and each year, I examined residuals to look for any abnormalities and to consider the amount of unexplained variance left in the models. I then estimated the impact of after school participation for prototypical students based on predictors included in the best-fit models. Results of fitted achievement models of math and reading achievement for each year of the study are displayed for participants and non-participants, as well as by other key variables.

Considering interaction effects is critical in research on prevention programs that build resiliency. This type of statistical analysis helps researchers to understand why some children maintain a positive developmental trajectory in spite of adversity, while others do not. Identifying these moderators of risk provides a foundation for the development of sound preventive interventions (Roosa, 2000).

Examining the interaction of participation at different grade levels or for different demographic groups is a piece of this research with the potential for applications in policy and practice. If participation is most significant at a particular grade level, then programs can target this grade. If participation shows differential effects for particular groups of students, then programs could be expanded for this demographic. Alternately, this could prompt educators to consider how program design could be altered to better address the needs of students that aren't seeing gains from participation in the program. Considering differential impacts of participation leads me to consider the impact of different levels of program participation, or dosage, as laid out in questions three and four.

Question 3**: Do the effects of attending the 21st Century After school program differ based on student attendance in the program within an academic year (# of days attended)?**

In order to answer question three, I conducted the final analysis that considers the data on a year to year basis, as was done in questions one and two. I examined if effects of program attendance varied based on the number of days students attended within each year. For example, would a student who attended 100 days be predicted to have higher math achievement that year than a student who attended 50 days? How about 150 days?

In this question, the thinking about participation is expanded to consider it beyond the dichotomous variable of participant or not. I now considered the level of participation, as measured by the number of days each participant attended the program within each school year. At this point, dosage (days attended within the year) is also considered in interaction with other critical variables (such as poverty status), as was done in question two.

I fit a series of simple and multiple regression models to math and reading achievement data for each year of the study, resulting in 10 sets of analyses. Participation and dosage were added first, as they are the key variables to answering question three. Next, control variables were added. Following this, interaction variables between days of attendance and school, demographics, and grade level were added, in that order. Finally, interaction effects were considered between school effects and key demographic variables – poverty status and status as a special learner (ELL or Special education). An example of a model follows:

$$
\begin{aligned}
MATHACH_i = \beta_0 + \beta_1(GRADE_i) \\
&+ \beta_2(FRL_i) + \beta_3(SPED_i) + \beta_4(DAYS_i) + \beta_5(DAYS_i * FRL_i) + \beta_6(S1_i) \\
&+ \beta_7(S2_i) + \beta_8(S4_i) + \beta_9(S5_i) + \varepsilon_i
\end{aligned}
$$

To answer question three, the key parameter estimates are β_4, which shows the difference in math achievement (positive or negative) for each additional day of attendance during that school year for a student who does not qualify for free or reduced price lunch, controlling for grade, status as student with a disability, and program of attendance. Parameter estimate β_5 indicates any differential effects of program dosage on math achievement for children qualifying for free or reduced price lunch. Analysis of the estimates predicted by the models in question three provide answers regarding the impact of intensity of participation within a school year.

As I proceeded with model building, I examined the R^2statistics and residuals from the fitted models to consider how much explained and unexplained variability there is in NECAP scores. Once a final model was fit for each subject and each year, I examined residuals to look for any abnormalities and to consider the amount of unexplained variance left in the models. After residuals were considered for each model, I graphed results showing math and reading achievement for prototypical students as an outcome of days of attendance. Question four considers the effects on academic achievement of program intensity and duration over time.

***Question 4*: Do attendees show greater growth in NECAP scores over time than non-attendees? Do the effects of attendance differ based on cumulative attendance over the elementary school career?**

To answer question four, I conducted multilevel modeling of change to examine growth in math and reading achievement over the five years for participants versus non-participants. Multilevel modeling for change is an analytic approach which allowed me to longitudinally model achievement over time. The model can include student-level variables other than time, such as program attendance and demographics. Multilevel modeling can be used even with unequally spaced measurement occasions and varying numbers of measurement occasions. Using a multilevel model allows me to look at how attendance over time impacts student outcomes (Luke, 2004).

Given that this is a longitudinal study, time is treated as the level one variable. Student academic growth trajectories were modeled over time, and I specifically considered whether trajectories vary based on program participation and dosage, and if these impacts differ based on student characteristics. Multilevel modeling allowed me to determine whether effects increase, decrease, or remain stable over time, and weather the trends are linear or non-linear. This technique allowed me to include predictors that change over time, such as program participation and dosage (Graham, Singer, & Willett, 2008). The answers gleaned from this analysis shed light on questions such as the need for dosage over time. Analyzing the longitudinal data available answers my research questions more richly.

The first step in answering question four was to convert the year by year data sets into one person period data set. Once data was arranged longitudinally by student ID, I was able to see which students had three or more years of data. There were no students with more than three waves of data and there were many students with fewer than three years. Students with only one year of achievement data were immediately eliminated. I then looked at the ratio of students with two years of data compared to three years and found that there were almost three times as many students with two waves of data as three. Including these students would negatively affect the accuracy of my fitted models and might even prevent them from converging. Therefore, I only included students with three waves of data in the person-period data sets.

This resulted in data sets with 367 students with three waves of math scores and 353 students with three waves of reading scores. Gender and race variables were constant for students across waves, but several other demographic factors were allowed to vary across waves of data. These included school of attendance, grade level, poverty status, and status as a special learner (ELL or SPED), as these factors can change for a student from one year to the next. Participation status and days of attendance were also allowed to vary across waves of data.

Once the data set was formed, the next step was to achieve balance in the data sets on key demographic factors. I did this by estimating propensity scores for participation for each wave of data. The data was then weighted by students' inverse propensity scores, in the same manner as was done for the year by year data sets. The waves of data were not evenly matched in terms of number of students included or student demographics. For math data, wave one (2009-11) had 195 students, wave two (2010-2012) had 45 students, and wave three (2011-2013) had 127 students. The breakdown was similar for reading data with 212, 26, and 115 students respectively. Multilevel modeling for change allows for unequal numbers of data points across waves (Luke, 2004).

The propensity of participating in the after school program was modeled using logistic regression for each wave of reading and math data. A student was considered a participant if they participated for 30 or more days during any of the three years. The propensity scores estimated by these models for each student were then used to calculate the Inverse Propensity Weight for each control student (Prop/(1-Prop). Since I calculated the average treatment effect on the treated (ATT), all treatment students were assigned a weight of 1 (Freedman & Berk, 2008). Once students were weighted by their IPW, demographic statistics were generated and compared to those calculated before weighting to ensure equivalency of the treatment and control group on key demographic features.

With balance achieved, question four begins by examining growth in math and reading achievement for participants versus non-participants over time. This question views participation as a dichotomous variable (participant or not), in the same manner in which it is considered in questions one and two. In this model, participation is treated as a time-varying predictor and is therefore included in level one of the multilevel model. This allowed for students' participation status to change over time. For example, a student may have participated in 2008, not in 2009, and then again in 2010. Student academic growth trajectories are modeled over time, specifically considering if trajectories vary based on program participation, and if these impacts differ based on student characteristics.

Before adding predictors, I first fit unconditional growth models to determine if math and reading achievement scores are linear or curvilinear over time. I then added participation as a variable to see if results vary over time for participants versus non-participants. Dummy variables for school of attendance, as well as demographic variables were next added to the model. Next, dosage over time was considered as a time varying predictor. Days, total days, and years of attendance were added to the model separately in order to examine dosage effects over time. Finally, interaction variables between days of attendance and school, gender and poverty status were added in order to determine the effect of dosage on growth trajectories in reading and math achievement, and whether these trajectories vary by gender or poverty status.

This analysis resulted in two sets of fitted models – one for math and one for reading. I compared differences in initial achievement levels between treatment and control students, as well as rates of change over time. An example of a model that was fit is:

Level 1 Model: $MATHACH_{ij} = \pi_{0i} + \pi_{1i}(wave - 1_{ij}) + \pi_{2i}(PARTIC_{ij}) + \pi_{3i}(wave - 1_{ij}) * (PARTIC_{ij}) + \varepsilon_{ij}$

Level 2 Models:

$$\pi_{0i} = \gamma_{00} + \zeta_{0i}$$
$$\pi_{1i} = \gamma_{10} + \zeta_{1i}$$
$$\pi_{2i} = \gamma_{20}$$
$$\pi_{3i} = \gamma_{30}$$

Composite model: $MATHACH_{ij} = [\gamma_{00} + \gamma_{10}(wave - 1_{ij}) + \gamma_{20}PARTIC_{ij} + \gamma_{30}(wave - 1_{ij}) * (PARTIC_{ij})] + [\zeta_{0i} + \zeta_{1i}(wave - 1_{ij}) + \varepsilon_{ij}]$

To interpret this model, the parameters are explained in Table 4:

Table 5: Explanation of sample parameters for the multilevel model for change in math achievement

π_{0i}	Individual i's true initial math achievement, when PARTIC=0
π_{1i}	The slope of individual i's true rate of change in math achievement, when PARTIC=0
π_{2i}	The impact of participation on initial math achievement

π_{3i}	The impact on rate of change for individual i due to participation in the after school program
γ_{00}	The population average initial math achievement for non-participants
γ_{10}	Population average rate of change in math achievement for non-participants
γ_{20}	The population average difference in rate of change in math achievement for participants
γ_{30}	Population average rate of change due to participation in the after school program

This model demonstrates the impact of participation in 21st Century after school programs on math achievement over time. I evaluated goodness-of-fit using pseudo R^2statistics, and plotted residuals by participants/non-participants, and by school. I also compared the -2Log Likelihood, AIC, and BIC statistics to compare goodness-of-fit between models. I tested the structural and stochastic tenability of the model assumptions (the normality of the residuals) by graphing histograms and Normal Q-Q plots of residuals. Finally, I graphed growth trajectories for reading and math for prototypical students.

Conclusion

Exploring this existing databank of students who have attended 21st Century after school programs in Nashua over time and comparing this data to the same data for comparable students, allows me to tentatively make causal inferences about the impacts of after school programs at the elementary level. It demonstrates whether or not these programs improve school outcomes for children considered at-risk for educational difficulties, and if so, for which children it produces the best results. Finally, the study shows if dosage matters (within one school year and/or over time), and if so, how much and for which students? The answers to these questions are presented in the Results section that follows.

Results

The methods described in chapter three yielded results that build the knowledge base with regard to the impact of participation in after school programs on student achievement. First, results of descriptive analyses are presented, along with results of exploratory analyses of relationships between variables. Next, results of OLS regression analyses are presented for each of the five years of data. These results demonstrate the impact of participation in the after school program on academic achievement, including impacts of dosage, and differential impacts for different demographic groups. Finally, results of hierarchical linear modeling are presented which model student achievement over time for non-participants, and participants at varying dosage levels.

Descriptive Analyses

Student Characteristics

Demographic data was examined by school and by participation status. Demographics varied slightly by school. As illustrated in Table 6, School 4 had the highest percentage of students identified as Hispanic, at 30%, and School 1 had the lowest at 22%. Schools 1 and 4 had the highest percentages of students identified as African American, at 11%, with School 5 having the lowest at 6%. Free/Reduced price meal qualification had variability and is worth noting. Schools 3 and 5 had the lowest percentages of qualification at 67%, with Schools 1, 4, and 2 being higher at 72%, 75%, and 76% respectively. Variations are evident in ELL enrollment as well, with School 4 having 17% of students enrolled in ELL services, with the other schools being much lower (7-11%). Special Education percentages were approximately level across schools ranging from School 1 at 10% to School 3 at 14%.

Table 6: Demographic data (percentages), by school for 2008-2014 (*n* = 2917)

	School 1	School 2	School 3	School 4	School 5
Male	51	52	50	50	50
Hispanic	22	27	24	30	23
African American	11	7	9	11	6
Free or Reduced Lunch	72	76	67	75	67
Poverty	64	69	59	69	60
ELL	7	10	11	17	9
SpEd	10	12	14	12	13

Table 7 presents demographic data by year. From this table, we can see that demographics were fairly stable over the five years included in this study, with the exception of small increases in the percentage of students identified as Hispanic, and the percentage of students qualifying for free or reduced price lunch. The percentage of students identified as English Language Learners increased by 10 percentage points over the five years of this study.

Table 7: Demographic data (percentages) for students attending the five elementary schools offering 21st Century after school programming, by year (*n* = 2917)

	Male	Hispanic	African American	Free/Reduced Lunch	Poverty	ELL	Special Education
2009	50	26	9	70	62	5	14
2010	50	32	9	72	65	7	14
2011	50	36	9	74	67	9	13
2012	49	41	9	76	69	13	12
2013	49	43	9	77	70	15	11

Table 8 presents demographic information separately for study participants and non-participants, along with results of statistical comparisons of the percentages across categories. Notable differences are found between the demographic characteristics of participants in the 21st Century after school program versus non-participants. Participation rates were higher among females, Hispanics, African Americans, students qualifying for free or reduced lunch and among students with special learning needs (ELL and Special Education).

Table 8: Demographic information for participants and non-participants of the 21st Century after school program in Nashua, NH between 2008-2014 (*n* = 2917)

	Participants		Non-participants	
	Number	**Percentage**	**Number**	**Percentage**
Male	805	49%	664	52%
Hispanics* **	462	28%	280	22%
African Americans*	166	10%	95	7%
Qualify for free or reduced lunch* **	1230	75%	851	66%
Qualify for free lunch* **	1102	67%	768	60%
Enrolled in ELL program	195	12%	133	10%
Enrolled in Special Education program	210	13%	144	11%

* Difference in means between treatment and control group is significant at the $p < .05$ level .
** Difference in means between treatment and control group is significant at the $p < .01$ level .
*** Difference in means between treatment and control group is significant at the $p < .001$ level .

As part of the exploratory analysis, participation was examined at varying levels. Students were initially coded as non-participants if they did not participate in the program at all in a given year. Participants were coded as Level 1 or Level 2 participants, with Level 1 participants attending between one and 29 days, and Level 2 participants attending 30 or more days. Thirty days was chosen as the cut-point, as this is the dosage used at the federal level to define a participant in the 21st Century after school program (U.S. DOE, 2003). Some analyses below compare statistics for participants across level of participation. If level of participation is not noted, then Participation refers to Level 2 participants only.

Table 9 breaks down demographic information for 2010 to examine differences between non-participants and participants at Level 1 and Level 2. Level 1 participants are more demographically similar to non-participants than to Level 2 participants. This holds true for free/reduced lunch and poverty status, two of the key variables in this study. These same trends exist for each year of data. Therefore, when coding participation to answer my research questions, Level 1 participants were ultimately coded as non-participants.

Table 9: Demographic information for participants at Level 1 and Level 2 and non-participants of the 21st Century after school program in Nashua, NH (2010) (n = 1893)

	Non-Participants	Level 1 Participants	Level 2 Participants
Male	**51%**	**49%**	**48%**
Hispanics	**25%*****	**18%**	**37%*****
African Americans	**9%**	**8%**	**11%****
Qualify for free or reduced lunch	**69%**	**71%**	**83%*****
Qualify for free lunch	**63%**	**64%**	**74%*****
Enrolled in ELL program	**12%*****	**5%**	**13%*****
Enrolled in Special Education program	**11%*****	**16%**	**17%**

* Difference in means between treatment and control group is significant at the $p < .05$ level .
** Difference in means between treatment and control group is significant at the $p < .01$ level .
*** Difference in means between treatment and control group is significant at the $p < .001$ level .

Level of Participation

Table 10 displays the average days and years of participation in the after school program by demographic category. It is evident from this table that different demographic groups participated at different levels of intensity. Students identified as Hispanic participated, on average 106 total days, compared to 53.6 total days for non-Hispanics (t(1572) = -14.99, $p < .001$), and 1.45 years, compared to 1.03 years for non-Hispanics (t (1572) = -11.95, $p < .001$). Similar results are found for African Americans. This group participated 97 total days, on average, compared to 68.2 total days for students not identified as African American (t(1572) = 4.66, $p < .001$), and 1.33 years compared with 1.15 years (t(1572) = -3.22, $p < .01$). Students qualifying for free or reduced price lunch also participated at a higher intensity level than full pay students. Students on free/reduced lunch participated 79.1 total days, on average, compared with 48.1 total days for full pay students (t(1572) = -10.36, $p < .001$), and 1.24 years compared with .95 years (t(1572) = -8.37, $p < .001$). Students identified as ELL participated more total days, on average, than non-ELL students (91.1 total days, compared with 68.5 total days (t(1572) = -4.16, $p < .001$). Finally, students identified as part of the special education program participated more years, on average, at 1.25 years compared with 1.15 years (t(1572) = -2.08, p

< .05). There was no significant difference in total days of attendance for students in the special education program.

Table 10: Participation rates by demographics, for participants in the 21st Century after school program in Nashua, NH between 2008-2014 (*n* = 2917)

	Mean Days	Mean Years		Mean Days	Mean Years
Male	59.5	1.02	**Females**	62	1.06
Hispanics*	106***	1.45***	**Non-Hispanics**	53.6***	1.03***
African Americans*	97***	1.33	**Non-African Amer**	68.2***	1.15**
Qualify for free or reduced lunch*	79.1***	1.24***	**Full Pay Lunch**	48.1***	.95***
Enrolled in ELL program	91.1***	1.16	**Non-ELL**	68.5***	1.02
Enrolled in Special Education program	60.9	1.25*	**Non-Special Ed.**	60.7	1.15*

* Difference in means between treatment and control group is significant at the $p < .05$ level .
** Difference in means between treatment and control group is significant at the $p < .01$ level .
*** Difference in means between treatment and control group is significant at the $p < .001$ level .

Next, I considered demographic differences among students who participated in the after school program during different grade levels. From this analysis, it is notable that a higher percentage of males attended the program in grade one (52% of participants), but that this then shifted to higher participation rates by females in grade 3 (53% of participants) where it remained for fourth and fifth grades. Another notable trend is that students identified as Hispanic participated at high rates in the younger grades, but participation waned as the grade level increased. In this data set, the trend went as follows: 57% of participants in grade one were Hispanic, this dropped to 52% in grade two, 40% in grade three, 29% in grade four, and 21% by grade five. The percentage of participants identified as African American increased in grade three (12% of participants vs. 8% of non-participants), and stayed at this level through grade five.

Students receiving free or reduced priced lunch participated at higher rates than full pay students in all grade levels in this data set. For example, among fourth grade participants, 70% of non-participants qualified for free or reduced price lunch, while 84% of participants qualified.

Students identified as part of the special education program were equally represented among participants and non-participants until grade five, where we saw a higher representation among participants (18% of participants vs. 11% of non-participants). Students identified as English Language Learners had the opposite trend – participating at high levels in the younger grades and declining in the upper grades. For example, among grade one participants, 22% of participants were identified as ELL, while only 10% of non-participants were identified. However, among fifth grade participants, 6% were identified as ELL, while 13% were ELL among non-participants.

School of Attendance

Table 11 illustrates participation in the 21st Century after school program by school. Of the students included in this data set, 1,636 (56.1%) participated in the 21st Century program at some point over the five years, and 1,281 (43.9%) did not participate at all during these five years. Participation rates varied somewhat by school. School of attendance is included in later analyses and recommendations are made for further research in this area.

Table 11: Participation rate in 21st Century after school programming in Nashua, NH (2008-2013), by school (*n* = 2917)

	Percentage of enrolled students who participated in 21st Century at any point	Percentage of total participants enrolled at that school
School 1	50%	11%
School 2	63%	21%
School 3	50%	25%
School 4	56%	24%
School 5	63%	19%

Schools also varied in their intensity of participation. Students attending the after school program at School 5 participated a higher number of total days (83.2 compared with 68.2 ($t(1572) = -7.19$, $p < .001$)) and years (1.42 compared with 1.11 ($t(1572) = -3.67$, $p < .001$)), on average, than the other schools. School 2 also had higher participation intensity, on average, when compared with the other schools. Students there averaged 81.7 total days, compared with

68.4 days for the other schools ($t(1572) = -3.26$, $p < .01$), and 1.34 years compared with 1.13 years ($t(1572) = -4.9$, $p < .001$). School 1 had mixed results with higher total days of attendance (89.8 compared with 68.1 ($t(1572) = -4.17$, $p < .001$)), but lower total years, on average with 1.05 years compared with 1.18 ($t(1572) = 2.69$, $p < .01$). Finally, School 4 had lower total days of attendance than the other schools at 60 days compared with 74.3 days ($t(1572) = 4.49$, $p < .001$).

Grade Level

The grade levels in which students participate in after school programming is another important demographic characteristic to consider. Table 12 presents the percentage of students participating in the after school program at each grade level. Students participated in grades one through five. The average grade level for participants was 3.2.

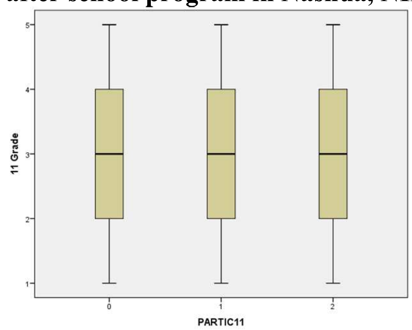
Table 12: Percentage of students participating in the 21st Century after school program in Nashua by grade level (2008-2013) (n = 2917)

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Non-participants	87	80	77	77	79
Participation – level 1	5	7	8	11	8
Participation – level 2	8	13	15	12	13

As seen above, the highest levels of participation are found in grades three and four, and the lowest levels are found in grade one. For all grade levels, less than 25% of the total population participated in the after school program. Among participants, more participated at level two (30 days or more), than at level one.

Participation levels were close to level across grade levels for each year, reflecting demographic shifts across years. For example, a slightly higher number of participants were in grades two and three in 2009. In 2010, slightly higher participation was found in grades three and four. As found in Figure 1, there is very little difference between mean grade level for participants (2), low participants (1), and non-participants (0). Average grade of participation was also very similar across schools, with no notable differences.

Figure 1: Box Plot demonstrating mean grade of enrollment for varying participation levels in the 21st Century after school program in Nashua, NH in 2011 (*n* = 1849)



As noted earlier, students who participated in 21st Century programs in grade one need to be identified so that they can be included in the data analysis. Of the 2,917 students, 661 of them participated in 21st Century programming during their first grade year during the period of this data collection. There likely are students in this data set who participated in grade one outside of the range of this study.

Dosage

Simply identifying a student as a participant in the 21st Century program is not adequate, as participation levels differ greatly among participants. Within the parameters of the data included in this study, on average, participants attended the 21st Century program for 1.85 years, with 75% of students attending two years or less. Looking at total days of attendance highlights the wide variability in attendance levels. Participants averaged 108.3 total days of attendance over the years included in this study. However, the standard deviation is 125.3 days, indicating the variability of this metric. In fact, total days of attendance ranged from one day to 796 days. Twenty-five percent of students attended 22 days or less, fifty percent attended 58 days or less, and seventy-five percent of participants attended 145 or less total days. Mean days of participation varied by year and again varied significantly. The highest mean days of participation was in 2012, with participants attending 71.1 days, on average (*SD* = 55.3 days). The lowest mean days of participation was in 2009, with participants attending 45 days, on average (*SD* = 40.1 days). Figure 2 presents the distribution of days of attendance for 2012 and is representative of the shape of the distribution for all five years of participation data in this study.

Figure 2: Histogram illustrating distribution of days of attendance in the 21st Century after school program in Nashua, NH in 2012 (*n* = 480)

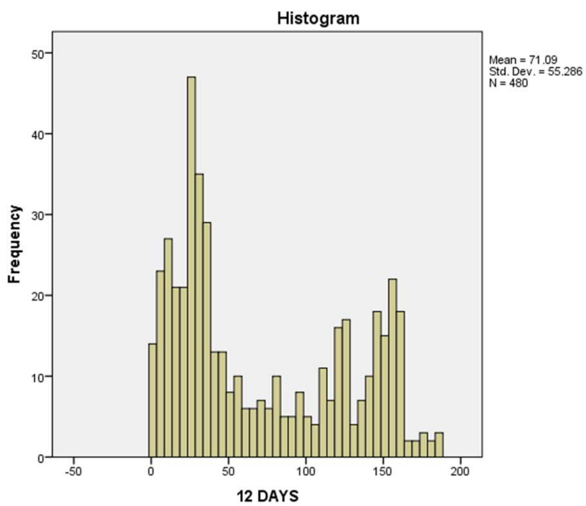
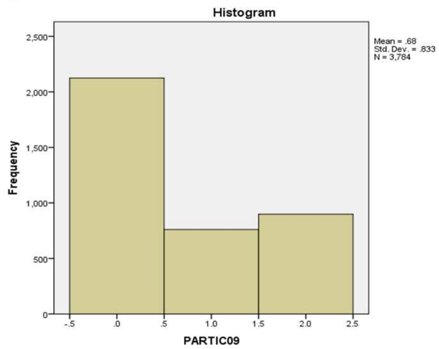


Table 13 below further outlines differences in participation levels for students across the five years of this study. For each year of the study, the greatest percentage of students did not participate in the after school program (participation level = 0). The next highest percentage is students who participated in the program 30 days or more that year (participation level = 2). Participation rates range from 43.8% (2009) to 29.7% (2011). The highest percentage of students participated in the program at least 30 days in 2013 (27%). Figure 3 below displays the distribution of attendance for one sample year, 2009.

Figure 3: Histogram displaying number of students at varying levels of participation in the 21st Century after school program in Nashua, NH in 2009 (*n* = 1956)



Also represented in Table 13 is the mean days that students attended the 21st Century after school program at each participation level for each year examined in this study. As expected, students coded with a participation level of 0 had zero mean days of participation. Students coded as level 1 participants, by definition, attended less than 30 days that year. Mean days of attendance for level 1 participants ranged from 12.6 days in 2009 to 17.4 days in 2012. Finally, our participants coded as level 2 had higher mean days of attendance for each year. In 2009, level 2 participants attended 73.9 days, on average, while the highest levels were found in 2013 with 101.4 days of attendance, on average.

Table 13: Mean days of participation by participation level at schools in Nashua, NH offering the 21st Century after school program, by year (2008-2013) (*n* = 2917)

Participation Level		**2009**	**2010**	**2011**	**2012**	**2013**
0	%	56.2	63.8	70.3	70.1	60.1
	Mean Days attending (*SD*)	0	0	0	0	0
1	%	20.1	14.7	10.3	9.8	12.4
	Mean Days attending (*SD*)	12.6 (8.2)	14.1 (7.8)	14.2 (8.4)	17.4 (8.8)	15.1 (6.5)
2	%	23.7	21.5	19.4	20.1	27
	Mean Days attending (*SD*)	73.9 (33.8)	90 (43.3)	86.7 (46.1)	99.5 (48.3)	101.4 (49.3)

Figures 4 through 6 below illustrate the distribution of days of attendance for participants at level one and level two for the 2010 school year. Days of participation at level one range from one to twenty-nine days, with a main cluster between 10 and 20 days. This is consistent with the mean for 2010 of 14.1 days of attendance, on average, for students at level one. Level two participants range from 30 to 175 days of attendance in the program in 2010. The distribution is skewed to the left, with peaks of participation between 30 and 50 days. Participation peaks again in the range of 120 to 150 days. While the mean number of days of participation for students at level two in 2010 was 90 days, it is interesting to note that this is due to the 'U' shaped nature of this distribution, rather than a peak of students who attended about 90 days that year.

Figure 4&5: Histograms illustrating the distribution of days attended for low and high level participants in the 21st Century program in Nashua, NH in 2010 (n = 669).

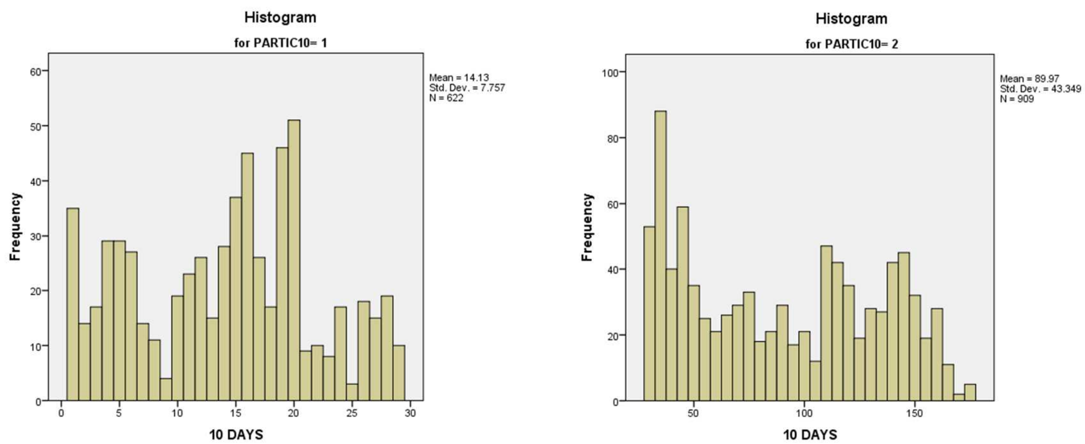
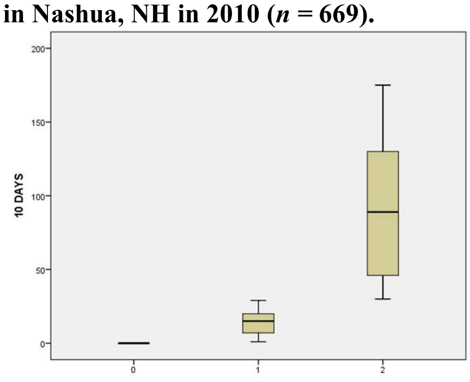


Figure 6: Box plot displaying the distribution of days of attendance in the 21st Century after school program in Nashua, NH in 2010 (n = 669).



Participation by Grade Level

Students who participated in the 21st Century after school program in first grade had higher overall participation than participants who did not attend in grade one. Grade one participants averaged a total of 247.5 total days of attendance over the span of this data set and averaged 2.75 years of attendance in the program. Participants who did not attend in grade one averaged 84.7 total days of attendance and 1.7 years of attendance in the program. Figures 7-9 illustrate the differences in the distribution of days and years of attendance for students who did and did not attend the program in grade one. Students who attended the program in grade one attended more days and years in the program over time, on average, than participants who did not attend in grade one.

Figure 7: Box plot of the distribution of total days of attendance (2008-2013) in the 21st Century after school program in Nashua, NH for students who attended the program in grade one (1), and participants in the program who did not attend in grade one (0) (*n* = 1636)

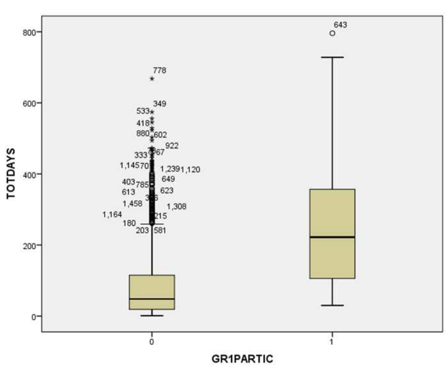
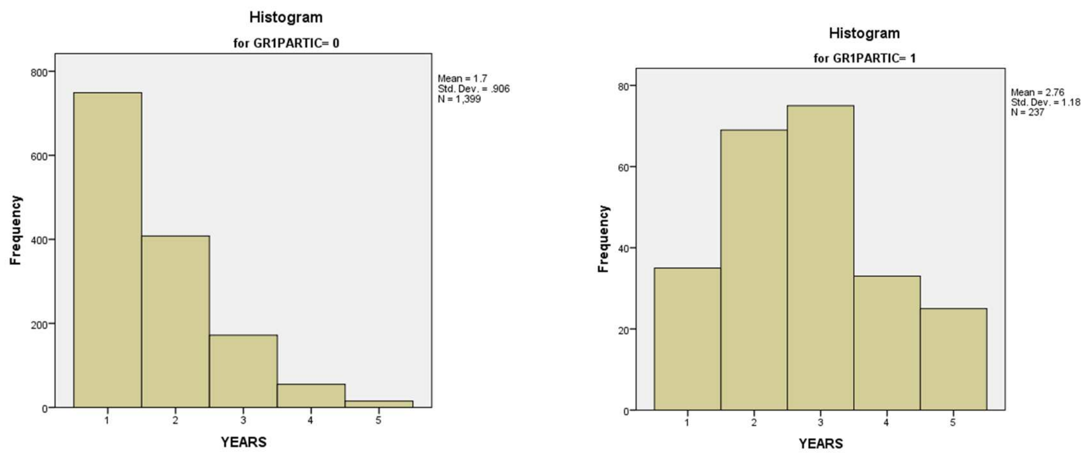


Figure 8&9: Histograms of the distribution of years of attendance in the 21st Century program in Nashua NH for students who participated in grade one (1) and participants who did not attend in grade one (0) (2008-2013) (*n* = 1636)



In addition, I considered level of participation by grade level. Table 14 charts the mean total days and years of participation for students of different participation levels in each grade.

Table 14: Mean years and total days of participation in the 21st Century after school program in Nashua, NH by level of participation and grade level (2008-2013) (*n* = 2917)

		Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Level 1 participants	Mean years	2.6	2.2	2.2	2.1	1.9
	Mean total days	97	60.9	57.1	58.1	47.7
Level 2 participants	Mean years	2.8	2.7	2.6	2.6	2.3
	Mean total days	247.5	232.1	225.1	221.4	174.5

As found above, students who participate in the after school program in grades one or two have higher total days and years of participation in the program than those who don't. Students who participate for at least 30 days in any grade (Level 2) have higher total days of attendance in the program than those that participate at lower levels. Next, I ran t-tests for equality of means looking at grade level of participation and total days of participation in the after school program. Reflecting what is found in Table 13, Level 2 participants at all grade levels had statistically significant *t* scores for total days of attendance compared with Level 1 participants.

Outcome Measures

The outcome measures in this study are students' NECAP scores for Reading and Math. Scores can range from 0-80 with 40 representing the cut point for 'Proficient.' Table 15 includes descriptive information for NECAP scores for all students in this data set, and then breaks it down by participants and non-participants. It is notable that there are slightly more Math scores (5597) in the data set than Reading (5523). This is due to the fact that non-English speakers are not required to take the Reading test during their first year in the country, but are required to take the Math test. Therefore, there are 74 students in this data set with a Math, but not Reading score, for one year of data. Furthermore, there are more scores for participants (3362) than for non-participants (2161) represented in the NECAP data. Figures 10 and 11 illustrate the distribution of Reading and Math scores for the full data set, and Figures 12 and 13 include box plots comparing Reading and Math scores for participants and non-participants.

Table 15: Descriptive statistics for NECAP scores (Reading & Math) for students who did and did not participate in the 21st Century after school program in Nashua, NH between 2008 and 2013 (n = 2917)

	Mean	Standard Deviation	Range	25th percentile	Median (50th percentile)	75th percentile
Reading – all students	43.3	11.2	0-80	37	44	50
Math – all students	41.5	11	0-80	35	42	49
Reading – participants	42.8	11	0-80	36	43	50
Reading – non-participants	44***	11.4	0-80	38	45	51
Math – Participants	41.4	10.9	0-80	35	42	49
Math – non-participants	42	10.9	0-80	35	43	49

*** Difference in means between treatment and control group is significant at the $p < .001$ level .

Figure 10&11: Histograms of the distribution of NECAP scores for Reading and Math for five elementary schools in Nashua, NH (2009-2014) (*n* = 2917)

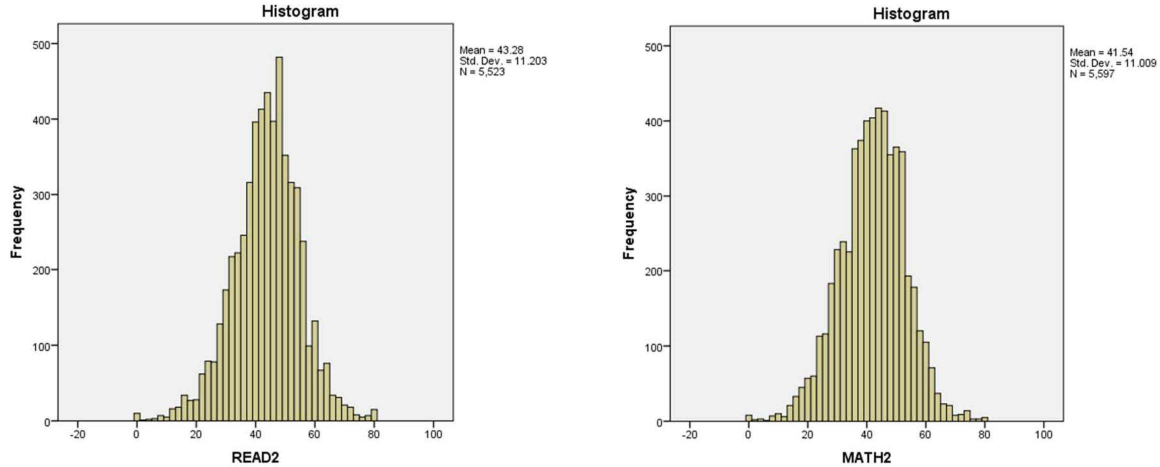
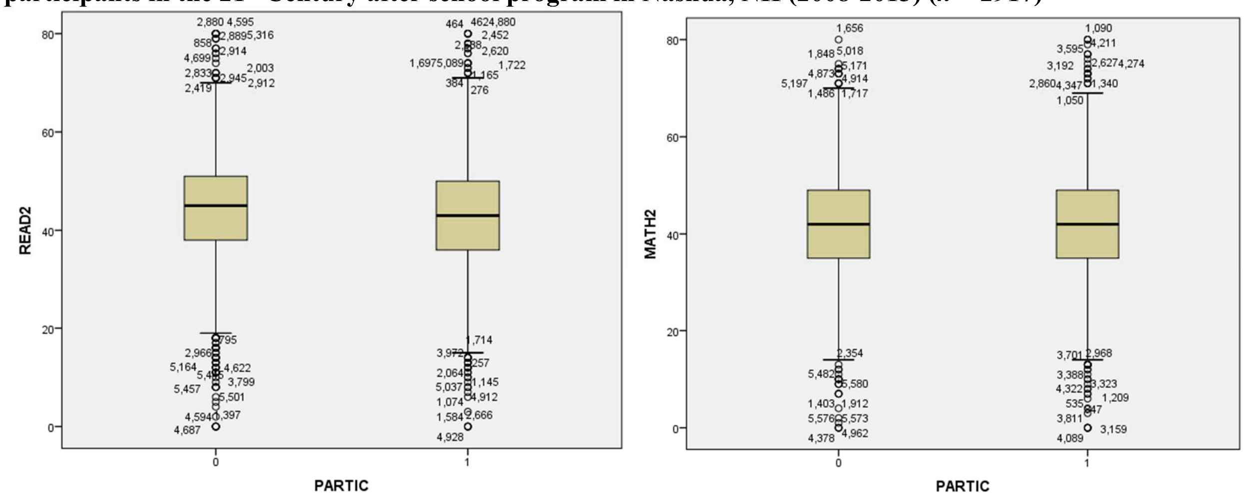


Figure 12&13: Box plots of the distribution of Reading and Math NECAP scores for participants and non-participants in the 21st Century after school program in Nashua, NH (2008-2013) (*n* = 2917)



As found in Table 15 above, students, on average, had higher scores on the Reading assessment than the Math assessment during this period of time (2009-2014). This was also true for participants and non-participants. In this sample, non-participants had higher scores in Math and Reading than did participants, on average. This gap was wider in Reading (1.2 point difference) than Math (0.6 point difference). The difference in means for Reading scores is significant at the $p < .001$ level, while the difference in means for Math is not significant at the p

< .05 level. The distribution for Reading and Math scores is distributed normally, as found in Figures 10 and 11 above.

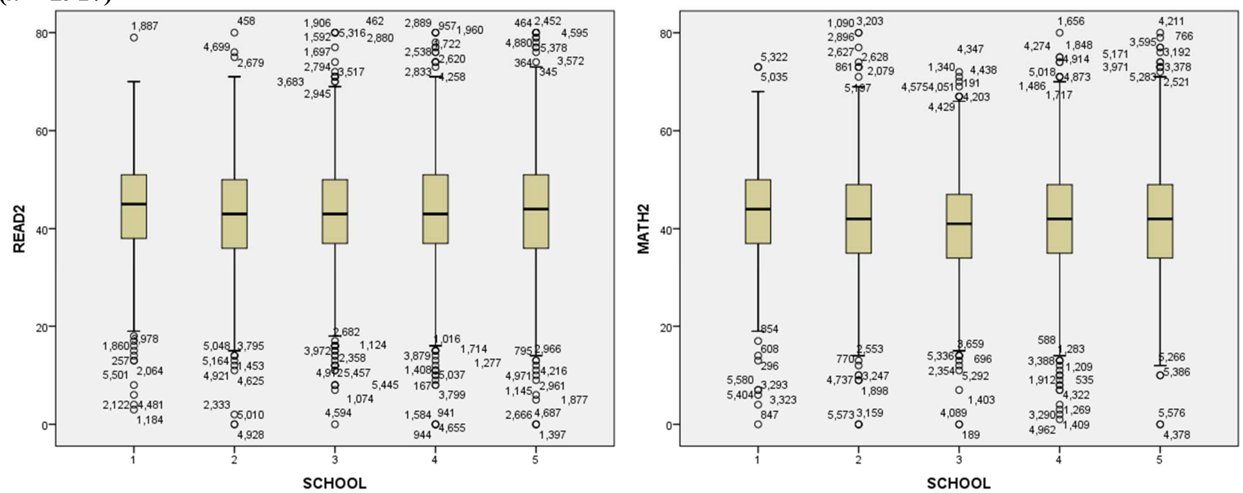
Reading scores remained stable over the span of this study. The lowest scores are found in 2013 with a mean Reading score of 41.8 in this sample, and the highest scores are in 2012 with a mean Reading score of 45.4. Math scores were similarly consistent with the lowest average scores found in 2014 with a mean Math score of 39.8 in this sample, and the highest scores in 2012, with a mean Math score of 44.3.

Table 16 contains mean NECAP scores for Reading and Math for participants and non-participants broken down by school of attendance. These represent mean scores over the period 2009-2014. The only significant differences of note between schools is that School 1 had significantly higher math scores, on average, than the other schools ($t(6366) = -3.62$, $p < .001$,) and School 2 had significantly lower Reading scores, on average, than the other schools ($t(6366) = 2.26$, $p < .05$). Figures 14 and 15 visually represent the distribution of NECAP scores by school.

Table 16: Mean NECAP scores (Reading & Math) for students who did and did not participate in the 21st Century after school program in Nashua, NH, by school (2009-2014) (*n* = 2917)

	School 1	School 2	School 3	School 4	School 5
Reading – Participants	42.9	42.5	42.6	42	44.3
Reading – Non-participants	45.2	42.8	43.5	45.9	42.3
Math – Participants	42.7	41.7	40.7	40.3	42.7
Math – Non-participants	43.7	41.8	40.4	44.4	40

Figures 14&15: Box plots of the distribution of NECAP scores for Reading and Math, by school (2009-2014) (*n* = 2917)

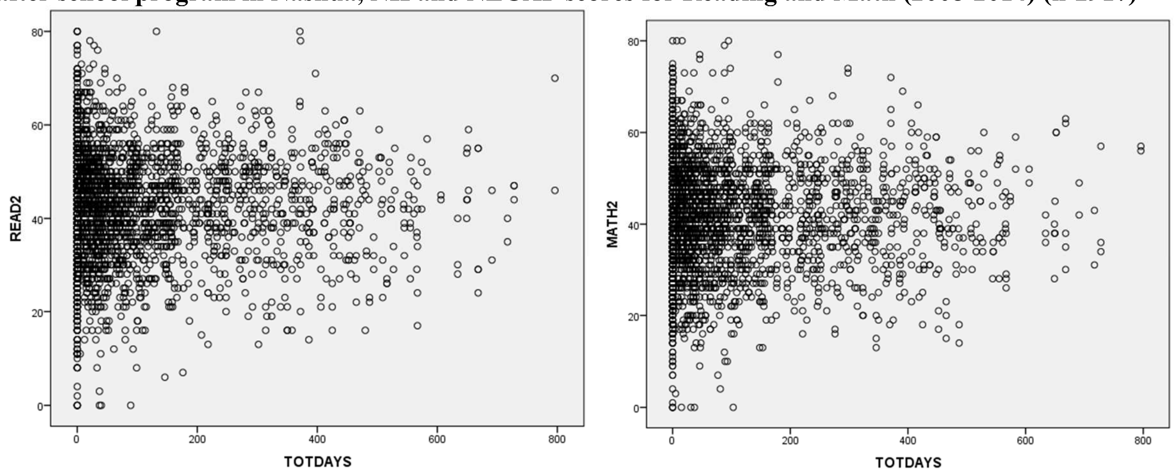


The trends found in the overall data hold true as well when data is broken down by school. Overall, Reading scores are higher than Math scores, and non-participants scored higher than participants. The exception to this is School 5 where participants outscored non-participants in both Reading and Math.

Relational Analyses

Before fitting models, I examined the relationships between key variables, such as student characteristics, participation, and academic achievement. Figures 16 and 17 illustrate the relationship between total days of after school program attendance and math and reading achievement.

Figures 16&17: Scatterplots illustrating the relationship between total days of attendance in the 21st Century after school program in Nashua, NH and NECAP scores for Reading and Math (2008-2014) (n-2917)



As can be found in the scatter plots above, there is no clear relationship between intensity of participation in the 21st Century program and achievement on the Math or Reading NECAP test. Non-participants (zero total days) have scores ranging from 0 to 80. There are students with high levels of participation with both high and low scores in Reading and in Math. I also ran *t*-tests for equivalence of means for grade level of participation and Math and Reading scores. Participation in the 21st Century program in grade two was positively related to Math scores ($t(6366) = -2.14$, $p < .05$). No other grade levels of participation were related to Reading or Math scores at a significant level.

Table 17 contains the mean Math and Reading NECAP scores for each demographic category. Several significant relationships were found when comparing mean achievement scores for different demographic groups. Males scored lower in Reading, on average, than females ($t(5521) = 7.6$, $p < .001$), and higher in Math, on average ($t(5595) = -4.09$, $p < .001$). Students identified as Hispanic scored lower, on average, on the Reading and Math assessments ($t(5521) = 10.89$, $p < .001$; $t(5595) = 10.61$, $p < .001$). There was not a significant difference in average Reading test scores for African Americans when compared with other students in this sample, however, African Americans did score lower, on average, in Math ($t(5595) = 4.84$, $p < .001$). Students who qualified for free or reduced price lunch scored lower on average on Reading and Math ($t(5521) = 15.46$, $p < .001$; $t(5595) = 16.18$, $p < .001$). Finally, students identified as English Language Learners (ELL) or as a student in special education scored significantly lower in Reading and Math, on average, than students not in these programs (ELL=$t(5521) = 16.1$, $p < .001$; $t(5595) = 13.84$, $p < .001$; SpEd=$t(5521) = 24.32$, $p < .001$; $t(5595) = 17.6$, $p < .001$). The table below contains the mean scores for each group for Reading and Math:

Table 17: Mean scores on the NECAP Reading and Math assessment by demographic category for Nashua, NH (2009-2014) (n = 2917)

		Mean score for test group	**Mean score for all others**
Male	Reading***	42.2	44.4
	Math***	42.1	40.9
Hispanic	Reading***	40.9	44.4
	Math***	39.3	42.6
African American	Reading	43	43.3
	Math***	39.2	41.8
Free/Reduced Lunch	Reading***	41.9	47
	Math***	40.1	45.3
Poverty	Reading***	41.7	46.3
	Math***	40	44.5
ELL	Reading***	34.9	44
	Math***	34.4	42.2
SpEd	Reading***	33.4	44.8
	Math***	34.6	42.6

*** Difference in means is significant at the $p < .001$level.

Next, I examined estimated correlations between demographic and participation variables and NECAP scores in Reading and Math (see Table 18 below). NECAP Math scores, in this study, were positively correlated with student identification as a Male ($r(3782) = .055$) and with participation in the 21st Century after school program in grade one ($r(3782) = .038$) (both significant at the .01 level.) On the other hand NECAP math scores were negatively correlated, significant at the .01 level, with student identification as Hispanic ($r(3782) = -.140$), African

American ($r(3782) = -.068$), Free/reduced lunch ($r(3782) = -.211$), Free lunch (Poverty) ($r(3782) = -.196$), English Language Learner (ELL) ($r(3782) = -.192$), or Special Education ($r(3782) = -.248$). Math scores were also negatively correlated with student participation in the 21st Century program in 2010 and 2011($r(3782) = -.049$; $r(3782) = -.057$). No significant correlation was found between math scores and school of attendance, or total days or years attending the 21st Century program.

There were similar correlations between reading NECAP scores and the demographic and participation variables. Reading scores were positively correlated with participation in the 21st Century after school program in grade one ($r(3782) = .049$, $p < .01$). Reading scores, however, were negatively correlated (significant at the .01 level) with student identification as a Male ($r(3782) = -.102$), Hispanic ($r(3782) = -.145$), Free/Reduced lunch ($r(3782) = -.204$), Poverty ($r(3782) = -.195$), ELL ($r(3782) = -.212$), or Special Education ($r(3782) = -.348$). Reading scores were also negatively correlated with some variables related to participation in the 21st Century program, including participation in 2010, 2011, 2012, or 2013, and total days ($r(3782) = -.055$) and years ($r(3782) = -.068$) of participation. No correlation was found in this study between Reading scores and school of attendance or student identification as African American.

Next, correlations between participation and demographic variables were estimated. Total days of participation was positively correlated in this study with student identification as Hispanic ($r(3782) = .209$), African American ($r(3782) = .070$), Free/Reduced Lunch ($r(3782) = .116$), Poverty ($r(3782) = .085$), and ELL ($r(3782) = .058$) (significant at the .01 level). Total years of program attendance had similar positive correlations with student identification as Hispanic ($r(3782) = .157$), African American ($r(3782) = .040$), Free/Reduced Lunch ($r(3782) = .102$), and Poverty ($r(3782) = .070$); however, ELL status was not correlated with total years of attendance, and Special Education status was correlated with total years of attendance ($r(3782) = .026$, $p < .05$), but not with total days. Participation in the 21st Century program in grade one was positively correlated with school of attendance, NECAP Math and Reading scores, and student identification as Hispanic, African American, Free/Reduced Lunch, Poverty, ELL, and Special Education.

Table 18: Estimated bivariate correlations between demographic factors, participation in the 21st Century after school program in Nashua, NH and NECAP scores in Reading and Math (2008-2014) (*n* = 2917)

	School	**Male**	**Hisp**	**AfAm**	**FRL**	**POV**	**ELL**
Math	-.018	.055**	-.140**	-.068**	-.211**	-.196**	-.192**
Read	.014	-.102**	-.145**	.009	-.204**	-.195**	-.212**
TotD	-.034**	-.005	.209**	.070**	.116**	.085**	.058**
Years	.049**	-.008	.157**	.040**	.102**	.070**	.011
P 13	.009	-.012	.161**	.052**	.059**	.036**	.093**
P 12	.038**	-.023	.196**	.048**	.114**	.087**	.090**
P 11	-.034**	.002	.135**	.014	.101**	.084**	.007
P 10	.008	-.033	.106**	.028*	.093**	.077**	-.014
P 09	-.009	.012	-.017	.033**	.040**	.027*	-.073**
Gr 1	.037**	.005	.067**	.031*	.040**	.026*	.106**
Sped	.016	.122**	-.043**	.014	.062**	.066**	.028*
ELL	.044**	-.045**	.388**	-.074**	.154**	.154**	1
POV	-.034**	-.051**	.228**	.061**	.841**	1	
FRL	-.050**	-.039**	.263**	.062**	1		
AfAm	-.019	-.021	-.178**	1			
Hisp	.012	.001	1				
Male	-.014	1					
School	1						

SpEd	Gr. 1	P 09	P 10	P 11	P 12	P 13	Years	Tot D	Read	Math
-.248**	.038**	-.013	-.049**	-.057**	-.005	.009	-.025	-.019	.680**	1
-.348**	.049**	-.024	-.071**	-.087**	-.029*	-.040**	-.068**	-.055**	1	
-.003	.105**	.357**	.513**	.648**	.653**	.527**	.765**	1		
.026*	.066**	.571**	.651**	.689**	.600**	.485**	1			
-.022	.318**	-.135**	-.004	.251**	.527**	1				
-.023	.240**	-.014	.117**	.417**	1					
.043**	-.042**	.189**	.386**	1						
.064**	-.130**	.501**	1							
.013	-.139**	1								
-.047**	1									
1										

** Correlation is significant at the .01 level (2-tailed).
*Correlation is significant at the .05 level (2-tailed).

Next, I disaggregated the data and looked only at relationships between students who participated in the 21st Century program in the years leading up to the NECAP. Table 19 presents the results of these analyses:

Table 19: Percentage of students from each grade level participating in the 21st Century program by grade of NECAP administration, as well as estimated bivariate correlations between 21st Century program participation by grade level and Math and Reading scores on the NECAP (2009-2014)

Students who took the Grade 3 NECAP (*n* = 1733)			
Grade one participants	**Percentage participating at different levels**	**Correlation with Grade 3 Reading NECAP**	**Correlation with Grade 3 Math NECAP**
• Level 0 • Level 1 • Level 2	81.4% 6.3% 12.3%	-.045	-.019
Grade two participants			
• Level 0 • Level 1 • Level 2	68.7% 10.6% 20.7%	-.039	.008
Students who took the Grade 4 NECAP (*n* = 1873)			
Grade one participants	**Percentage participating at different levels**	**Correlation with Grade 4 Reading NECAP**	**Correlation with Grade 4 Math NECAP**
• Level 0 • Level 1 • Level 2	86.9% 4.9% 8.2%	-.06*	-.005
Grade two participants			
• Level 0 • Level 1 • Level 2	77.3% 8.5% 14.2%	-.01	.039
Grade three participants			
• Level 0 • Level 1 • Level 2	70.1% 10.5% 19.4%	-.061**	-.031
Students who took the Grade 5 NECAP (*n* = 1973)			
Grade one participants	**Percentage participating at different levels**	**Correlation with Grade 5 Reading NECAP**	**Correlation with Grade 5 Math NECAP**
• Level 0 • Level 1 • Level 2	87.4% 5% 7.6%	-.096**	-.038
Grade two participants			
• Level 0 • Level 1 • Level 2	83.1% 6.6% 10.3%	-.044	-.016
Grade three participants			
• Level 0 • Level 1 • Level 2	77.3% 7.6% 15.1%	-.054*	-.042
Grade four participants			
• Level 0 • Level 1 • Level 2	69.5% 13.2% 17.3%	-.062**	-.035

** Correlation is significant at the .01 level (2-tailed).

*Correlation is significant at the .05 level (2-tailed).

Looking at Table 19, we again see higher rates of participation among third and fourth graders. The strength and significance of correlations between participation and Reading and Math scores is different when looking at it here by grade level of participation than when looking by year of participation. While most of the correlations are still negative, very few are significant at the .05 or .01 levels. There are no significant correlations between participation in the 21st Century program in grades one or two and Reading or Math scores in grade three. There are no significant correlations between participation and Math scores at any level. For students taking the NECAP in grade 4, we do see significant negative correlations between participation in 21st Century in grade one ($r(1871) = -.060$, $p < .05$) and three ($r(1871) = -.061$, $p < .01$) with reading scores, but not with grade two participation. For students taking the NECAP in grade 5, there are significant negative correlations between Reading achievement and participation in the 21st Century program in grades one ($r(1971) = -.096$, $p < .01$), three ($r(1971) = -.054$, $p < .05$), and four ($r(1971) = -.062$, $p < .01$).

Balancing the Data Set

Differences in demographic characteristics seem to impact the choice to participate in the after school program (see Table 8). Demographic differences are also correlated with performance on the outcome measures, the Math and Reading NECAP scores (see Table 18). Demographic differences between the two groups are therefore a confounding factor and this could impact my ability to reasonably estimate the treatment effect of attendance in the after school program on achievement outcomes.

From the descriptive analysis thus far, it is clear that the 21st Century after school programs in Nashua during this time enrolled a disproportionate number of students identified as Hispanic, and students who qualified for free or reduced price lunch. Enrolling underserved populations such as these is a stated goal of the federal 21st Century grant (U.S. DOE, 2003). Students identified as Hispanics or students qualifying for free or reduced price lunch performed lower, on average, on the NECAP Reading and Math assessment. Looking at initial analyses, participants in the 21st Century program in Nashua performed lower on the NECAP assessments than non-participants. Teasing out achievement effects predicted by after school participation versus demographic predictors is an important consideration in later analyses. These relationships are further explored using ordinary least squares regression analysis and multilevel modeling to determine effects of participation in the 21st Century after school program on student achievement in Reading and Math. Forming the data described above into matched treatment and control groups is the first critical step of this process.

As described in Chapter 3, propensity scores were estimated for each student and then these scores were used to compute an inverse propensity weight for each student in the control group. Given the recommendation to include all relevant demographic characteristics in the propensity score model (Ho et al., 2007; Thoemmes, 2012), I included the following variables in the logistic regression model for each year of data: *Male, African American, Hispanic, Free/Reduced Lunch, Poverty, SpEd, ELL, Grade, and School.* Variables were included one at a time, with the goodness of fit examined after each addition using the -2Log Likelihood statistic. I conducted tests of the change in the -2 Log Likelihood statistic between subsequent models. Equations of the best fitting models with associated goodness-of-fit statistics by year are presented in Table 20.

Table 20: Best fitting binary logistic regression models estimating the propensity for a student to participate in the 21st Century Program based on demographic characteristics for school years 2009-2013 (*n* = 2917)

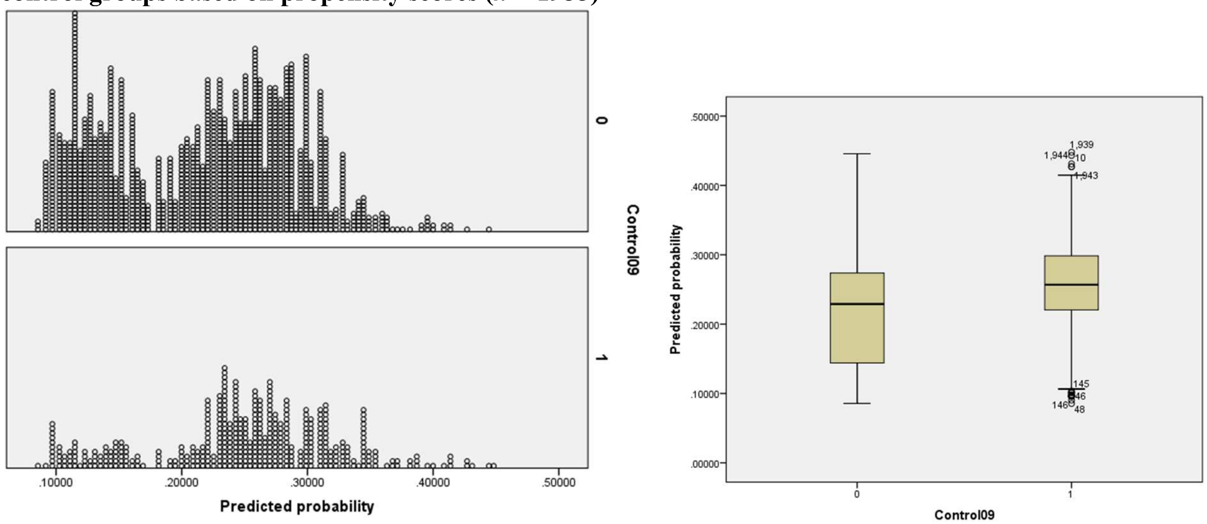
YEAR	BEST FIT BINARY LOGISTIC REGRESSION MODEL	-2LOG LIKELIHOOD STATISTIC
2009	Estimated LOG(ODDS)= -2.117-.076(SCHOOL)+.059(MALE) +.182(HISPANIC) +.431(AFAM)+.864(FRL) - .009(POVERTY)+.268(ELL) - .115(SPED) +.126(GRADE)	2017.214

2010	Estimated LOG(ODDS)= -2.206+.742FRL+.082ELL-.107SCHOOL-.159POV +.872HISP+.370SPED+.538AFAM-.084MALE+.116GRADE	1800.762
2011	Estimated LOG(ODDS)= -1.948+.864FRL+.567HISP-.212POV-.161SCH +.29SPED-.033MALE+.142AFAM+.041GRADE+.019ELL	1658.823
2012	Estimated LOG(ODDS)= -1.926+.393FRL-.083POV+.595HISP +.097ELL +.672AFAM-.078SPED-.066GRADE-.07MALE +.017SCHOOL	1625.010
2013	Estimated LOG(ODDS)= -1.078+.285FRL-.353POV+.588HISP +.903AFAM -.028ELL+.037SPED-.089MALE-.069GRADE-.036SCHOOL	1464.959

The logistic regression models indicate which demographic groups of students are more likely to have attended the 21st Century program during each academic year. In looking at the fitted models above, there are some consistencies across years. Students qualifying for free or reduced price lunch, Hispanics, and African Americans were more likely to be in the program for each year of data. Other characteristics fluctuate year to year. Most years, males were less likely to attend the program, however, in 2009, they were more likely to attend. In 2009-2011, older students were more likely to attend, but this switched in 2012 and 2013 to younger students being more likely to attend. Students receiving special services, such as ELL or Special Education also varied year to year. Students in the English Language Learner program were more likely to attend every year except 2013, and students enrolled in Special Education were more likely to attend every year except 2009 and 2012.

Once propensity scores were estimated, I checked to see that there was common support between the treatment and control groups, meaning that there was sufficient overlap between the estimated propensity scores between the two groups. This was checked for each year by examining box plots and dot plots displaying the distribution of propensity scores for the treatment and control groups. Common support between the treatment and control group for 2009 is demonstrated in Figures 18 and 19 below. Common support was established for each year of data (see Appendix B for figures for subsequent years of data).

Figures 18 & 19: Dot plot (18) and Box plot (19) demonstrating common support between the treatment and control groups based on propensity scores (*n* = 1955)



Inverse Propensity Weights (IPW) were then calculated for all students in the Control group for each year of data (P/(1-P)). I checked the range of weights for the controls for each year. Weights higher than .99 were removed from the data set as they can create bias (Guo & Fraser, 2014). After examining weights, four students were removed from the control group for 2013- #335,446, 880, 1216 – who all had weights greater than one (1.01-1.47). All other weights fell under .99. Students in the Treatment group were assigned a weight of one. The resulting weights were used to conduct subsequent analyses.

Finally, I rechecked the demographic information for each year to see if balance was achieved. After balancing, no statistically significant differences were found in the means for any of the demographic variables. Distributions for Grade and School were examined using histograms, and no noticeable differences were observed. The distribution of propensity scores was also examined using box plots and dot plots. Distributions looked equivalent, and there were no statistically significant differences in mean propensity scores between the treatment and control groups for any year of this study. Year by year data is displayed below in Tables 21 through 25.

Table 21: Mean demographic characteristics for treatment and control groups for 2008-2009 school year before and after balancing using Inverse Propensity Weights (*n* = 1955)

	N	%MALE	%HISP	%AFAM	%FRL	%POV	%ELL	%SPED	SCHL	GRADE	PROP SCORE
Control	1516	50	17	8**	64***	57***	5	14	3.13*	2.92	.2171***
Treat	439	51	19	12	81	72	7	14	2.99	3.05	.2502
After weighting on Inverse Propensity Scores											
Control	439	51	19	12	81	72	7	14	3.00	3.04	.2484
Treat	439	51	19	12	81	72	7	14	2.99	3.05	.2502

* Difference in means between treatment and control group is significant at the $p < .05$ level .
** Difference in means between treatment and control group is significant at the $p < .01$ level .
*** Difference in means between treatment and control group is significant at the $p < .001$ level .

Table 22: Mean demographic characteristics for treatment and control groups for 2009-2010 school year before and after balancing using Inverse Propensity Weights (*n* = 1893)

	N	%MAL	%HISP	%AFM	%FRL	%POV	%ELL	%SPD	SCHL	GRAD	PROP SCR
Control	1519	50	21***	8	68***	61***	8**	12*	3.19*	3.12	.1889***
Treat	374	48	37	11	83	74	13	17	3.02	3.08	.2328
After weighting on Inverse Propensity Scores											
Control	373	48	37	11	82	73	13	16	3.03	3.08	.2314
Treat	374	48	37	11	83	74	13	17	3.02	3.08	.2328

* Difference in means between treatment and control group is significant at the $p < .05$ level .
** Difference in means between treatment and control group is significant at the $p < .01$ level .
*** Difference in means between treatment and control group is significant at the $p < .001$ level .

Table 23: Mean demographic characteristics for treatment and control groups for 2010-2011 school year before and after balancing using Inverse Propensity Weights (*n* = 1848)

	N	%MALE	%HISP	%AFAM	%FRL	%POV	%ELL	%SPED	SCHL	GRADE	PROP SCR
Control	1523	50	29***	9	71***	64***	11*	12	3.24**	3.14	.1701***
Treat	325	49	44	9	85	76	16	15	2.97	3.10	.2027
After weighting on Inverse Propensity Scores											
Control	325	49	43	9	85	76	16	15	2.98	3.10	.2021
Treat	325	49	44	9	85	76	16	15	2.97	3.10	.2027

* Difference in means between treatment and control group is significant at the $p < .05$ level .
** Difference in means between treatment and control group is significant at the $p < .01$ level .
*** Difference in means between treatment and control group is significant at the $p < .001$ level .

Table 24: Mean demographic characteristics for treatment and control groups for 2011-2012 school year before and after balancing using Inverse Propensity Weights (*n* = 1764)

	N	%MALE	%HISP	%AFAM	%FRL	%POV	%ELL	%SPED	SCHL	GRADE	PROP SCR
Control	1445	50	37***	9*	74**	67**	15**	11	3.18	3.16*	.1766***
Treat	319	48	52	13	83	75	22	10	3.18	2.98	.2000
After weighting on Inverse Propensity Scores											
Control	320	48	53	13	83	75	22	10	3.18	2.97	.2015
Treat	319	48	52	13	83	75	22	10	3.18	2.98	.2000

* Difference in means between treatment and control group is significant at the $p < .05$ level .
** Difference in means between treatment and control group is significant at the $p < .01$ level .
*** Difference in means between treatment and control group is significant at the $p < .001$ level .

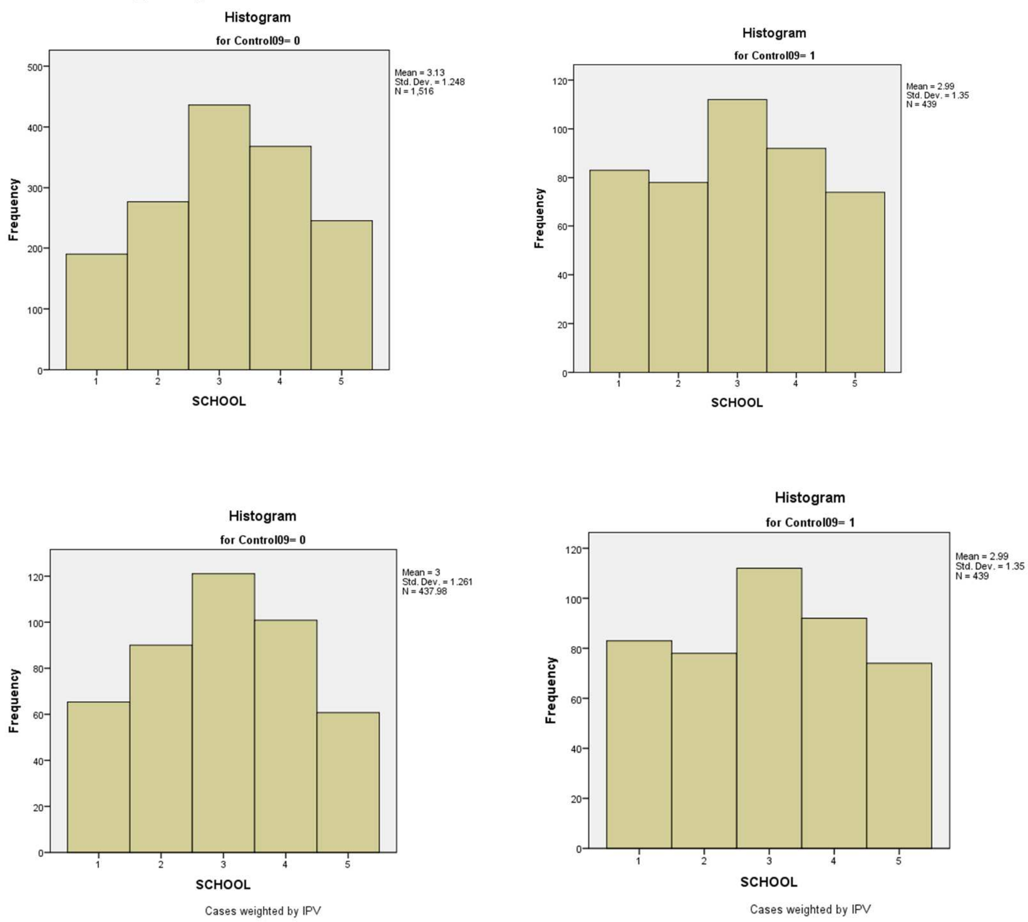
Table 25: Mean demographic characteristics for treatment and control groups for 2012-2013 school year before and after balancing using Inverse Propensity Weights (*n* = 1355)

	N	%MALE	%HISP	%AFAM	%FRL	%POV	%ELL	%SPED	SCHL	GRADE	PROP SCR
Control	1028	50	40**	8**	76	69	18	10	3.18	3.58	.2344***
Treat	327	48	50	14	79	70	21	10	3.11	3.50	.2595
After weighting on Inverse Propensity Scores											
Control	324	48	50	13	79	70	22	10	3.11	3.51	.2584
Treat	327	48	50	14	79	70	21	10	3.11	3.50	.2595

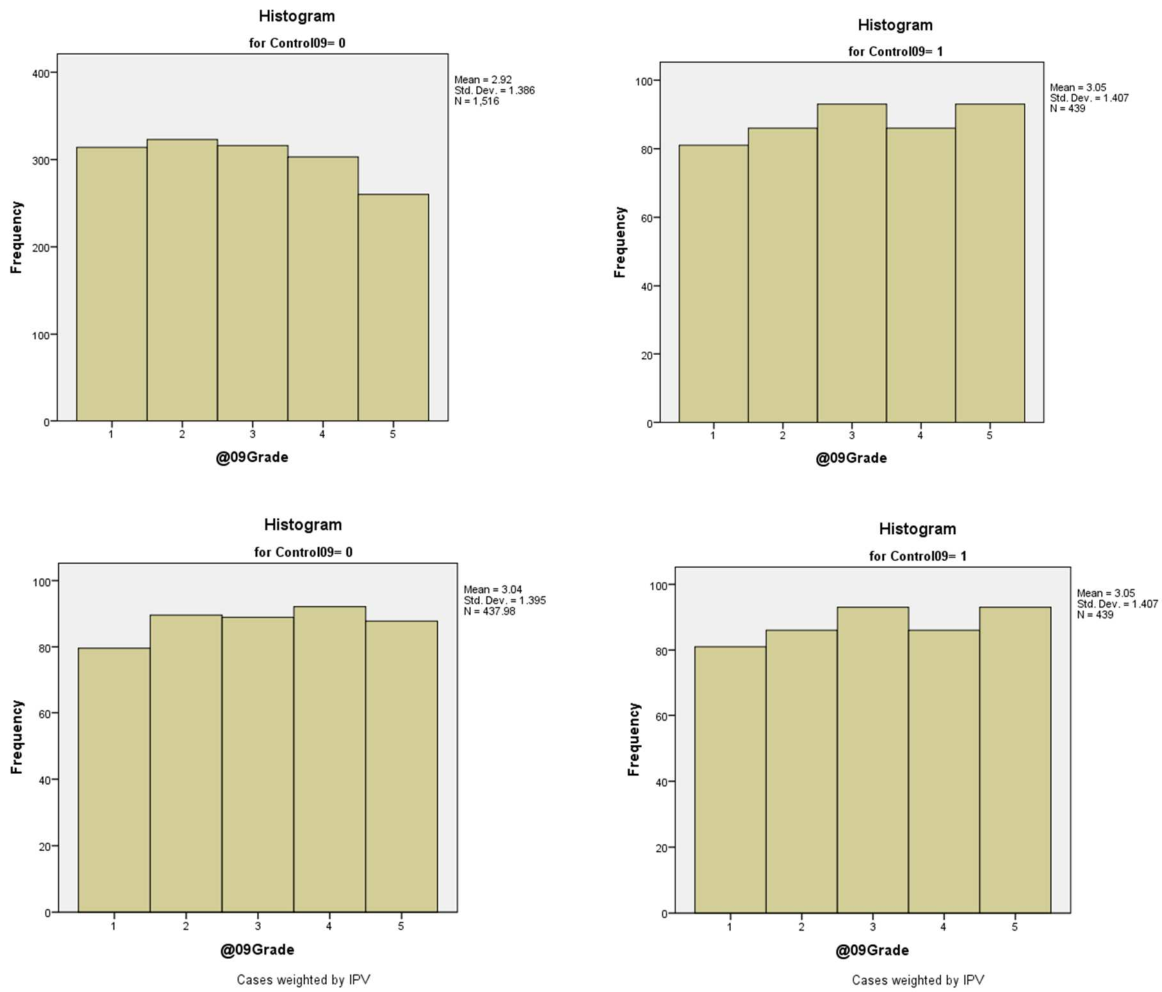
** Difference in means between treatment and control group is significant at the $p < .01$ level .
*** Difference in means between treatment and control group is significant at the $p < .001$ level .

It is clear from the tables above that balance was improved on all measures, with the treatment and control groups looking nearly identical on all key demographic variables. The histograms below display the distribution of Schools (Figures 20-23) and Grades (Figures 24-27) for the treatment and control groups before and after weighting for 2009.

Figures 20-23: Distribution among schools for control and treatment groups before (20&21) and after (22&23) weighting for 2009 (*n* = 1955)

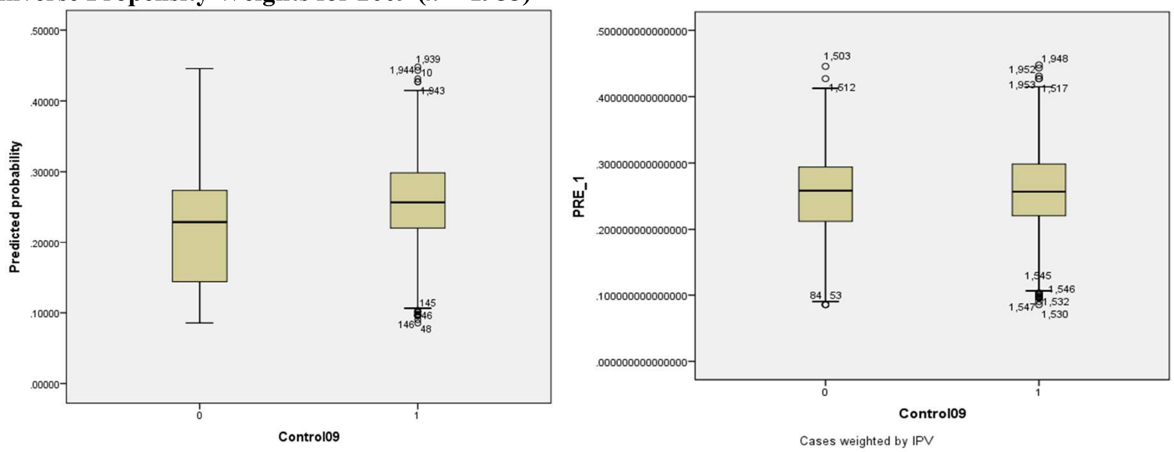


Figures 24-27: Distribution among grade level for control and treatment groups before (24&25) and after (26&27) weighting for 2009 (*n* = 1955)



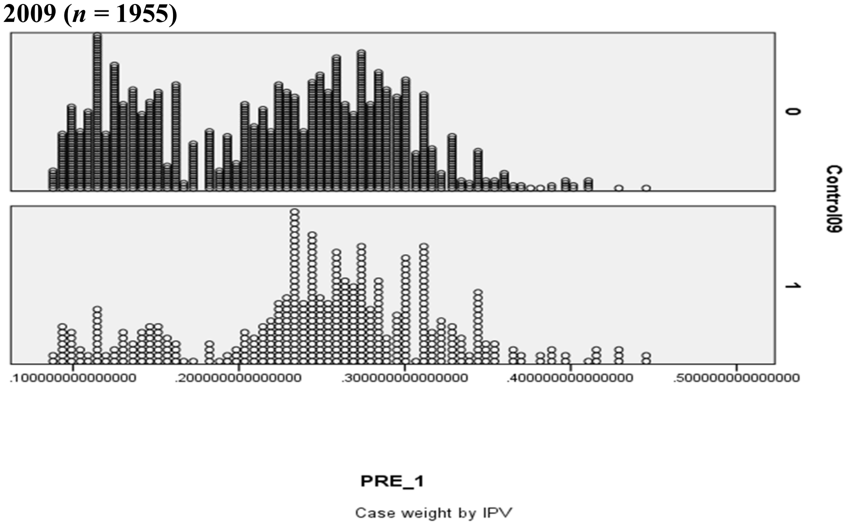
As a next point of comparison, I compared the distributions of the propensity scores for the treatment and control groups before and after matching. It is easy to see from the box plots in Figures 28 and 29 that, with the use of IPW, the mean propensity scores for the treatment and control groups are more similar, as is the distribution of propensity scores between the 25^{th} and 75^{th} quartiles, as illustrated by the shaded box in each diagram.

Figures 28&29: Box Plots comparing distributions of propensity scores before and after application of Inverse Propensity Weights for 2009 (*n* = 1955)



As a final check, I recreated the dot plot of propensity scores for the treatment and control groups, with the IPW weights applied to ensure that the weighted distributions look comparable. From looking at Figure 30 below, it is evident that the distributions are comparable. Comparative dot and box plots for subsequent years of data are included in Appendix C.

Figure 30: Dot plot of the distribution of propensity scores for the treatment and weighted control groups for 2009 (*n* = 1955)



<u>Causal Analyses</u>

Now that the treatment and control groups were statistically balanced on key demographic variables, I could conduct statistical analyses to answer my research questions.

<u>Question 1:</u> Do students who attend 21st Century After school programs perform better academically, on average, based on scores on the grades 3-6 Math and Reading NECAP tests, than students in the same schools who do not attend these after school programs?

Question one examines the effect of participation in the 21st Century after school program on math and reading achievement. Participation is considered in the broadest sense here, with a student either being a participant, or not a participant, for each year of the study. The federal government uses a threshhold of 30 days of attendance in the program to define a participant, therefore the same criterion was used here. Table 26 includes the total number of students, as well as the percentage of participants for each year of this study.

Table 26: Total students in sample and participation rate, by year (2008-2013)

	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013
Total students	1559	1876	1627	1499	1316
% Participants	23	20	18	18	25
% Non-Participants	77	80	82	82	75

To determine the impact of participation, I fit a series of multiple regression models for math and reading for each year, controlling for school, grade, and key demographic factors, as shown in Table 27.

Table 27: Variables included in taxonomy of fitted regression models

Model 1	Participation (0,1)
Model 2	Model 1 + School of attendance – Dummy variables for Schools 1, 2, 4, and 5 School 3 serves as reference point
Model 3	Model 2 + Current Grade (1-5)
Model 4	Model 3 + Key Demographics – Gender (MALE); Race (Dummy variables for Hispanic & African American – White as reference point); Poverty Status (FRL); Special programming (Designation as English Language Learner (ELL) or student with a disability (SPED))

Doing this resulted in five fitted models for math achievement and five for reading achievement. These models are represented below in Tables 28 and 29, and subsequently discussed. The taxonomy of fitted models for each year for math and reading are included in Appendix D. In both tables, 'Intercept' represents the predicted score for a White, female, non-participant attending School 3 who does not qualify for free or reduced price lunch or ELL or special education services.

Table 28: Parameter estimates for predicted math achievement, controlling for participation in the 21st Century after school program, school, grade, and demographic factors

	$\widehat{MATHACH}09$	$\widehat{MATHACH}10$	$\widehat{MATHACH}11$	$\widehat{MATHACH}12$	$\widehat{MATHACH}13$
Intercept	48.49***	49.16***	48.87***	46.14***	41.52***
Participation	-1.27~	.34	-1.68*	1.22	.56
School 1	3.55**	2.88*	1.92	3.12~	4.88***
School 2	1.57	1.25	-.88	2.26~	2.52*
School 4	-1.98~	.07	-.29	2.26~	3.46**
School 5	.13	1.18	-.38	.64	.97
Grade	-1.05*	-1.23***	-.73*	-.61	.61~
Male	1.98**	1.79*	.59	2.54**	1.59~
Hispanic	-1.25	-3.22**	-.77	-1.81~	-1.41
African Am	-2.37*	-4.33***	-2.37	-3.65*	-3.73**
FRL	-3.83***	-3.87***	-3.23**	-5.54**	-3.63**
ELL	-7.33***	-5.89***	-5.54***	-6.43***	-5.28***
Special Ed.	-6.27***	-6.64***	-5.24***	-6.80***	-9.73***

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 29: Parameter estimates for predicted reading achievement, controlling for participation in the 21st Century after school program, school, grade, and demographic factors

	$\widehat{READACH}09$	$\widehat{READACH}10$	$\widehat{READACH}11$	$\widehat{READACH}12$	$\widehat{READACH}13$
Intercept	53.56***	49.10***	51.92***	49.31***	47.33***
Participation	-1.30~	.05	-1.49~	.76	-.79
School 1	.50	1.26	.25	.92	4.81***
School 2	-.71	1.21	-3.09**	1.38	4.15***
School 4	-2.99**	.59	-2.12~	2.03	2.91*
School 5	-.74	1.76	-1.57	-.29	2.42~
Grade	-1.42***	-.37	-.35	-.64~	-.12
Male	-1.63*	-1.08	-1.00	-.10	-1.87*
Hispanic	-2.69~	-3.03**	-1.48	-1.91~	-.88
African Am	.38	-1.38	-.75	-1.84	-1.26
FRL	-2.69**	-3.45**	-3.76**	-3.67*	-3.88***
ELL	-8.42***	-6.62***	-4.46**	-7.56***	-4.72***
Special Ed.	-8.95***	-7.77***	-7.28***	-7.67***	-10.12***

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Impact of Participation

The estimated impact of participation on math achievement varies year to year from a negative impact of -1.68 points to +1.22 points. However, the only years in which participation in the 21st Century program had a statistically significant impact on math achievement was in 2009 and 2011. The impact of participation in both of these years is negative. In 2009, the estimated impact of participation is -1.27 points, which is statistically significant at the $p < .10$ level. In 2011, the estimated impact of participation is -1.68 points, which is statistically significant at the $p < .05$ level. This model predicts that a student who participates in the 21st Century program would score 1.68 points lower on the Math NECAP assessment, on average, than a student who had not participated in the program, controlling for grade, school, and key demographic features.

Similar results were found for the estimated impact of participation on reading achievement. Impacts ranged from -1.49 points in 2011 to +.76 points in 2012. Statistically significant impacts of participation were found in 2009 and 2011. In 2009, the estimated impact of participation in the 21st Century program was -1.30 points, which was statistically significant at the $p < .10$ level. In 2011, the estimated impact of participation on reading achievement was -1.49 points, which was significant at the $p < .10$ level. This would estimate that a student participating in the program would be predicted to score 1.49 points lower on the Reading NECAP than a non-participant, controlling for grade, school, and key demographic features.

Impact of Grade Level

The estimated impact of grade is negative for each year except for 2013. This indicates that on average, students score lower on the Math NECAP in higher grades than in lower grades, controlling for participation and school of attendance. Grade had a statistically significant impact on math achievement in 2009 and 2013. In 2009, the estimated impact of grade level is -.99 points, which is statistically significant at the $p < .05$ level. In 2013, the estimated impact of grade level is +.77 points, which is significant at the $p < .05$ level. Similar results are found for reading achievement. In 2009, the estimated impact of grade level is -1.07 points, which is statistically significant at the $p < .01$ level. This model predicts that a fourth grader would score 1.07 points lower, on average, on the Reading NECAP than a third grader, controlling for participation and school.

The tables below display the predicted math and reading achievement for participants and non-participants by grade level for each year. Table 30 delineates estimated results for White, male students living in poverty and attending School 3. These variables were selected to represent a prototypical student because these are the largest groups, demographically, in this study. As discussed, the table presents lower predicted scores in math and reading for participants and non-participants in grade 5 than in grade 2, with the exception of 2013.

Table 30: Estimated math and reading NECAP scores for participants and non-participants in the 21st Century after school program, by grade level, for white, male students living in poverty at School 3, by year (2008-2013)

2009

Math	NON-PARTICIPANTS	PARTICIPANTS	**Reading**	NON-PARTICIPANTS	PARTICIPANTS
GR.2	44.54	43.27	GR.2	46.40	45.10
GR.3	43.49	42.22	GR.3	44.98	43.68
GR.4	42.44	41.17	GR.4	43.56	42.26
GR.5	41.39	40.12	GR.5	42.14	40.84

2010

Math	NON-PARTICIPANTS	PARTICIPANTS	**Reading**	NON-PARTICIPANTS	PARTICIPANTS
GR.2	44.62	44.96	GR.2	43.83	43.88
GR.3	43.39	43.73	GR.3	43.46	43.51
GR.4	42.16	42.50	GR.4	43.09	43.14
GR.5	40.93	41.27	GR.5	42.72	42.77

2011

Math	NON-PARTICIPANTS	PARTICIPANTS	**Reading**	NON-PARTICIPANTS	PARTICIPANTS
GR.2	44.77	43.09	GR.2	46.46	44.97
GR.3	44.04	42.36	GR.3	46.11	44.62
GR.4	43.31	41.63	GR.4	45.77	44.28
GR.5	42.58	40.90	GR.5	45.42	43.93

2012

Math	NON-PARTICIPANTS	PARTICIPANTS	**Reading**	NON-PARTICIPANTS	PARTICIPANTS
GR.2	43.92	45.14	GR.2	44.86	45.62
GR.3	43.31	44.53	GR.3	44.22	44.98
GR.4	42.70	43.92	GR.4	43.58	44.34

GR.5	42.09	43.31	GR.5	42.94	43.70

2013

Math	NON-PARTICIPANTS	PARTICIPANTS	**Reading**	NON-PARTICIPANTS	PARTICIPANTS
GR.2	40.70	41.26	GR.2	41.34	40.55
GR.3	41.31	41.87	GR.3	41.22	40.43
GR.4	41.92	42.48	GR.4	41.10	40.31
GR.5	42.53	43.09	GR.5	40.98	40.19

School Differences

School of attendance has a statistically significant impact on achievement scores in all models except 2012 Reading. With School 3 being used as the reference point, estimated coefficients for each school indicate the estimated difference in score, on average, of students at that school compared to students at School 3, controlling for participation, grade, and demographic features. Students at School 1 scored higher, on average, than the students at School 3. This difference in math achievement was statistically significant for each year of the study, and was significant in reading achievement in 2010 and 2013. As an example, in 2010, a typical, non-participating student at School 1 would be predicted to score 46.27 points on the Math NECAP compared with a typical student at School 3 being predicted to score 43.39 points ($p < .01$ level).

Students at School 2 had statistically significant ($p < .05$ level) higher scores in Math than students at School 3, on average, in 2012 and 2013. Scores in Reading were lower for students at School 2 by -2.77 points, on average, in 2011 ($p < .05$ level), but higher by 4.46 points, on average, in 2013 ($p < .001$ level). School 4 also had mixed results in comparison with School 3. Students at School 4 scored lower on the NECAP assessment, on average, in Reading and Math in 2009 than students at School 3. Students at School 4 would be estimated to score 3.31 points lower on the Reading NECAP than a student at School 3, controlling for participation and grade level ($p < .01$ level). However, in 2013, students at School 4 had significantly ($p <$

.05 level) higher scores, on average, in Reading and Math than students at School 3. A student at School 4 would be estimated to score 3.05 points higher, on average, than a student at School 3, controlling for participation and grade level.

Students at School 5 scored higher than students at School 3 at a statistically significant level in 2010 (Reading and Math), and 2013 (Reading). No other years had statistically significant differences. In 2010, a student at School 5 would be predicted to score 3.18 points higher, on average, on the Reading NECAP than a student at School 3, controlling for grade and participation ($p < .05$ level). Table 31 displays predicted NECAP scores for Reading and Math for each year of the study for White, male third graders living in poverty. The scores are predicted for participants and non-participants in the 21st Century after school program, by school.

Table 31: Predicted Math and Reading NECAP scores for prototypical students, by school and participation (2009-2013)

2009

	SCHOOL	NON-PARTICIPANTS	PARTICIPANTS		SCHOOL	NON-PARTICIPANTS	PARTICIPANTS
Math	1	47.04	45.77	**Reading**	1	45.48	44.18
	2	45.06	43.79		2	44.27	42.97
	3	43.49	42.22		3	44.98	43.68
	4	41.51	40.24		4	41.99	40.69
	5	43.62	42.35		5	44.24	42.94

2010

	SCHOOL	NON-PARTICIPANTS	PARTICIPANTS		SCHOOL	NON-PARTICIPANTS	PARTICIPANTS
Math	1	46.27	46.61	**Reading**	1	44.72	44.77
	2	44.64	44.98		2	44.67	44.72
	3	43.39	43.73		3	43.46	43.51
	4	43.46	43.80		4	44.05	44.10
	5	44.57	44.91		5	45.22	45.27

2011

	SCHOOL	NON-PARTICIPANTS	PARTICIPANTS		SCHOOL	NON-PARTICIPANTS	PARTICIPANTS
Math	1	45.96	44.28	**Reading**	1	46.36	44.87
	2	43.16	41.48		2	43.02	41.53
	3	44.04	42.36		3	46.11	44.62
	4	43.75	42.07		4	43.99	42.50
	5	43.66	41.98		5	44.54	43.05

2012

	SCHOOL	NON-PARTICIPANTS	PARTICIPANTS		SCHOOL	NON-PARTICIPANTS	PARTICIPANTS
Math	1	46.43	47.65	**Reading**	1	45.14	45.90
	2	45.57	46.79		2	45.60	46.36
	3	43.31	44.53		3	44.22	44.98
	4	45.57	46.79		4	46.25	47.01
	5	43.95	45.17		5	43.93	44.69

2013

	SCHOOL	NON-PARTICIPANTS	PARTICIPANTS		SCHOOL	NON-PARTICIPANTS	PARTICIPANTS
Math	1	46.19	46.75	**Reading**	1	46.03	45.24
	2	43.83	44.39		2	45.37	44.58
	3	41.31	41.87		3	41.22	40.43
	4	44.77	45.33		4	44.13	43.34
	5	42.28	42.84		5	43.64	42.85

Student Characteristics

The final consideration in question one is the impact of key demographic features on math and reading achievement. Gender, race, poverty status and special learning needs are considered. In the analyses conducted, males had higher math scores, on average, than females, and lower reading scores, on average, controlling for participation, grade, and school of attendance. This gender gap is statistically significant in four of the five years of analysis for math, with the exception of 2011. All differences are positive, and range from an estimate of 1.59 points (2013, $p < .10$) to 2.54 (2012, $p < .01$). Based on this model, it would be predicted that a male student would score 2.54 points higher on the 2012 Math NECAP than a female student, on average, controlling for participation, grade, school of attendance, and other demographic factors. In Reading, females score higher than males, on average, at a statistically significant level in 2009 (1.63 points, $p < .05$), and 2013 (1.87 points, $p < .05$).

On average, students of color are predicted to score lower on the Math NECAP based on the models generated in question one. African American students have significantly lower scores than White students in 2009 (-2.37 points), 2010 (-4.33 points), 2012 (-3.65 points), and

2013 (-3.73 points). Hispanic students have significantly lower scores, on average, than White students in 2010 (-3.22 points) and 2012 (-1.81 points). On the Reading assessment, Hispanic students scored lower at a statistically significant level than White students in 2009, 2010, and 2012 (-2.69, -3.03, and -1.91 points, respectively). For example, based on 2010 data, a Hispanic student would be predicted to score 3.03 points lower on the Reading NECAP than a White student, controlling for participation, grade, school, and other demographic factors. African American students' scores in Reading did not differ from those of White students at a statistically significant level for any of the years of this study.

Socioeconomic Status, as measured by qualification for Free or Reduced price lunch in this study, has a significant impact on student performance in Math and Reading. Estimates for Math and Reading achievement for all five years indicate lower scores at a statistically significant level, on average, for students living in poverty (FRL). The estimated negative impact on math scores ranged from -3.23 points in 2011 ($p < .01$) to -3.87 points in 2010 ($p < .001$). In reading, the estimates range from -2.69 points in 2009 ($p < .01$) to -3.88 points in 2013 ($p < .001$). To illustrate, a student qualifying for free or reduced price lunch in 2013 would be predicted to score 3.88 points lower on the Reading NECAP than a student that did not qualify, controlling for participation, grade, school, and other demographic factors.

Finally, some students are identified as having special learning needs which can impact their academic progress. In this study, some students are identified as English Language Learners (ELL) and others are identified as students with a disability, meaning they qualify for special education services (SPED). Students that have either or both of these special learning needs are predicted to score lower on the Math and Reading NECAP assessments than students without special learning considerations. These estimated differences are statistically significant for all years of this study. For English language learners, the estimated difference ranges from -5.28 points in 2011 ($p < .001$) to -7.33 points in 2009 ($p < .001$) for the math assessment, and -4.46 points in 2011 ($p < .01$) to -8.42 points ($p < .001$) for the reading assessment. For students with special needs, the estimated differences range from -5.24 points in 2011 ($p < .001$) to -9.73 points in 2013 ($p < .001$) on the reading assessment, and -7.28 points in 2011 ($p < .001$) to -10.12 points in 2013 ($p < .001$) on the math assessment. Table 32 delineates predicted scores for third grade participants and non-participants at School 3 for each demographic factor for each year of the study.

Table 32: Predicted Math and Reading NECAP scores for prototypical students, by demographic characteristic and participation (2009-2013)

2009

Math	NON-PARTICIPANT	PARTICIPANT	**Reading**	NON-PARTICIPANT	PARTICIPANT
FEMALE	41.51	40.24	FEMALE	46.61	45.31
HISPANIC	42.24	40.97	HISPANIC	42.29	40.99
AFRICAN AMERICAN	41.12	39.85	AFRICAN AMERICAN	45.36	44.06
NON-FREE LUNCH	47.32	46.05	NON-FREE LUNCH	47.67	46.37
ELL	34.91	33.64	ELL	33.87	32.57
SPED	37.22	35.95	SPED	36.03	34.73

2010

Math	NON-PARTICIPANT	PARTICIPANT	**Reading**	NON-PARTICIPANT	PARTICIPANT
FEMALE	41.60	41.94	FEMALE	44.54	44.59
HISPANIC	40.17	40.51	HISPANIC	40.43	40.48
AFRICAN AMERICAN	39.06	39.40	AFRICAN AMERICAN	42.08	42.13
NON-FREE LUNCH	47.26	47.60	NON-FREE LUNCH	46.91	46.96
ELL	34.28	34.62	ELL	33.81	33.86
SPED	36.75	37.09	SPED	35.69	35.74

2011

Math	NON-PARTICIPANT	PARTICIPANT	**Reading**	NON-PARTICIPANT	PARTICIPANT
FEMALE	43.45	41.77	FEMALE	47.11	45.62
HISPANIC	43.27	41.59	HISPANIC	44.63	43.14
AFRICAN AMERICAN	41.67	39.99	AFRICAN AMERICAN	45.36	43.87
NON-FREE LUNCH	47.27	45.59	NON-FREE LUNCH	49.87	48.38
ELL	37.73	36.05	ELL	40.17	38.68

SPED	38.80	37.12	SPED	38.83	37.34

2012

Math	NON-PARTICIPANT	PARTICIPANT	**Reading**	NON-PARTICIPANT	PARTICIPANT
FEMALE	40.77	41.99	FEMALE	44.32	45.08
HISPANIC	41.50	42.72	HISPANIC	42.31	43.07
AFRICAN AMERICAN	39.66	40.88	AFRICAN AMERICAN	42.38	43.14
NON-FREE LUNCH	46.85	48.07	NON-FREE LUNCH	47.29	48.05
ELL	35.07	36.29	ELL	34.75	35.51
SPED	36.51	37.73	SPED	36.55	37.31

2013

Math	NON-PARTICIPANT	PARTICIPANT	**Reading**	NON-PARTICIPANT	PARTICIPANT
FEMALE	39.72	40.28	FEMALE	43.09	42.30
HISPANIC	39.90	40.46	HISPANIC	40.34	39.55
AFRICAN AMERICAN	37.58	38.14	AFRICAN AMERICAN	39.96	39.17
NON-FREE LUNCH	44.94	45.50	NON-FREE LUNCH	45.10	44.31
ELL	34.62	35.18	ELL	35.62	34.83
SPED	31.58	32.14	SPED	31.10	30.31

Question One Summary

Question one asks if students who attend 21st Century after school programs perform better on Math and Reading assessments than non-participants. Based on the analyses presented above, the answer would be that they do not. While results vary year to year, the only statistically significant impacts of participation are negative. Results suggest that program participants score one to two points lower, on average, than non-participants, controlling for school, grade, and demographic factors. These results hold for Reading and Math assessments.

In this analysis, the variables included explain less than 20% of the variance (R^2) in achievement scores. For math achievement, the variables included (participation, grade, school, and demographics) explain between 11.3% (2011) and 18% (2013) of the variation in Math NECAP scores. For reading achievement, the variables explain between 16.3% (2011) and 19.3% (2013) of variation in Reading NECAP scores. Through the progression of model fitting, it is evident that the majority of the variance is explained by demographic factors. Adding gender, race, poverty status, and special learning factors to the model explained between 10 and 16% of the variance, on average. With over 80% of the variance in achievement scores left unexplained, it is important to explore other factors that may explain math and reading achievement. In question two, differential impacts of participation are considered.

<u>Question 2:</u> Do the effects of attending the 21st Century After School program in selected elementary schools in Nashua, New Hampshire differ based on student characteristics such as race, gender, free/reduced lunch status, status as a student with a disability or English language learner, or grade of attendance?

To answer question two, differential effects of program participation were considered based on school, demographic factors, and grade level. First, school effects were considered. I explored differences in achievement based on which school's after school program was attended. This was done by multiplying participation by school to get a new variable for each school (e.g., P1=participation*School 1). Next, interactions between participation and demographics were considered. This was calculated in the same manner. For example, the differential impact of participation on male students would be modeled by including a variable PARTMALE= MALE*participation. Finally, for each year, grade level of participation was considered. This was calculated by first creating dummy variables for grade of participation. Each new variable (Gr.3, for example) was multiplied by participation. This makes it possible to examine if participation in the after school program has a bigger impact during some grade levels than others. Table 33 outlines the progression of models that were fit to answer question two.

Table 33: Variables included in taxonomy of fitted regression models

Model 5	Program Effects (School*Participation)
Model 6	Demographic Interactions (MALE*participation), (HISP*participation), (AFAM*participation), (FRL*participation), (ELL*participation), (SPED*participation)

Model 7	Grade Level of Participation (2nd Grade*participation), (3rd Grade*participation), (4th Grade*participation), (5th Grade*Participation)
Model 8	Best fit model

Tables 34 and 35 display the results of the best fit regression models for each year of the data for math and reading. Following is an examination of the model residuals and a discussion of significant findings. The full taxonomy of fitted models can be found in Appendix D.

Table 34: Parameter estimates for predicted math achievement, controlling for participation in the 21st Century after school program, school, grade, and demographic factors

	$\widehat{MATHACH}09$	$\widehat{MATHACH}10$	$\widehat{MATHACH}11$	$\widehat{MATHACH}12$	$\widehat{MATHACH}13$
Intercept	46.63***	49.24***	46.98***	46.14***	42.63***
Participation	N/A	-.58	-.63	1.22	-1.79
School 1	1.50	2.98*	2.12	3.12~	4.82**
School 2	1.72	1.39	-.83	2.26~	2.56*
School 4	-1.64	.17	-.18	2.26~	3.54**
School 5	-.04	1.13	-.19	.64	1.03
Grade	-.49	-1.25***	-.23	-.61	.63~
Male	2.01**	1.81*	.78	2.54**	1.51~
Hispanic	-3.93**	-3.53***	-1.31	-1.81~	-3.75**
African Am	-2.36*	-4.47***	-2.66~	-3.65*	-3.53**
FRL	-3.75***	-3.86***	-2.93*	-3.54**	-3.68**
ELL	-6.85***	-5.68***	-5.35***	-6.43***	-5.17***
Special Ed.	-6.15***	-6.57***	-5.19***	-6.80***	-9.75***
GR2 PARTIC	3.71**	N/A	N/A	N/A	N/A
GR3 PARTIC	-4.53**	3.58**	N/A	N/A	N/A
GR4 PARTIC	-2.03	N/A	N/A	N/A	N/A
GR5 PARTIC	-3.87**	N/A	-4.63**	N/A	N/A
*P*SCHOOL1*	4.05*	N/A	N/A	N/A	N/A
*HISP*PART*	N/A	N/A	N/A	N/A	4.71**

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 35: Parameter estimates for predicted reading achievement, controlling for participation in the 21st Century after school program, school, grade, and demographic factors

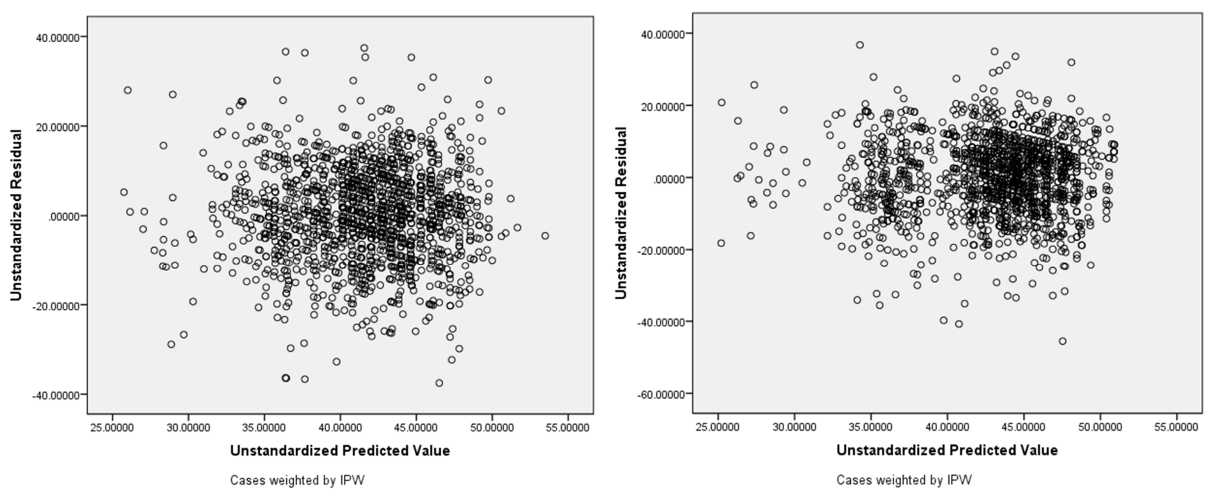
	$\widehat{READACH}09$	$\widehat{READACH}10$	$\widehat{READACH}11$	$\widehat{READACH}12$	$\widehat{READACH}13$
Intercept	54.52***	49.47***	53.31***	50.75***	47.33***
Participation	-.45	-1.34	-1.08	.04	-.79
School 1	.29	1.45	-.31	.74	4.81***
School 2	-.82	1.30	-3.39**	1.19	4.15***
School 4	-2.90**	.71	-2.09~	1.94	2.91*
School 5	-.75	1.80	-1.49	-.53	2.42~
Grade	-1.68***	-.41	-.69~	-1.06*	-.12
Male	-1.61*	-.95	-.77	-.17	-1.87*
Hispanic	-3.56*	-3.23**	-1.71~	-1.95~	-.88
African Am	.44	-1.46	-3.05	-1.92	-1.26
FRL	-2.61**	-3.44**	-3.71**	-2.94*	-3.88***
ELL	-8.59***	-6.70***	-3.90**	-7.48***	-4.72***
Special Ed.	-8.98***	-9.75***	-7.48***	-7.71***	-10.12***
GR2 PARTIC	N/A	N/A	-5.44**	N/A	N/A
GR3 PARTIC	-3.24*	2.81*	N/A	N/A	N/A

GR4 PARTIC	N/A	N/A	N/A	N/A	N/A
GR5 PARTIC	N/A	N/A	N/A	3.88	N/A
*SPED*PART*	N/A	3.99~	N/A	N/A	N/A
*AFAM*PART*	N/A	N/A	4.68~	N/A	N/A

~ $p <$ .10, * $p <$ .05, ** $p <$ .01, *** $p <$ .001

Examining the residuals for the final models, there are no signs of homoscedasticity or non-linearity. Residuals appear to be randomly distributed. The models meet the assumptions for linear statistical models. Figures 31 and 32 contain the scatterplots of unstandardized residuals for 2012 for Math and Reading to illustrate the random distribution of residuals. Residual plots for the other eight models can be found in Appendix E.

Figures 31&32: Bivariate scatterplots of unstandardized residuals versus unstandardized predicted values for Math (31) and Reading (32) for 2012 *(n* = 1499)



In fitting final, best fit models for each year of data for reading and math, statistically significant interactions with participation were found for math achievement for each year of the data with the exception of 2012, and for reading achievement for each year except 2013. For these two years, the best fit model was the model presented in question one.

School Effects

The first set of interactions examined was the differential impact of attending the after school program at a particular school. Were some schools' after school programs associated with higher levels of student achievement in reading or math than others? In fitting the models with the school interaction predictor, one statistically significant effect was found. For school year 2008-2009, the estimated effect of attending the after school program at School 1 on math achievement was statistically significant at the $p < .05$ level. The effect of participating in the program for these students is estimated to be +4.05 points on the Math NECAP, controlling for school, grade, demographic factors, and grade level of participation. Prior to including the interaction in the model, the estimated impact of attending School 1 was 3.68 points. Upon inclusion of the interaction variable for School 1 after school participation, the school effect dropped to 1.50 points, suggesting that some of the higher achievement in math of students at School 1 is actually attributable to student participation in the after school program.

This impact of participation by school is found in Table 36. Non-participating students at School 1 are predicted to score .22 points lower on the math assessment than students at School 2, and at the other end, 3.14 points higher than students at School 4. However, after school participants at School 1 are predicted to score higher than participants at any of the other schools, with the largest estimated differential being 7.19 points higher than School 4 participants.

Table 36: Predicted Math NECAP scores (2009) for prototypical students, by school and participation (*n* = 1559)

SCHOOL	NON-PARTICIPANTS	PARTICIPANTS
1	44.92	44.44
2	45.14	40.61
3	43.42	38.89
4	41.78	37.25
5	43.02	38.85

Demographic Interactions

Next, interactions between participation and key demographic factors were included in each model. In looking at interaction effects on math achievement, a statistically significant impact ($p < .01$) is found for Hispanic students who participate in the after school program for the 2012-2013 school year. This model estimates that Hispanic students who participated in the program score 4.71 points higher on the Math NECAP than Hispanic students who did not participate, controlling for school, grade, and demographic factors. While Hispanic students, on average, are predicted to score lower on the Math NECAP than White students, participation in the after school program is associated with predicted math scores that exceed the predicted scores of White students participating in the program by almost one point (.96). While participants in the program, on average, have lower predicted math achievement than non-participants, Hispanic participants are estimated to have achievement close to that of White, non-participants, coming within one point (.83). Table 37 presents a comparison of predicted achievement levels in math for 2013 for various demographic groups.

Table 37: Predicted Math NECAP scores (2013) for prototypical students, by demographic factors and participation (*n* = 1316)

	NON-PARTICIPANT	PARTICIPANT
WHITE	42.35	40.56
FEMALE	40.84	39.05
HISPANIC	38.6	41.52
AFRICAN AMERICAN	38.82	37.03
NON-FREE LUNCH	46.03	44.24
ELL	33.43	36.35
SPED	32.60	30.81

For reading achievement, two demographic interactions were found. In 2010, students identified with a disability who participated in the after school program were predicted to score 3.99 points higher, on average, than students with disabilities who did not participate ($p < .1$), controlling for participation, school, grade, demographic factors, and third grade participation.

While students with disabilities are predicted to score lower on the Reading NECAP, on average, than their non-disabled peers, participation in the after school program closed this gap by 2.65 points, on average. Table 38 highlights these estimated impacts on reading achievement for the 2009-2010 school year.

Table 38: Predicted Reading NECAP scores (2010) for prototypical students, by demographic factors and participation (*n* = 1876)

	NON-PARTICIPANT	PARTICIPANT
WHITE	43.85	45.32
FEMALE	44.80	46.27
HISPANIC	40.62	42.09
AFRICAN AMERICAN	42.39	43.86
NON-FREE LUNCH	47.29	48.76
ELL	33.92	32.58
SPED	34.1	39.56

A demographic interaction was also found in the model for school year 2010-2011. African American students who participated in the after school program were predicted, on average, to score 4.68 points higher on the Reading NECAP than African American students who did not participate ($p < .10$), controlling for school, grade, demographic factors and grade 2 participation. While African American students are predicted to score lower on the Reading NECAP, on average, than White students (43.71 points vs. 46.76 points), participation in the after school program is associated with higher predicted scores than White students. In the final model, African American students that participate in the after school program are estimated to outscore White participants by 1.63 points, on average, and White non-participants by .55 points, on average. Table 39 contains predicted Reading NECAP scores for various demographic groups for 2010-2011.

Table 39: Predicted Reading NECAP scores (2011) for prototypical students, by demographic factors and participation (*n* = 1627)

	NON-PARTICIPANT	PARTICIPANT
WHITE	46.76	45.68
FEMALE	47.56	46.48
HISPANIC	45.05	43.97
AFRICAN AMERICAN	43.71	47.31

NON-FREE LUNCH	50.47	49.39
ELL	41.15	40.07
SPED	39.28	38.20

Grade level interactions

The final interaction with participation that was considered was the differential impact of the grade level in which participation occurs. For example, is there a greater impact of after school program participation in grade 3 than grade 5? Several statistically significant interaction effects were found with grade level of participation for both Math and Reading achievement. Fourth grade participation was not found to have a significant impact on math or reading achievement during any year of this study. Participation during the second grade year had mixed results. Based on 2009 data, students who participated in the after school program during grade two were predicted to score 3.71 points higher on the Math NECAP ($p < .01$), on average, than students who did not participate in grade two, controlling for school, grade, demographic features, School 1 participation, and participation at other grade levels. However, grade two participation predicted a negative association with reading achievement using data from the 2011 school year. Participating students were predicted to score 5.44 points lower on the Reading NECAP ($p < .01$), on average, than non-participants, controlling for school, grade, and demographic features.

Grade three participation similarly was associated with mixed results. With regards to math achievement, grade three participation was associated with higher Math NECAP scores (+3.58 points, $p < .01$) in 2010, but lower Math NECAP scores (-4.53 points, $p < .01$) in 2009. Mixed results were also found for Reading scores. Higher results were predicted by grade three participation in 2010 (+2.81 points, $p < .05$), but lower scores in 2009 (-3.24 points, $p < .05$).

Participation in the 21st Century after school program in grade five had a statistically significant negative association with Math NECAP scores in the final models for 2009 and 2011 data. Based on 2009 data, a student who participated in the after school program in grade five would be predicted to score 3.87 points lower, on average, on the Math NECAP than a fifth grade student that did not participate, controlling for school, demographic features, and School 1 participation. Similarly, based on 2011 data, a participating grade 5 student would be predicted to score 4.63 points lower, on average, on the Math NECAP than a non-participating fifth grader, controlling for participation, school, grade, and demographic factors. In contrast, based on 2012 data, students that participated in the program during their grade five year were predicted to have higher scores on the Reading NECAP than their non-participating peers (+3.88 points, $p < .05$), controlling for participation, school, grade, and demographic factors. Table 40 presents predicted scores for Reading and Math achievement based on grade level of participation for the years and subjects in which there was an applicable interaction effect.

Table 40: Predicted NECAP scores for prototypical students, by grade and participation

2009 Math			**2010 Math**		
	NON-PARTICIPANTS	PARTICIPANTS		NON-PARTICIPANTS	PARTICIPANTS
GR.2	43.91	47.62	GR.2	44.62	44.96
GR.3	43.42	38.89	GR.3	43.39	43.73
GR.4	42.93	40.90	GR.4	42.16	42.50
GR.5	42.44	38.57	GR.5	40.93	41.27

2011 Math			**2009 Reading**		
	NON-PARTICIPANTS	PARTICIPANTS		NON-PARTICIPANTS	PARTICIPANTS
GR.2	44.37	43.74	GR.2	46.94	46.49
GR.3	44.14	43.51	GR.3	45.26	41.57
GR.4	43.91	43.28	GR.4	43.58	43.13
GR.5	43.68	38.42	GR.5	41.90	41.45

2010 Reading			**2011 Reading**		
	NON-PARTICIPANTS	PARTICIPANTS		NON-PARTICIPANTS	PARTICIPANTS
GR.2	44.26	42.92	GR.2	47.45	40.93
GR.3	43.85	45.32	GR.3	46.76	45.68
GR.4	43.44	42.10	GR.4	46.07	44.99

GR.5	43.03	41.69	GR.5	45.38	44.30

2012 Reading

	NON-PARTICIPANTS	PARTICIPANTS
GR.2	45.52	45.56
GR.3	44.46	44.50
GR.4	43.40	43.80
GR.5	42.34	46.26

Question Two Summary

Question two explores differential effects of participation in the after school program on Math and Reading scores. Differences are considered by school, grade level and student characteristics. The results described above indicate that some differential effects of participation exist, however, results are not consistent from year to year. A differential effect of participation on Math scores was found for students attending School 1 in 2009. However, no other differential effects were found for school of participation for any other year or subject.

Differential effects of grade level of participation were mixed across years and subject areas. Statistically significant positive effects were found on math achievement for participation in grades two and three (2009, 2010), but negative effects were found for participation in grades three, four, and five (2009, 2011). Similarly mixed results were found for reading achievement, with positive impacts for participation during grades three and five some years (2010, 2012), and negative impacts for participation in grades two and three in other years (2009, 2011).

Positive differential effects of participation were found for certain groups in certain years for one subject area. Hispanic students who participated in the after school program scored significantly higher in math than Hispanic students who did not participate (2013). Similarly, participating African American students outscored their non-participating peers in Reading (2011), and students in special education programs who participated in the after school program scored higher than their non-participating peers (2010). While these findings indicate that after school program participation may impact subgroups of students differently, the inconsistency of the findings make it difficult to generalize results.

Examining the R^2statistics for these ten models, more of the variation in Math and Reading NECAP scores is explained by the models answering question two than those presented in question one, for most years. In the models fit to look at the differential impacts of participation on math achievement, the model fit for 2009 explains the most variation in scores at 20.1%, while the model for 2011 explains the least amount of variation at 12.5%. The addition of interaction variables for participation improved the fit of four of the five models for math, with the biggest improvement in the 2009 model (4% increase in R^2). Results were similar for reading achievement scores with the 2009 model explaining the most variation at 19.8% and the 2012 model explaining the least at 17.6%. Improvement in the R^2 statistic was more modest for reading scores with the greatest growth found in 2011 at 1.9 percentage points.

While modest improvements in the fit of our models is encouraging, there is still a large amount of unexplained variation in Math and Reading NECAP scores. Considering the impact of different levels of program participation, or dosage, is explored next to see if this improves the fit of the statistical models.

<u>Question 3:</u> Do the effects of attending the 21st Century after school program differ based on student attendance in the program within an academic year (# of days attended)?

To answer question three, the same ten data sets (math and reading for each of the five years of the study) were used to examine effects of program dosage on math and reading achievement. The data sets were weighted by the Inverse Propensity Score of each student, to achieve the balanced data described in the beginning of this chapter. Taxonomies of fitted models are presented that examine first the effect of the number of days of attendance in the after school program within a school year, and then any differential impact of days of attendance for different types of students. These analyses explore whether or not a high dosage of attendance has a greater benefit for certain types of students than others or for students at a certain grade level.

Table 41 highlights the model progression used in fitting the taxonomies of models. Participation and dosage were added first, as they are the key variables to answering question three. Next, control variables were added. Following this, interaction variables between days of attendance and school, demographics, and grade level were added, in that order. Interaction effects for all three of these categories were found with participation, and they were added in the same progression in the model building for question three. Finally, interaction effects were considered between school effects and key demographic variables – poverty status and status as a special learner (ELL or Special education). These are groups that have been targeted for academic intervention by the 21st Century program for academic improvement (U.S. DOE, 2003). These interaction variables examine if students in these categories have better academic results from attending the after school program at some schools than others.

Table 41: Variables included in taxonomy of fitted regression models

Model 1	Participation & days of attendance
Model 2	Demographic controls: School, Grade, Gender, Race, FRL/Poverty status, ELL, Special Education
Model 3	School Effects: School*Days (School 1, 2, 4, 5)
Model 4	Demographic Interactions: MALE*DAYS, HISP*DAYS, AFAM*DAYS, ELL*DAYS, SPED*DAYS, FRL*DAYS
Model 5	Grade Interaction: Grade of attendance*Days
Model 6	School effects*FRL: School*Days*FRL
Model 7	School effects*ELL: School*Days*ELL
Model 8	School effects*SpEd: School*Days*SpEd
Model 9	Best Fit Model

Fitting this progression of models for each data set has yielded 10 best fit models, presented in Tables 42 and 43. A detailed analysis follows. The full taxonomy for each model is presented in Appendix F.

Table 42: Parameter estimates for predicted math achievement, controlling for dosage of participation in the 21st Century after school program, school, grade, and demographic factors

	$\widehat{MATHACH}09$	$\widehat{MATHACH}10$	$\widehat{MATHACH}11$	$\widehat{MATHACH}12$	$\widehat{MATHACH}13$
Intercept	49.33***	48.80***	46.16***	45.77***	42.79***
Participation	-2.04	.31	-1.22	-1.59	-2.90*
Days	-.01	.00	.04*	.02	.01
Grade	-1.17**	-1.25***	-.41	-.49	.62~
School 1	1.22	2.81*	1.94	3.27*	4.05*
School 2	1.91~	1.22	-.95	2.74*	2.43*
School 4	-1.41	2.00	1.37	2.65*	3.70**
School 5	-.03	.88	-.20	1.03	.82
FRL	-4.32***	-3.74***	-2.72*	-3.36**	-3.86***
African Am	-2.96**	-4.36***	-2.24	-3.70*	-3.70**
Hispanic	-4.36**	-3.55***	-1.43	-2.08~	-3.78**
Male	2.12**	1.86*	2.75*	2.70**	1.35~
ELL	-7.50***	-5.87***	-5.01***	-8.05***	-4.93***
Special Ed.	-6.58***	-6.65***	-5.56***	-7.52***	-8.66***
Gr2 Days	.06**	N/A	N/A	N/A	N/A
Gr3 Days	-.03~	.03*	N/A	N/A	N/A
Gr5 Days	N/A	N/A	-.04*	N/A	N/A
*FRL Days*Sch1*	.07**	N/A	N/A	N/A	.04~
*Days*Sch4*	N/A	-.05**	-.04*	N/A	N/A
*Male*Days*	N/A	N/A	-.05**	N/A	N/A
*ELL*Days*	N/A	N/A	N/A	.04~	N/A
*Hisp*Days*	N/A	N/A	N/A	N/A	.04**

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 43: Parameter estimates for predicted reading achievement, controlling for dosage of participation in the 21st Century after school program, school, grade, and demographic factors

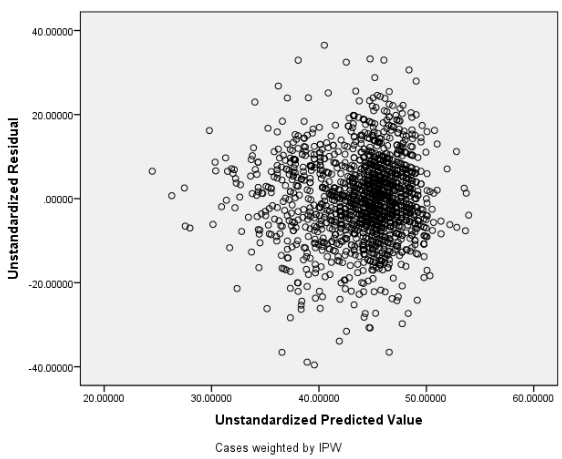
	$\widehat{READACH}09$	$\widehat{READACH}10$	$\widehat{READACH}11$	$\widehat{READACH}12$	$\widehat{READACH}13$
Intercept	51.12***	49.30***	50.06***	51.17***	47.23***
Participation	-2.64*	-.38	-1.73	.32	-4.01**
Days	.00	.00	.05**	.01	.03**
Grade	-.76~	-.34	-.56	-1.12**	-.06
School 1	.66	.31	-.34	.90	4.90***
School 2	-.58	1.51	-2.50~	.63	3.48**
School 4	-2.69*	1.30	-.73	2.80~	3.15**
School 5	-.90	2.34~	.28	-1.53	2.29~
FRL	-2.65**	-4.11***	-2.34*	-2.69*	-3.72***
African Am	.39	-1.86	-.78	-1.34	-1.31
Hispanic	-3.75**	-3.06**	-1.57~	-2.41*	-1.12
Male	-1.54*	-1.20	.84	-.55	-1.91*
ELL	-8.17**	-6.14***	-3.81**	-7.29***	-5.35***
Special Ed.	-8.99***	-6.81***	-7.70***	-8.55***	-9.71***
Gr2 Days	.07***	N/A	-.05**	N/A	N/A
Gr3 Days	N/A	.03*	N/A	N/A	N/A
Gr5 Days	N/A	N/A	N/A	.04*	N/A
*FRL Days*Sch1*	N/A	.04~	N/A	N/A	N/A
*FRL Days*Sch2*	N/A	N/A	-.03~	N/A	N/A
*FRL Days*Sch4*	N/A	N/A	-.05*	-.05*	N/A
*FRL Days*Sch5*	N/A	N/A	-.05*	N/A	N/A
*Sped Days*Sch4*	N/A	-.06*	N/A	.13**	N/A

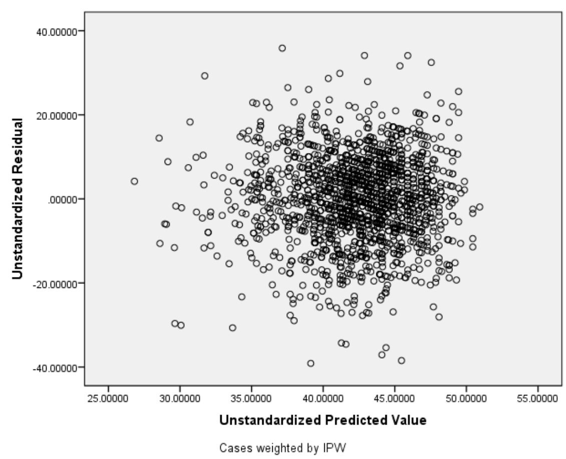
*Sped Days*Sch5*	N/A	-.08*	N/A	N/A	N/A
*ELL Days*Sch2*	N/A	N/A	N/A	N/A	.05*
*Male*Days*	N/A	N/A	-.03*	N/A	N/A

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Residuals for all ten models displayed no signs of non-linearity or homoscedasticity. The Shapiro-Wilk test of normality was not significant at the $p < .05$ level for any of the final models, indicating that the residuals are normally distributed. Residuals were also examined by key predictors (school, grade, participation, and FRL status), and all were normally distributed. Residual plots for 2011 are illustrated below in Figures 33 and 34. The plots for the other four years are included in Appendix G.

Figures 33 & 34: Raw residuals from final fitted regression models (2011) in which math (33) and reading (34) achievement is predicted by participation in the 21st Century after school program, controlling for demographic characteristics (*n* = 1627)





Sensitivity tests were conducted for each model to look for any atypical data points. Analyses were conducted that examined Cook's D, Press Residuals, and HAT statistics. These analyses did result in some adjustments being made to final models. In the 2009 Reading data, School 4 Days*SpEd and School 1 Days*FRL had statistically significant effects. However, influence statistics indicated that these effects were each unduly influenced by one student. Therefore, these variables were not included in the final model. In the 2011 Math data, School 2

Days*SpEd had statistically significant effects. However, when examining influence statistics, it was influenced by one student's score, and was therefore excluded from the final model. Similarly, in the 2013 Math and Reading data, School 1 Days*SpEd was statistically significant, however, upon further examination was influenced by one student and was not included in the final model.

Dosage

When considering a student as a participant in the program, he or she could have attended the program anywhere between 30 and 180 days during that school year. Including the number of days a student attended in the regression model considers whether or not the number of days attending the program is associated with a difference in math or reading achievement. In seven out of ten of the final models, no statistically significant impact was found for days of attendance. Days of attendance was found to be statistically significant ($p < .05$) using the 2012 math data. Days of participation had an estimated effect of .03 points per day. This is small, and must also be considered in conjunction with other parameter estimates. This model estimates that participants, on average, scored 1.50 points lower than non-participants. Therefore, we would predict that a student that attended the program more than 50 days would demonstrate higher levels of math achievement than a non-participant, controlling for school, grade, and demographic factors. The average participating student in 2012 attended the program 99.5 days. Based on this model, we would predict that an average participating student would score approximately 1.49 points higher on the Math NECAP than a non-participating student.

Similar results are found with the 2013 data for math and reading. The math data indicated a statistically significant effect of days ($p < .01$), with an estimated effect of .04 points/day. Participants in this model, however, were predicted to score 3.33 points lower than non-participants. Therefore, it would be predicted that a student attending the program more than 83 days would have higher math scores than a non-participating student. The average participating student in 2013 attended the program 101.4 days and would be predicted to score approximately .73 points higher, on average, than a non-participating student, controlling for school, grade, and demographic factors. The same estimated effect of days was found for reading achievement in 2013 (+.04, $p < .01$). However, the parameter estimated for participation was lower at -4.07 points. Therefore, a student would have to attend the program at least 102 days to have a predicted Reading NECAP score higher than a non-participant. While 102 days is slightly higher than the average participant attended, it is not outside of the range of participation for 2013.

Figures 35 and 36 below illustrate predicted achievement scores for participants and non-participants for 2013 as a function of the number of days attending. As you can see, participants with low levels of participation score lower than non-participants, on average. However, at higher dosages, participants are predicted to have higher achievement scores, on average, than their non-participant peers.

Figure 35: Fitted regression lines illustrating the relationship between level of participation in 21st Century programming and performance on the Math NECAP exam (2013)

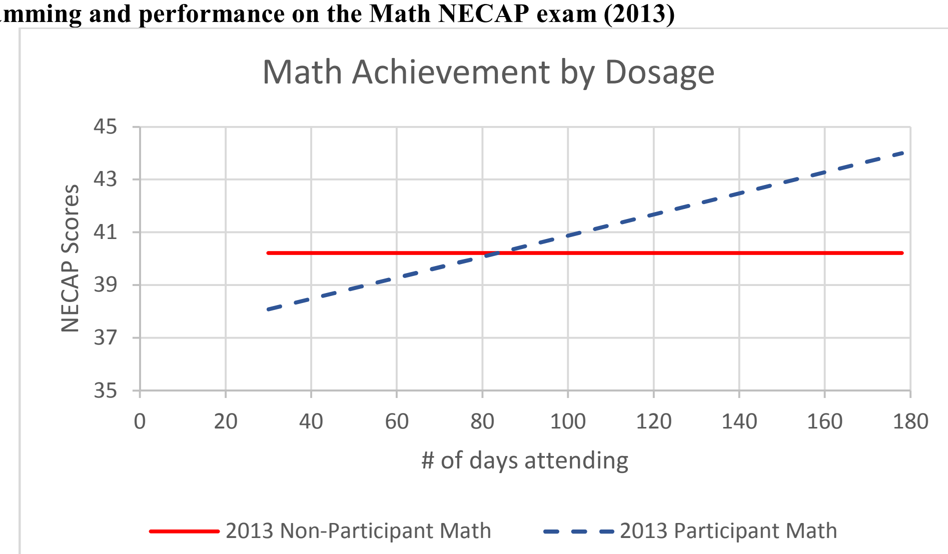
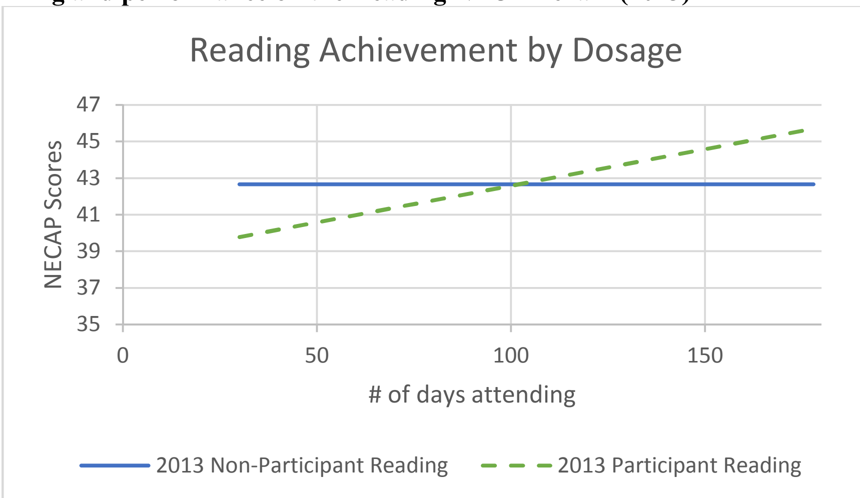


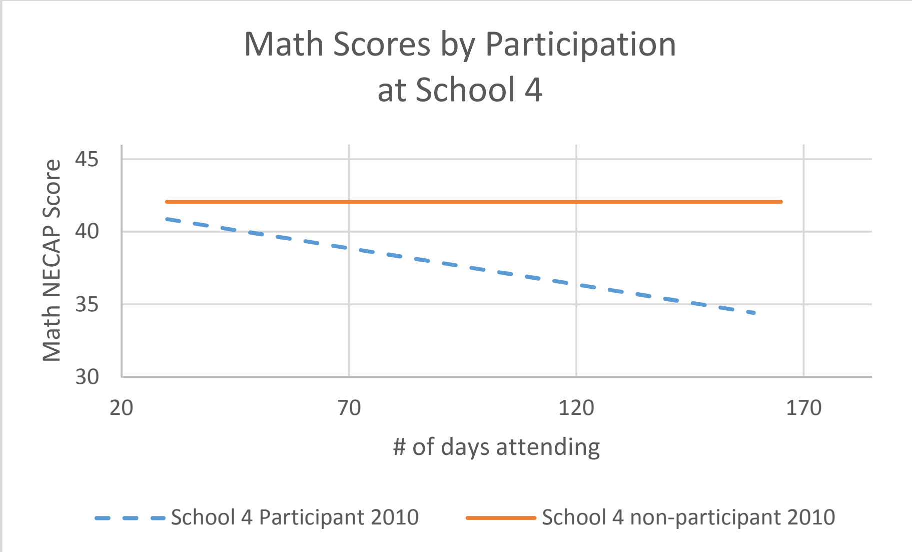
Figure 36: Fitted regression lines illustrating the relationship between level of participation in 21st Century programming and performance on the Reading NECAP exam (2013)



School Effects

In the next part of the analysis, interaction effects are considered between dosage and other key variables, such as grade, school and demographic factors. First, the impact of school of participation was considered. The number of days spent in after school programming was considered by school. Do some after school programs predict higher achievement scores than others? While no statistically significant impacts were found for program dosage on reading scores, statistically significant negative impacts were found on math achievement scores for participants at School 4 for 2010 and 2011. Participants at School 4 were predicted to score .05 points lower per day of attendance ($p < .01$) than their peers, controlling for grade, school, grade 3 dosage, and other demographic factors (2010), and .04 points lower/day ($p < .01$) than their peers in 2011, controlling for grade, school, male days, grade 5 dosage, and demographic factors. Figure 37 illustrates predicted achievement in math for participants and non-participants at School 4 for 2010, controlling for other model variables.

Figure 37: Fitted regression lines illustrating the relationship between level of participation in 21st Century programming at School 4 and performance on the Math NECAP exam (2010)



Next, models were fit including interaction variables between site dosage and key demographic features, including Free/reduced lunch status, designation as an English Language learner, and status as a student with special needs. First, interactions were considered between site effects and free/reduced lunch status (School*Days*FRL). Statistically significant interactions were found for each year of the data, as illustrated in Table 44.

Table 44: Estimated effect of after school participation for students qualifying for free or reduced price lunch by school, subject, and year

2009	Math	School 1	+.07 points/day ($p <$ .001)
2010	Reading	School 1	+.04 points/day ($p <$ .10)
2011	Reading	School 2	-.03 points/day ($p <$ .10)
2011	Reading	School 4	-.05 points/day ($p <$ .05)
2011	Reading	School 5	-.05 points/day ($p <$ .05)
2012	Reading	School 4	-.05 points/day ($p <$ .05)
2013	Math	School 1	+.04 points/day ($p <$ .10)

Figure 38 highlights the positive impact of days of attendance in the after school program at School 1 on math achievement for students qualifying for Free or Reduced Price Lunch, compared with comparable students participating at other schools. Figure 39 illustrates the positive impact of days of attendance in School 1's after school program on reading achievement for students qualifying for Free or Reduced Price Lunch, compared with comparable participants at other schools. As a counterpoint, Figure 39 also displays the negative impact of more days of attendance in the after school program at School 4 on reading achievement for students qualifying for Free or Reduced Price Lunch, compared with comparable peers participating at other schools.

Figure 38: Fitted regression lines illustrating the relationship between level of participation in 21st Century programming, by site and performance on the NECAP Math exam for students qualifying for free or reduced price lunch (2009)

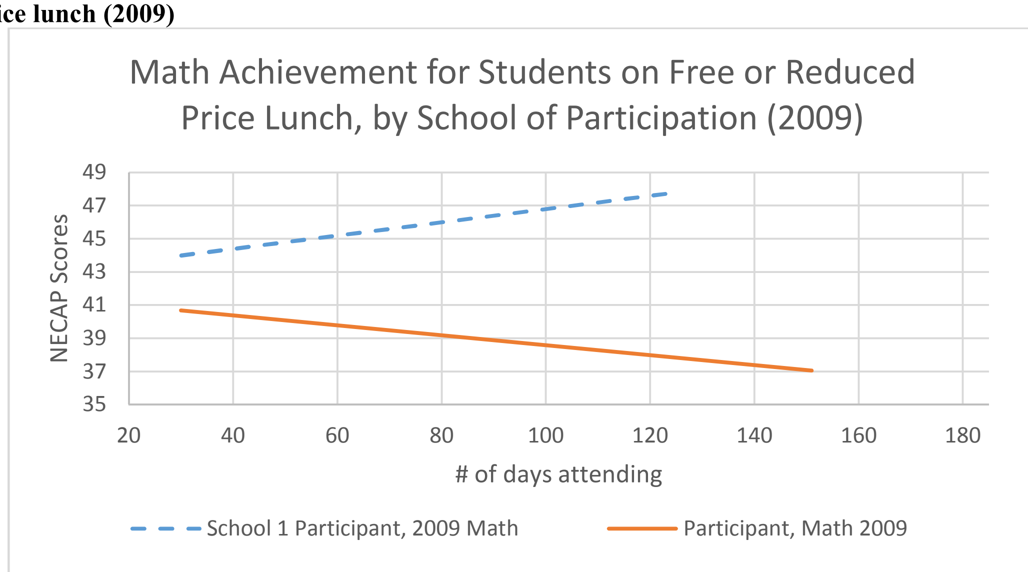
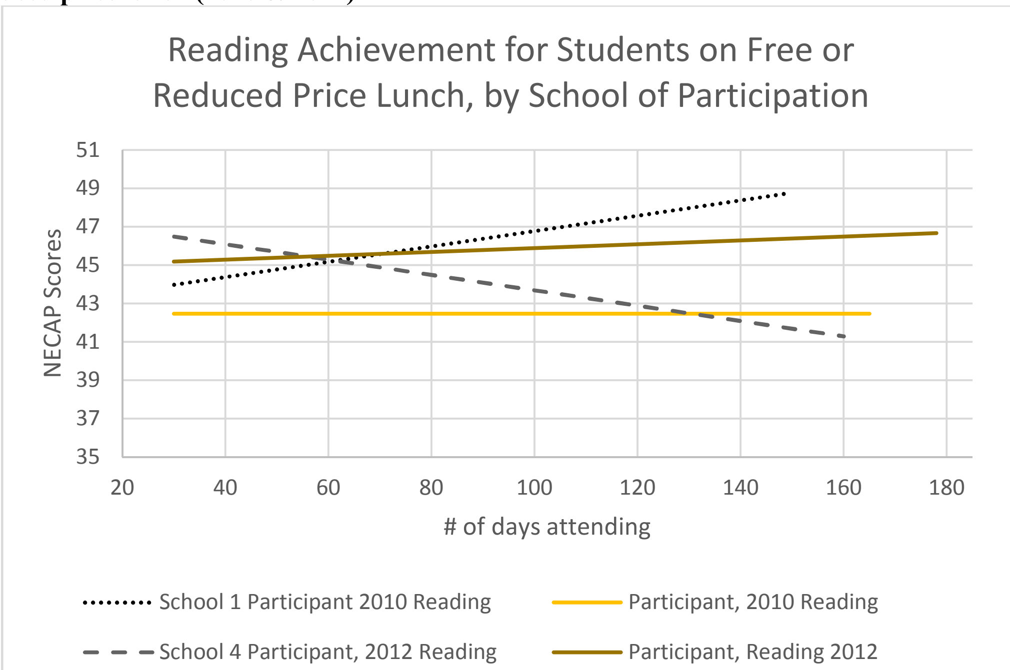


Figure 39: Fitted regression lines illustrating the relationship between level of participation in 21st Century programming, by site and performance on the NECAP Reading exam for students qualifying for free or reduced price lunch (2010 & 2012)



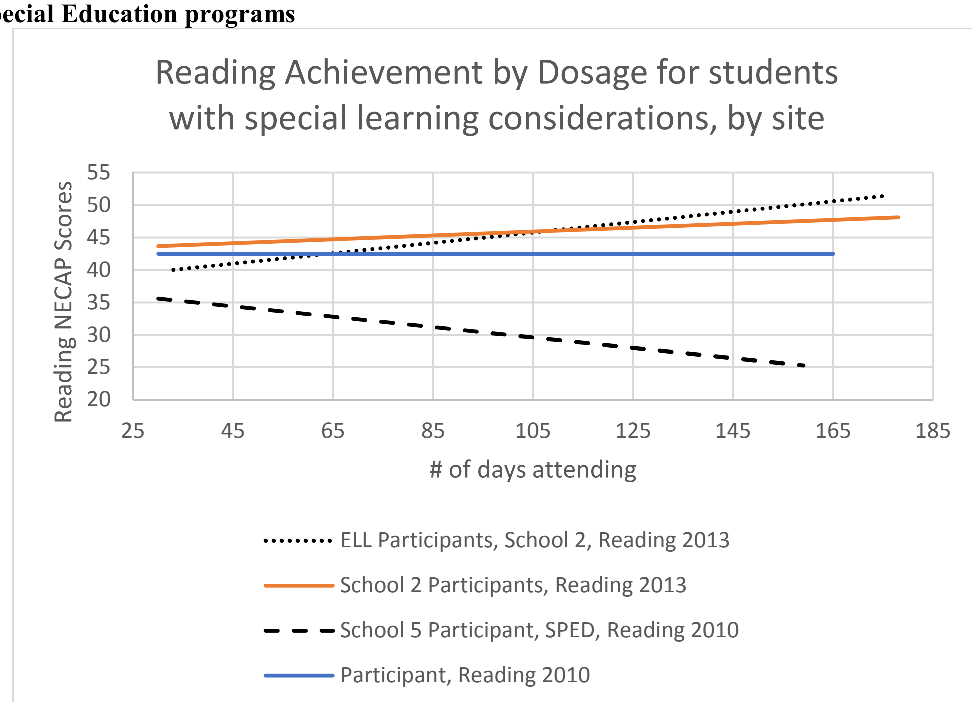
These interaction effects indicate that there are differential effects for students living in poverty the more days they attend the program at certain schools. Students experiencing poverty that attend the after school program at School 1 are predicted to outscore their peers that do not attend the program, as well as their peers that attend the program at other schools in Math (2009&2013), and in Reading (2010). In all three of these years, students with high levels of participation in the program (106, 135, 112 days, or more, respectively) would be predicted to score as high or higher than their peers who do not qualify for free or reduced price lunch, thus closing the achievement gap for these students.

Students qualifying for free or reduced price lunch did not display this differential level of achievement at all schools. In fact, students at School 2, School 4, and School 5 were predicted to score lower on the Reading NECAP than their peers at the same schools who didn't qualify for free or reduced price lunch, holding participation level constant, and lower than their qualifying peers at School 1 or School 3 in 2011. This remained true the following year for participating students experiencing poverty at School 4 with a differential negative impact of participation on their reading achievement scores, holding other factors constant.

Next, interaction variables were added that tested for differential effects of program participation for students with special learning designations, including English Language Learners (ELL) and students with special needs. One interaction effect was statistically significant for ELL students. A differential positive impact on reading achievement was found for participating students at School 2 (+.05, $p < .05$) in 2013, controlling for school, grade, and other demographic factors. Using this model, participating ELL students at School 2 would outscore their non-participating ELL peers, given 50 or more days attending the program, and would outscore their non-participating, non-ELL peers after 117 days of program participation, thus closing the often wide achievement gap between students designated as ELL and their peers.

Mixed results were found for differential impact of program participation for students identified with special needs. Statistically significant negative impacts on reading achievement were found for participants at School 4 (-.06, $p < .05$) and School 5 (-.08, $p < .05$) in 2010, controlling for grade, school, grade 3 dosage, School 1 FRL dosage, and other demographic factors. Participating students with special needs at these schools would be predicted to score lower than their non-participating peers, and lower than other students with special needs participating in the program at the other three schools, controlling for participation level. Positive results were found for reading achievement for students with special needs participating at School 4 in 2012 (+.13, $p < .01$), controlling for grade, school, grade 5 participation, School 4 FRL dosage, and other demographic features). Students with special needs attending the program at School 4 would be predicted to outscore their non-disabled peers on the reading assessment after 58 days of program participation, holding other factors constant. Figure 40, below, illustrates the differential effects of program participation in Schools 2 and 5 for students with special learning status.

Figure 40: **Fitted regression lines illustrating the relationship between level of participation in 21st Century programming, by site and performance on the Reading NECAP exam for students qualifying for ELL or Special Education programs**

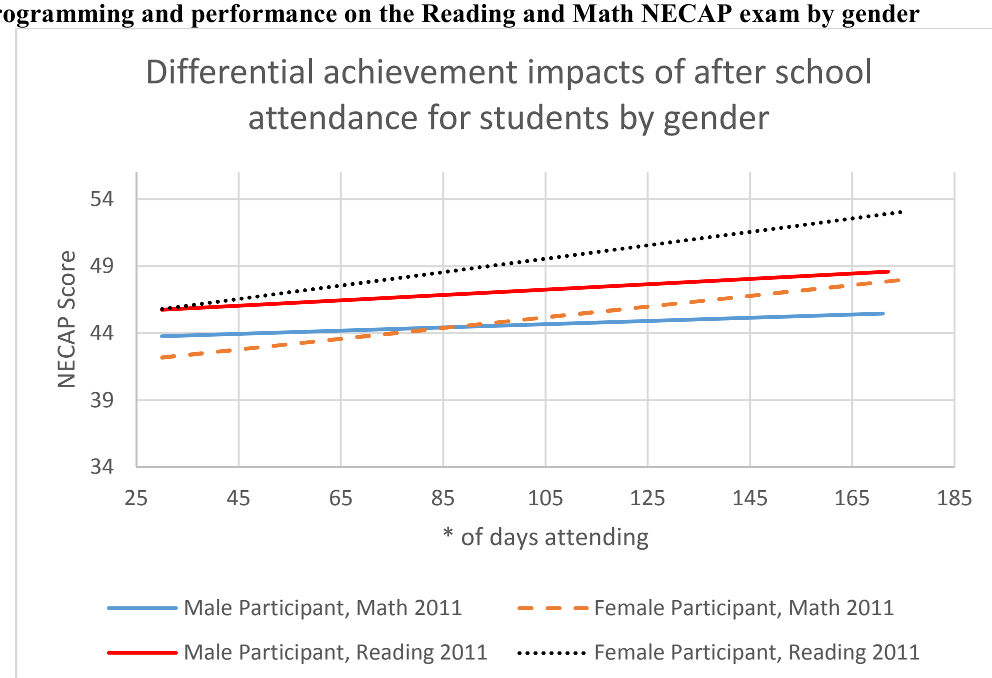


Demographic Interactions

Next, differential impacts of attending the after school program were considered for students of different demographic categories, including gender, race, poverty status, and status as a special learner. Statistically significant negative impacts were found for males in both math and reading in 2011. A male student attending the program 100 days would be predicted to score five points lower on the math assessment ($p < .01$) and three points lower than female participants on the reading assessment ($p < .05$), controlling for other model variables.

Positive differential impacts of participation were found on math achievement for Hispanic students (2013) and students participating in the ELL program (2012). ELL participants attending the after school program for 27 or more days would be predicted to outscore their non-participating peers, and participants attending more than 161 days would be predicted to outscore their non-ELL peers. Similar results were found for Hispanic students, with participants attending 134 or more days predicted to outscore their White peers in math, controlling for other factors. Figure 41 highlights the differential impacts of program attendance on males and females' math and reading achievement.

Figure 41: Fitted regression lines illustrating the relationship between level of participation in 21st Century programming and performance on the Reading and Math NECAP exam by gender



Grade Level Interactions

The final interaction effect considered in this analysis is grade level of participation. Included in the models were interaction variables which estimate the differential impact of attending the after school program at higher levels in certain grade levels. Are high levels of participation associated with higher or lower achievement in math or reading at some grade levels than in others?

Dosage of grade two participation had mixed results, with statistically significant positive effects on Math (+.06 points/day, $p < .01$) and Reading (+.07 points/day, $p < .001$) achievement in 2009. In both cases, this stands in sharp contrast to dosage in other grade levels which had no effect, or a negative effect on achievement, controlling for other model variables. In contrast, higher attendance in grade two predicted lower achievement scores in reading in the 2011 data (-.05 points/day, $p < .01$).

Contrasting results were also observed for grade three dosage. In the 2009 data, grade three participants scored .03 points lower on the Math NECAP, on average, for every additional day of program participation ($p < .1$), controlling for school, grade, grade two dosage, School 1 FRL dosage, and other demographic factors. Positive associations were found, however, for grade three dosage in 2010 for math and reading achievement. Participating grade three students scored +.03 points higher on the Math and Reading NECAP, on average, for each additional day of program participation ($p < .05$), controlling for other model variables.

No statistically significant interaction effects were found for grade four dosage for any year of this study. Grade five dosage, however, was associated with lower math scores in 2011 and higher reading scores in 2012. Based on 2011 data, grade 4 students participating 100 days would be predicted to score 4.41 points higher on the Math NECAP than grade five students participating 100 days, controlling for school, site effects (School 4), male dosage, and other demographic factors. Using the 2012 data, a grade five student participating 100 days would be predicted to score 2.88 points higher, on average, on the Reading NECAP than grade four students participating 100 days, controlling for school, free reduced lunch and special education dosage (School 4) and other demographic factors. Figure 42 visually illustrates the differential negative impact of grade five dosage for math in 2011. Figure 43 highlights the differential positive impact of grade five dosage on reading achievement in 2012.

Figure 42: Fitted regression lines illustrating the relationship between level of participation in 21st Century programming in grades 4 & 5 and performance on the Math NECAP exam

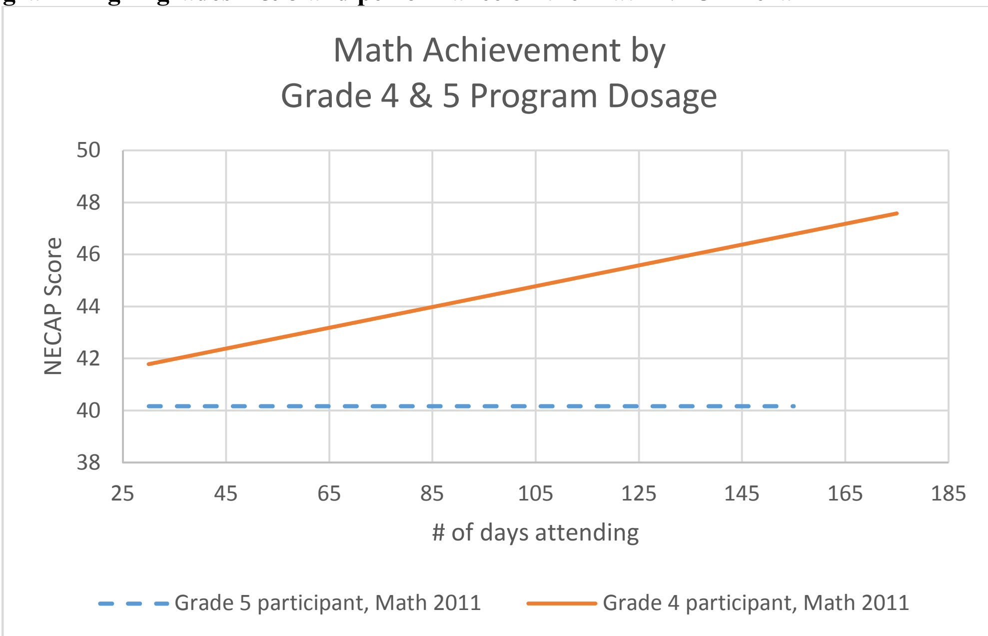
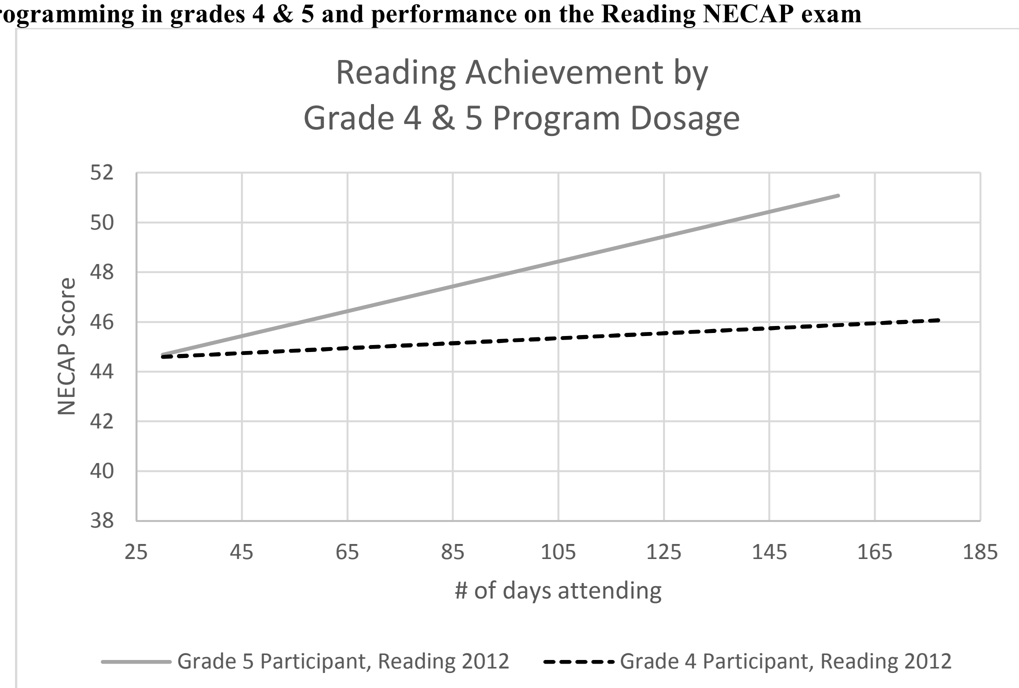


Figure 43: Fitted regression lines illustrating the relationship between level of participation in 21st Century programming in grades 4 & 5 and performance on the Reading NECAP exam



Question three summary

Question three explores the differential effects of dosage on student achievement in reading and math. The results described above indicate that in some years, there is a positive impact of increased days spent in the after school program, and that some years there is no impact, or a negative (not statistically significant) impact. Positive effects of days of attendance were found for math achievement in 2011 and reading achievement in 2011 and 2013. Due to the negative effect of participation in those years, students that participate at low dosages (e.g., less than 50 days) have lower achievement scores than their peers, while high dosage participants score higher than their peers in math and reading.

Different effects of dosage were found for some schools, demographic groups and grade levels, as described above. Some of these results were promising and hint at the possibility of after school programming being able to close the achievement gap for students living in poverty or having special learning needs. However, these results were inconsistent from year to year, preventing generalization of these findings.

Examining the R^2statistics for these ten models, inclusion of a dosage variable and related interaction effects have improved the goodness of fit of all of our fitted regression models. Increases in R^2statistics from question two models to question three ranged from .4 to 3.2 percentage points. The highest R^2statistic for math and reading achievement are the models for 2013, at 22.2% and 21% respectively. While the addition of dosage has improved the goodness of fit, there is still a lot of unexplained variance in math and reading achievement in the final models.

The final question examined in this study considers growth in math and reading achievement over time as a function of participation in the 21st Century after school program.

Question 4: Do attendees show greater growth in NECAP scores over time than non-attendees? Do the effects of attendance differ based on cumulative attendance over the elementary school career?

Answering question four required that the data be restructured into a person-period data set, as described in Chapter three. Data across years was combined into one data set for reading and one for math, so that the data could be examined longitudinally. As with the yearly data sets, the person-period data set was not demographically balanced between the treatment and control groups. Balance was achieved through the use of Inverse Propensity Weighting (IPW). The propensity of participating in the after school program was modeled using logistic regression for each wave of reading and math data. A student was considered a participant if they participated for 30 or more days during any of the three years of that wave. The models are included in Table 45.

Table 45: Fitted logistic regression models estimating the propensity of a student being a participant in the after school program for each wave of data, based on demographic characteristics

Reading – Wave 1	-.64+.48(HISP)-.22(AFAM)-.21(09SCHOOL)+1.69(09FRL)-.02(09POV)+.35(09ELL)+.36(09SPED)-.18(09GRADE)
Reading – Wave 2	.94+1.87(MALE)+.76(HISP)+.51(AFAM)-.79(10SCHOOL)+.73(10FRL)-1.9(10SPED)
Reading – Wave 3	-2.89-.22(MALE)-.01(HISP)+1.05(AFAM)+.15(11SCHOOL)-.20(11FRL)-.18(11POV)+.93(11ELL)-.05(11SPED)+.91(11GRADE)
Math – Wave 1	-1.65+.02(MALE)+1.43(HISP)+.24(AFAM)-.16(09SCHOOL)+1.51(09FRL) +.12(09POV) +.39(09ELL)+.31(09SPED)+.12(09GRADE)
Math – Wave 2	-1.88+2.23(MALE)+1.60(HISP)+1.70(AFAM)-.84(10SCHOOL)-1.10(10FRL)+2.22(10POV)+21.47(10ELL)+1.66(10SPED) +.47(10GRADE)
Math – Wave 3	-.72-.14(11SCHOOL)-.16(MALE)+.35(HISP)+.92(AFAM)-.23(11FRL)-.16(11POV)+.59(11ELL)+.10(11SPED) +.40(11GRADE)

The propensity scores generated by these models for each student were then used to calculate the Inverse Propensity Weight for each control student (P/(1-P). Since I am calculating the average treatment effect on the treated (ATT), all treatment students were assigned a weight of one. Once students were weighted by their IPW, demographic statistics were generated and compared to those calculated before weighting. The results are displayed in Tables 46 and 47. After weighting by IPW, no statistically significant differences remain for any of the demographic features included in this study.

Table 46: Math: Percentage of students in key demographic categores before and after weighting on Inverse Propensity Scores within each wave of data (*N* = 367)

Treat Status	# of waves	Male	Hispanic	AfAm	FRL	POV	ELL	SpEd	Grade	School	Prop Score
Cont.	1070	53	18***	7***	66***	60***	7*	11*	2.39	3.18***	.3633***
Treat	765	52	27	14	86	79	11	16	2.45	2.89	.4942
After weighting on IPW											
Cont.	735	52	25	12	86	78	10	17	2.42	2.96	.4708
Treat	765	52	27	14	86	79	11	16	2.45	2.89	.4942

* Difference in means between treatment and control group is significant at the $p < .05$ level .
** Difference in means between treatment and control group is significant at the $p < .01$ level .
*** Difference in means between treatment and control group is significant at the $p < .001$ level .

Table 47: Reading: Percentage of students in key demographic categores before and after weighting on Inverse Propensity Scores within each wave of data (*N* = 353)

Treat Status	# of waves	Male	Hispanic	AfAm	FRL	POV	ELL	SpEd	Grade	School	Prop Score
Cont.	1060	53	17***	8*	65***	59***	7*	12	3.68	3.18***	.3636***
Treat	705	51	25	11	86	78	11	15	3.66	2.94	.4860
After weighting on IPW											
Cont.	749	54	21	10	86	78	10	18	3.68	2.90	.4906
Treat	705	51	25	11	86	78	11	15	3.66	2.94	.4860

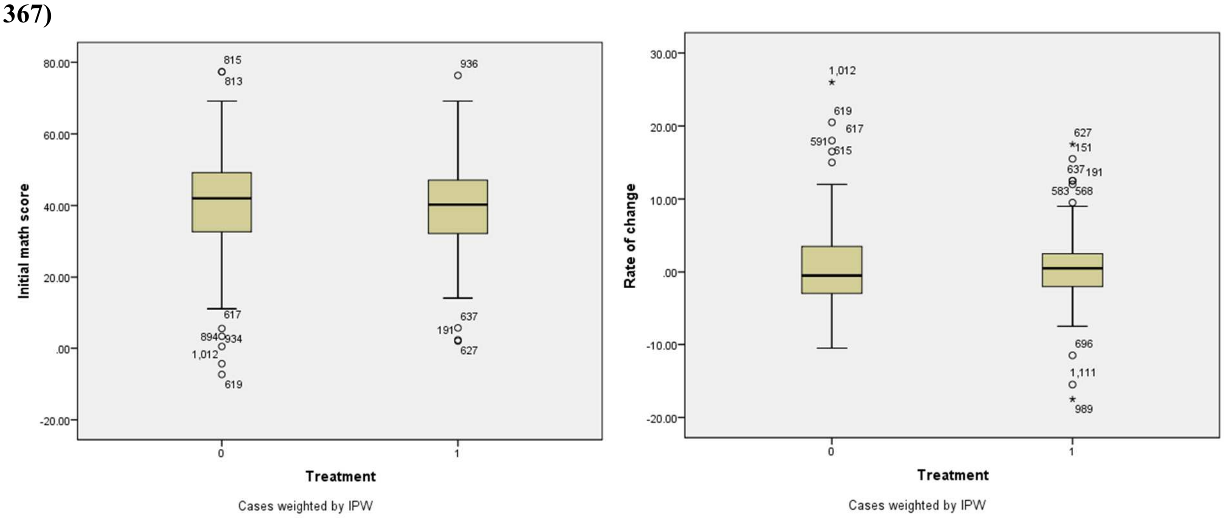
* Difference in means between treatment and control group is significant at the $p < .05$ level .
** Difference in means between treatment and control group is significant at the $p < .01$ level .
*** Difference in means between treatment and control group is significant at the $p < .001$ level .

Exploratory analysis of empirical growth trajectories for math achievement

With balance achieved, I began exploring growth trajectories for math and reading achievement for the treatment and control groups. I used OLS regression to estimate initial achievement scores and rates of change over time. Initial math scores were very similar for the treatment and control groups at 39.59 and 39.94, respectively. Participants in the after school program experienced a lower rate of growth in their math scores over time, on average, than non-participants. The average rate of change for the control group was .59 points, compared with .43 points for the treatment group.

Correlations between treatment status and initial score or rate of change were not statistically significant. However, there was a statistically significant correlation between initial math score and rate of change (r = -.723, p < .001). This indicates that students with higher initial math scores had less growth over time than those students with lower scores, independent of treatment status. The box plots in Figures 44 and 45 illustrate the distributions for initial math score and rate of change for the treatment and control groups in math.

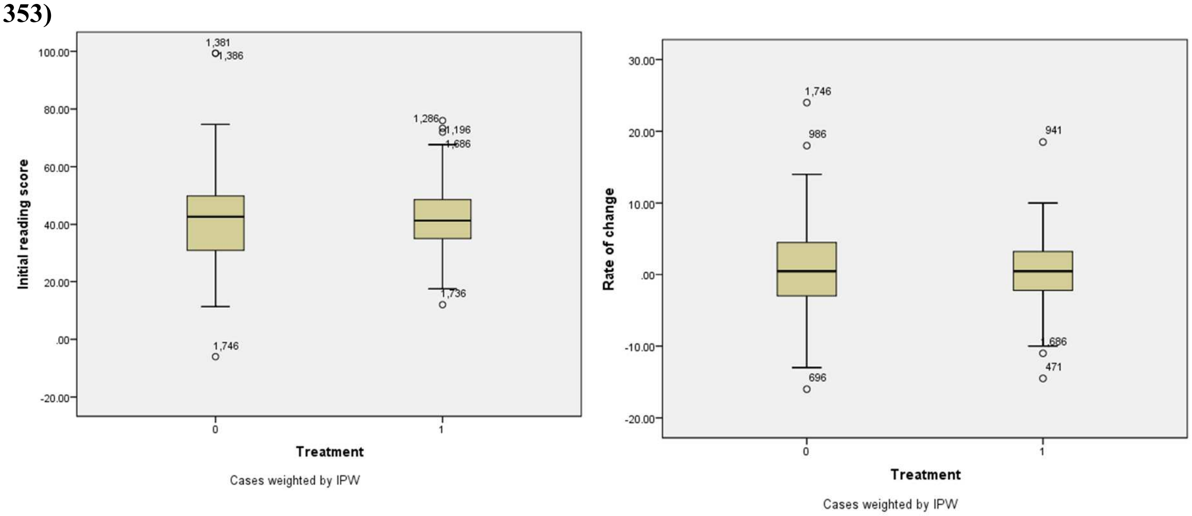
Figures 44 & 45: Box Plots of initial math score and rate of change for treatment and control groups (n = 367)



Exploratory analysis of empirical growth trajectories for reading achievement

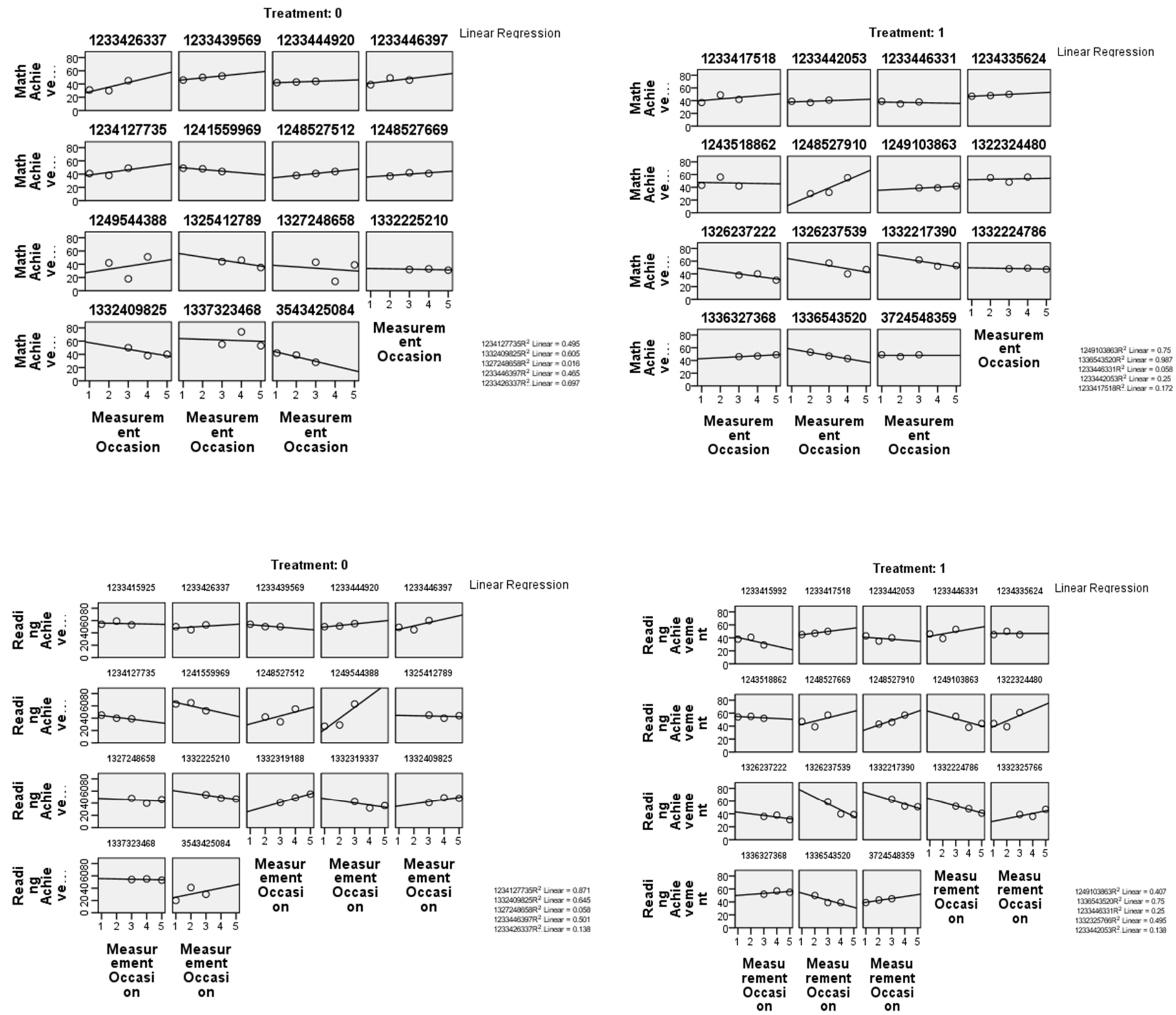
In reading, participants attending the after school program actually had higher initial reading scores than non-participants (41.65 compared with 41.36). Non-participants, however, had more than double the rate of growth in Reading NECAP scores over time, compared with their participating peers (1.02 points compared with .50 points). These relationships between program participation, initial score, and rate of change were not correlated at a statistically significant level. The relationship, however, between initial reading score and rate of change were significantly correlated. Students with higher initial reading scores experienced lower rates of growth, on average (r = -.650, p < .001), independent of treatment status. The box plots in Figures 46 and 47 illustrate the distributions for initial reading score and rate of change for the treatment and control groups.

Figures 46 & 47: Box Plots of initial reading score and rate of change for treatment and control groups (*n* = 353)



Next, growth trajectories were fit for 10% of the sample, at random, from the treatment and control groups to examine any patterns in the data. I used SPSS to graph observed achievement scores over time for each selected student, and then fit OLS regression models to estimate growth trajectories for each. Figures 48-49 display these graphs for 15 selected students in Math, and Figures 50-51 display these graphs for the same 15 students for reading achievement over time. The empirical growth trajectories looked linear and the superimposed OLS trajectories looked like they fit the observed data well. It is also clear that there was variability in initial scores and rates of change both between and within groups. These relationships were explored further in the multilevel models for change that follow.

Figures 48-51: Fitted growth trajectories by treatment status for math and reading achievement



School of attendance is a key demographic predictor in our models and was found to have a significant relationship with achievement in many of the models in the analyses to address research questions one, two, and three presented above. Therefore, I examined growth trajectories for the same sampling of students shown above, disaggregated by school for reading and math. The empirical growth trajectories appeared linear and demonstrated variability both within and between schools. From the selected sampling, however, it is notable that growth trajectories in math and reading for students at School 1, School 2, and School 5 appear flat or declining. Sampled students at School 3 appear to have more positive growth trajectories than the other schools, and School 4 appears more variable with some positive and negative growth. These growth trajectories by school and subject are included in Appendix H.

Model fitting

Next, I conducted analyses of change in math and reading achievement over time using multilevel modeling for change. The progression of models that were fit to answer question four are outlined in Table 48. The unconditional model was fit first and models achievement growth over time with no added variables. Participation is a primary variable of interest, so this was added in the second model. Effects of school, demographics, and grade were added next, in the same order as was done in research questions one through three. Next, dosage was added as a secondary variable of interest. Finally, interaction variables were added for key interactions found both in the research base (Petit et al., 1997) and in the analyses presented in questions two and three in this paper.

Table 48: Progression of fitted models estimating math and reading achievement over time

Model 1	Unconditional Model
Model 2	Participation effects
Model 3	School effects
Model 4	Demographic effects (Gender, race, poverty, ELL & Special Education status)
Model 5	Grade effects
Model 6	Impact of dosage
Model 7	School*Days
Model 8	Differential impact of dosage on students living in poverty
Model 9	Differential impact of dosage on boys
Model 10	Best fit model

Fitting the above models resulted in the two best fit longitudinal models displayed in Table 49. The variables included in these models best reduced the variance in the model, as discussed at greater length later in this chapter. The full taxonomy of fitted models is included in Appendix I.

Table 49: Parameter estimates for predicted math and reading achievement over time, controlling for participation in the 21st Century after school program, school, grade, and demographic factors

VARIABLE	$\widehat{MATHACH}$	$\widehat{READACH}$
Intercept	39.39***	44.36***
WAVE_C	3.08***	2.03**
PARTIC	N/A	-3.26*
*WAVE_C*PARTIC*	N/A	1.16~
SCHOOL 1	4.72*	2.28
SCHOOL 2	5.91**	4.22*
SCHOOL 4	5.99**	2.44
SCHOOL 5	7.29`~	2.44
*WAVE_C*SCHOOL 1*	-2.66*	-2.54*
*WAVE_C*SCHOOL 2*	-4.26***	-4.24***
*WAVE_C*SCHOOL 4*	-3.21***	-2.15**
*WAVE_C*SCHOOL 5*	-3.76**	-1.55
MALE	2.83**	-.69
HISPANIC	.34	2.59~
AFRICAN AM	-1.43	1.73
FRL	-4.55**	-2.00
ELL	-.81	-4.25*
SPED	-7.55***	-9.67***
DAYS	N/A	.03~
*WAVE_C*DAYS*	N/A	-.01~

In both data sets, there were three students with inverse propensity weights larger than two. Students with heavier weights have the potential of skewing the results of regression analyses. Due to this risk, I conducted analyses with and without these students included. No significant differences were found between the models, and therefore, the students were included in the best fit models.

Residuals for both final models are normally distributed and there are no apparent signs of homoscedasticity or non-linearity. I conducted a Shapiro-Wilk test of normality, which was not significant at the $p < .05$ level. Therefore, we do not reject the null hypothesis that the residuals are normally distributed. I also examined unstandardized residuals graphed by treatment status and school of attendance. Residuals were evenly distributed across schools and between participants and non-participants. All residual plots for question four are included in Appendix J.

In fitting the taxonomy of models above, first I fit the unconditional model, modeling initial scores and rates of change for math and reading achievement, without any additional predictors in the model. The estimated average grade three math score from this model was 39.26 and the estimated true rate of change in math scores was .53 points per year ($p < .1$). This predicts that students have positive growth, on average, in math achievement over time. The variance components of the model are statistically significant. The estimated individual variance was 48.57, significant at the $p < .001$ level. A high level of variance was also estimated in initial math scores (76.08, $p < .001$). The estimated variance for rate of change was 3.60, significant at the $p < .1$ level, as was the covariance between initial status and rate of change (-7.93). The negative estimated covariance suggests that students with higher initial test scores tend to have lower rates of change, on average. This is consistent with the descriptive correlation analyses presented above.

The unconditional model for reading scores estimated an initial reading score of 41.81, and an estimated rate of change of .37 points per year. The rate of change was not statistically significant, and also had a high level of estimated variance (9.35, $p < .001$), indicating much unexplained variability. Estimated individual variance was 40.42, and estimated variance in initial reading scores was 93.07. The negative estimated covariance (-16.71), demonstrates a negative relationship between initial reading score and rate of change. All variance components were significant at the $p < .001$ level, and with the exception of individual variance, all were higher than the comparable variance components for math achievement over time. This indicates greater variability in reading achievement than math.

The impact of participation

Next, participation in the after school program was added to the model as a predictor of initial score and rate of change. A student is considered a participant if they attended the program for 30 or more days during any of the three years of data used for that student. For math, a negative impact, though not statistically significant, was found on initial math scores, resulting in estimated initial math scores of 40.02 for non-participants, and 38.62 for participants. A similar negative, non-significant effect was found on rate of change for participants. Whereas non-participants are predicted to grow .54 points per year, participants would be predicted to grow .52 points per year. The addition of the participation variable did not alter the variance components of the model in any significant way.

Results were similar for reading with participants having an estimated initial reading score 3.26 points lower than non-participants ($p < .05$). In reading, participants were estimated to experience greater growth over time than non-participants (3.19 points vs. 2.03 points per year, $p < .10$). It should be noted that this estimated impact of participation on reading achievement was only found when days of attendance was included in the model. Parameter estimates should be considered in conjunction with parameter estimates for dosage, described below.

Impact of school of attendance

School of attendance was then added to the model as a predictor. Which school a child attended had significant effects on both their initial math and reading scores, as well as their growth over time. Table 50 illustrates the differences between schools for math and reading achievement for non-participants. The effect of participation was not significant in either model. As can be found below, students at School 3 had lower estimated initial test scores in math and reading than the other schools, but had significantly higher growth over the course of this study. The other four schools had rates of growth that ranged from slightly negative, to slightly positive. Initial estimated reading scores were higher than math scores, on average, for all five schools.

Table 50: Estimated average initial math and reading scores, and rates of change for non-participants, by school (Model 3)

School	Estimated initial math score	Rate of change, math	Estimated initial reading score	Rate of change, reading
School 1	41.36*	-.02**	42.66	-.3*
School 2	41.96**	-.92***	44.68*	-1.61***
School 3	35.69***	3.09***	40.40***	2.47***
School 4	42.39**	-.02***	43.64	-.03**
School 5	43.95~	-.68**	40.63	.38

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Estimated variance in initial math scores decreased, but was still significant at the $p < .001$ level. Estimated variance in rate of change dropped from 3.52 to 1.58 with the addition of the school predictors, and was no longer significant. Estimated covariance between initial score and rate of change also dropped from significance, decreasing to -4.51. For reading scores, the greatest impact was found on rate of change, with estimated variance decreasing from 9.36 to 7.51. All variance components for the reading model were still significant at the $p < .001$ level.

Impact of student demographics

In model 4, demographic variables were added to the model as predictors of initial score and rate of change. These predictors included gender, race, poverty status, and status as a special learner (ELL or special education). All of these except for gender and race were included as time varying predictors. Gender and special education status were found to have statistically significant impacts on initial math scores. Male students were predicted to score 3.03 points higher, on average, than their female peers on the third grade Math NECAP ($p < .05$), and students identified as part of the special education program were predicted to score 6.28 points lower than their non-identified peers, on average ($p < .01$). No significant impacts were found on rate of change for any of the demographic categories. Addition of the demographic factors lowered the estimated variance for initial test score from 70.21 to 62.33, however, this was still significant at the $p < .001$ level.

Similar results were found with Reading scores with some demographic factors having an impact on initial test scores, but none impacting rate of change in scores. Hispanic students were estimated to score higher on their initial reading assessment by 4.16 points, significant at the $p < .1$ level). Students with special learning needs were estimated to have lower initial reading scores, with students in the ELL program estimated to score 7.19 points lower ($p < .05$), on average, and students in the special education program estimated to score 12.07 points lower ($p < .001$), on average, than their peers. Addition of demographic factors reduced initial estimated variance from 92.72 to 69.48, however, this remains significant at the $p < .001$ level. Also reduced but still significant was the estimated covariance which was reduced from -15.38 to -12.71.

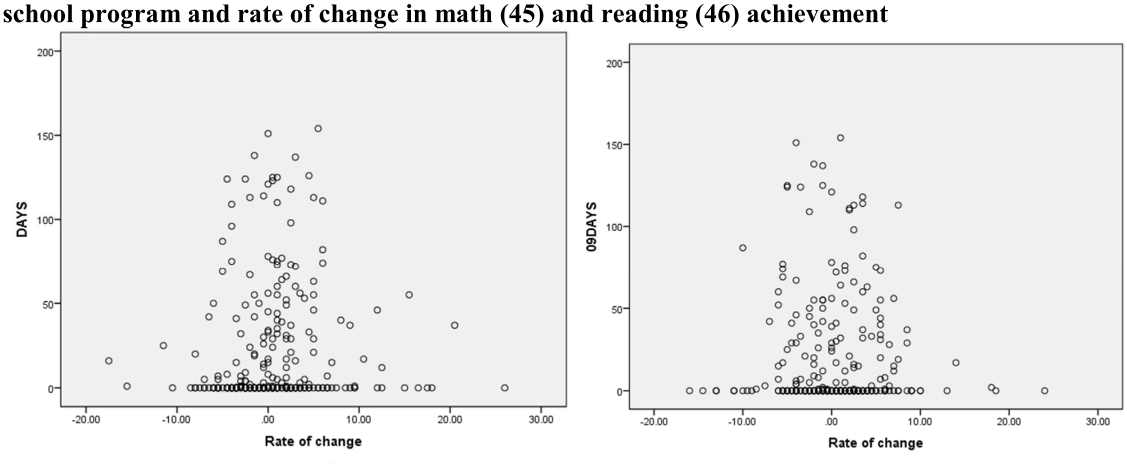
Students' grade level was added next as a predictor of their initial score and rate of change. No significant impacts of grade level were found for math achievement. Grade two participation had a positive impact on initial reading scores (8.47 points, $p < .1$), but a negative impact on rate of change (-5.18 points, $p < .05$). Inclusion of the grade level variables reduced the individual variance statistic slightly, however, all other variance components increased, indicating that inclusion of these variables may not improve the model's fit.

In answering the first part of question four, participation in the after school program does not impact initial math or reading scores, or students' growth rate in math or reading achievement. However, our initial analyses do indicate some interesting findings about what does impact achievement. Some demographic factors including gender, race, and special learning status impacted students' initial math or reading scores, but not their growth over time. Students' school of attendance, however, had significant impacts on both their initial achievement levels and their growth over time in reading and math. School effects will continue to be considered in part two of the analysis, which looks at the impact of program dosage on growth in reading and math over time.

Impact of dosage

First, I looked at the relationship between days of participation in the after school program and rate of change in reading and math achievement. Results were comparable for reading and math, both with a wider range of growth for non-participants than for high dosage participants. Students with zero days of participation had a range of growth from -20 points to +20 points for both reading and math. However, for students with high levels of participation in the program, growth rates narrowed, with a range of -10 points to +10 points for reading and math achievement. This indicates less variability in rate of change for participants than non-participants. Figures 52 and 53 provide an illustration of these relationships.

Figures 52&53: Scatterplots illustrating the relationship between days of attendance in the 21st Century after school program and rate of change in math (45) and reading (46) achievement



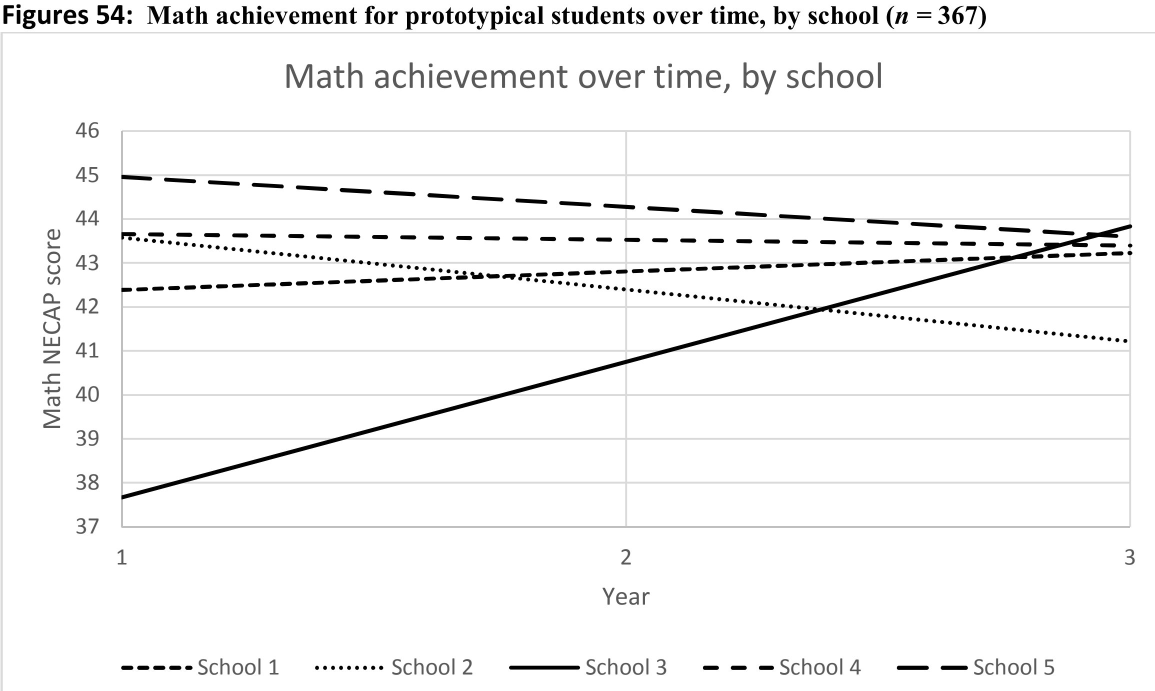
Overall, dosage had very little impact on math or reading achievement over time. The only statistically significant effects were found for the impact of days of attendance on reading achievement over time. Each day of attendance resulted in an estimated increase of .03 points, on average, in initial reading scores, and an estimated decrease of .01 points, on average, on rate of change over time. Both of these effects were significant at the $p < .1$ level. This indicates that students with higher levels of participation had higher initial Reading NECAP scores, on average, than students that did not participate or participated at low levels, but that they experienced lower rates of growth in reading achievement over time. The impact of days of attendance did not have any statistically significant effects on initial math achievement or growth over time. No effects were found for math or reading achievement for cumulative days of attendance or years of participation.

Next, I looked at interactions between dosage (days of attendance) and school, poverty status, and gender. Dosage effects for School 4 (days*School 4) resulted in a decline of .01 points per day ($p < .1$) on math achievement over time. However, the addition of this variable decreased the variance explained, and thus the pseudo-R^2 statistic. Because it did not improve the fit of the model, it was not included in the final model for math achievement over time. No other statistically significant impacts were found for school of participation for reading or math. I checked for differential impacts of program attendance for male students and students living in poverty (FRL). No effects were found for these predictors for math or reading achievement over time.

Analysis of the final models - Math

This results in a final model for math that includes estimated effects on initial math score from demographic controls (gender, race, poverty status, and special learning needs) and school of attendance. Inclusion of these variables reduces variance in initial math scores by 13.3%, although variance is still significant at the $p < .001$ level. The final model also includes effects on rate of change in math scores from school of attendance. Inclusion of school reduced variance in rate of change in math achievement by 68.1% (pseudo-R^2), and this estimated variance is no longer statistically significant. Estimated covariance in the model was reduced by 2.98 points and was no longer statistically significant. Estimated individual variance decreased by 3.27 points, and was still significant at the $p < .001$ level. Deviance statistics also demonstrate that the final model is a better fitting model, with improvement from the unconditional model (6290.120) to the final model (5995.506). Figure 54 illustrates predicted math achievement over time for prototypical students (White, male, non-participant students who qualify for free or reduced price lunch) from each of the five elementary schools included in this study.

Figures 54: Math achievement for prototypical students over time, by school (*n* = 367)



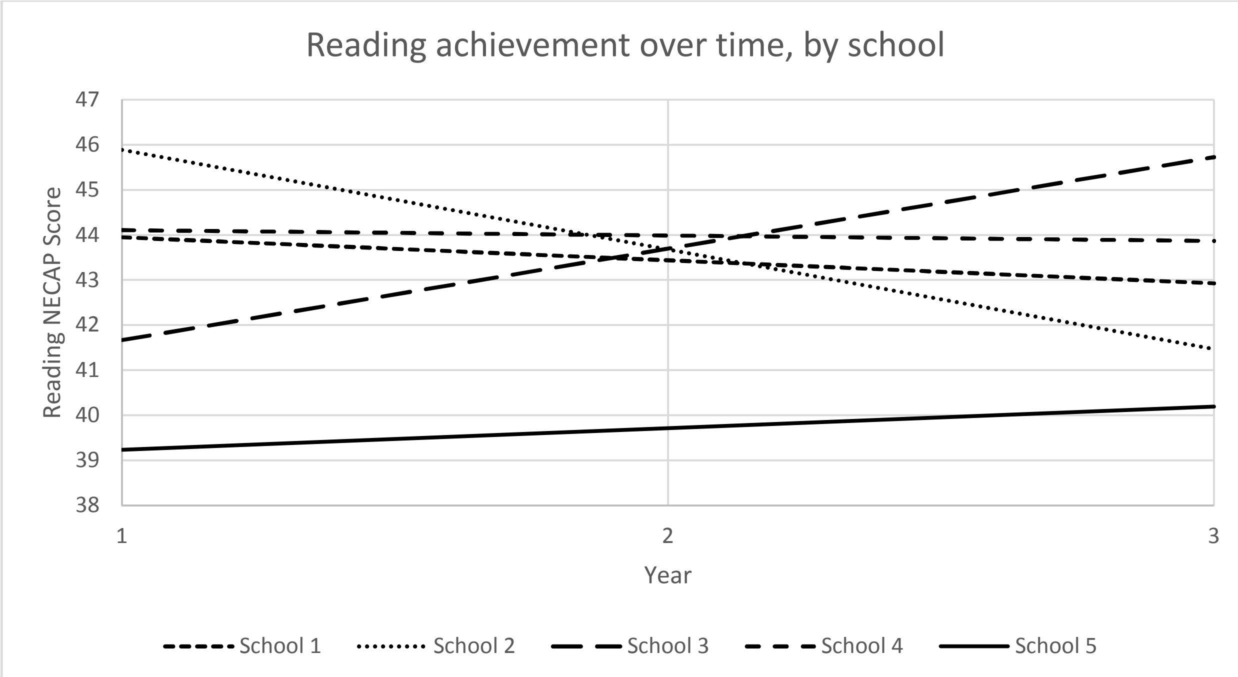
As is displayed in the graph, students at School 5 had the highest initial math scores, on average, and students at School 3 had the lowest initial scores, on average. Students at School 3, however, experienced the highest rate of change in math scores (3.08 points per year, on average). This results in School 3 students being predicted to have the highest math scores, on average, in the third wave of data. School 2, School 4, and School 5 all had negative growth, on average, in math achievement, while School 1 students demonstrated modest gains.

Analysis of the final models - Reading

The final model for reading achievement over time estimates initial reading scores predicted by demographic controls (gender, race, poverty status, and special learning needs), school of attendance, participation status, and the number of days the student attended the after school program. Inclusion of these variables decreased the estimated variance in initial math scores by 27.3%, although variance is still significant at the $p < .001$ level. The estimated rate of change in reading scores in the final model was effected by school of attendance, whether or not the student participated in the after school program, and how many days they attended. Inclusion of these predictors reduced variance in the rate of change in reading scores by 41.8%, with the remaining estimated variance significant at the $p < .01$ level. Estimated covariance in the model was reduced by 6.49 points ($p < .01$), but estimated individual variance increased by 1.34 points ($p < .001$). Deviance statistics indicate improvement in the model with improvement from the unconditional model (6034.698) to the final model (5752.400).

Figure 55 illustrates predicted reading achievement over time for prototypical students (White, male, non-participant students who qualify for free or reduced price lunch) from each of the five elementary schools included in this study.

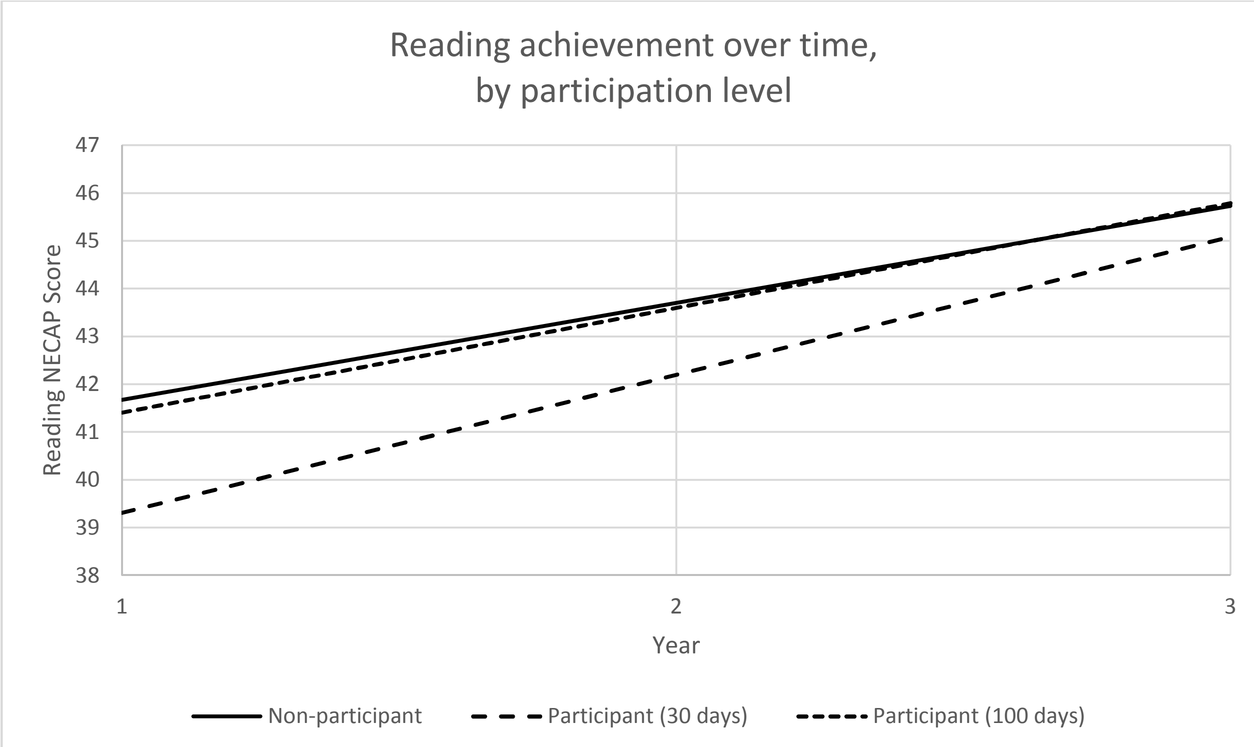
Figures 55: Reading achievement for prototypical students over time, by school (*n* = 353)



Students at School 5 had the lowest initial reading scores, at 39.23 points, on average, while students at School 2 demonstrated the highest average scores, at 45.89 points. Similar to math achievement, students at School 3 had the highest rate of growth in reading scores (2.03 points per year, on average). This resulted in School 3 having the highest predicted reading scores, on average, in the third wave of data. Students at Schools 1, 2, and 4 experienced negative growth in reading, on average, while School 5's students demonstrated modest gains.

Figure 56 highlights the estimated effects on reading achievement over time of participation in the after school program at high and low doses for prototypical students at School 3.

Figures 56: Reading achievement for prototypical students over time, by participation level (*n* = 353)



Students participating in the 21st Century after school program at low dosages (e.g., 30 days) are predicted to have lower initial reading scores than non-participants or high dosage participants. Non-participants and high dosage participants (100 days, for example) have almost identical initial reading scores (.26 points difference). Low dosage participants, however, are predicted to demonstrate the highest rates of growth in reading achievement (2.89 points per year, on average). Even with their higher growth rate, they do not catch up with their non-participating or higher dosage participating peers within the three waves of data. High dosage participants have the next highest rate of growth at 2.19 points per year, on average. After three waves of data, high dosage participants are predicted to outscore non-participating peers, on average, in reading. While non-participants have the highest average initial reading scores, they have the flattest estimated growth rate, at 2.03 points per year.

Summary

Student participation alone does not predict any significant difference in student achievement in reading or math over time. When program dosage, as measured by days of attendance within a school year, is added to the model, significant impacts are found on reading achievement over time, but not math achievement. Our models estimate that students participating at high levels start with initial reading scores similar to that of non-participants, but show a higher rate of growth in achievement. In contrast, low-level participants have lower initial achievement, on average, but show the highest rates of growth in reading achievement over time. No effects were found for cumulative attendance over time or years of attendance in the program. School of attendance was a significant factor impacting estimated initial scores and rate of change in math and reading achievement.

Throughout the analysis, participation in the 21st Century after school program has been considered in a variety of ways. Effects on math and reading achievement have been considered for participants and non-participants within an academic year and over time. Program dosage has been considered by adding days of attendance to yearly and longitudinal models. Finally, interaction effects of participation and dosage have been considered for models within the academic year and over time. Implications of these findings for educators and policy-makers will be discussed in chapter five.

Conclusion

Purpose of the Research

Childhood poverty remains a sociological concern in the United States. With childhood poverty rates of 22% (Laughlin, 2014), the needs of this group must be considered by U.S. leaders and policy-makers. Living in poverty has a negative impact on academic achievement (NCES, 2011a; NCES, 2011b; Sirin, 2005). These negative impacts hold across age and subject area. Over the years, the federal government has funded numerous programs to address the ill effects of childhood poverty on academics. These programs have included Title One services, after school tutoring, mentoring, school restructuring, and after school programs (Currie, 2000). Federal funds, however, are not infinite, and there is little consensus within the research community on which programs are most effective for improving academic outcomes for children living in poverty (Darling-Hammond, 2000). It is critical to determine which interventions are most effective so that federal funds can be strategically targeted for maximum benefit.

After school programs have been a popular intervention for students living in poverty for decades (Dynarski et al., 2004). They provide child care for working parents, provide safe havens for youth, and prevent delinquency (Lord & Mahoney, 1997). In the past 20 years, after school programs have taken on a more academic focus, often including tutoring or homework assistance (James-Burdumy et al., 2007). Finally, after school programs can reduce the extracurricular gap that exists for students living in poverty (Laughlin, 2014; Posner & Vandell, 1999) by providing enrichment activities that they might not otherwise be able to access.

The federal government began funding the 21st Century after school programs in 1994. The federal government currently spends over one billion dollars per year on the 21st Century program which serves over one million students annually (U.S. DOE, 2015). These programs target students attending high poverty schools with the specific primary goal of raising student achievement in reading and math.

Overview of the Study

My research study investigated the effect of participation in the 21st Century after school program on academic achievement in reading and math for students attending high poverty schools. Furthermore, I explored whether program dosage impacted outcomes, and if participation or dosage effects varied for students based on their demographic characteristics, grade level, or school of attendance. Finally, I considered whether participation impacted academic growth in reading or math over time.

This study was important due to methodological flaws in the existing research base. Very few studies of after school programs have utilized experimental or quasi-experimental designs, and those that have, often failed to use appropriate statistical matching techniques (Scott-Little et al., 2002). Another criticism of the existing research is a failure to consider program dosage in evaluations of after school programs' effectiveness (Chaput et al., 2004). Other researchers have called for more research on the differential impacts of program participation on subgroups of students, such as students living in poverty (Gardner et al., 2009). Finally, few studies have modeled effects of after school program attendance over time (Roth et al., 2010).

This research study addressed many of the limitations in the existing research base. This secondary analysis utilized a large existing data set containing information on students attending five high poverty elementary schools in Nashua, New Hampshire. These five schools have run federally funded 21st Century after school programs for over 15 years. The data studied spans the years from 2008-2013, allowing for longitudinal analysis of change over time. In this data set, program participants and non-participants differed on key demographic variables, including race and poverty status. As has been noted in previous studies (Petit et al., 1997; Russell & Woods, 2012), students identified as Hispanic or African American, and students qualifying for free or reduced price lunch participated in the after school programs at higher rates and intensities. Therefore, inverse propensity weighting was used to ensure proper statistical matching of the treatment and control groups.

This study is notable for viewing student participation in the after school program in multiple ways. Students were considered a participant if they attended the 21st Century after school program 30 days or more within one school year. Students might have participated one year, and not the next, causing their coding to change over time. Grade of participation and school of participation were also included in analysis. Intensity and duration of program attendance were also included in this study. Intensity was considered as the number of days the student attended the program within one school year. In longitudinal analysis, yearly dosage was allowed to vary year to year, and cumulative days of attendance was also included. Duration of attendance was included in the longitudinal analysis as total years having attended the program.

As with all studies, this study has limitations related to lack of access to all possible relevant information. It needs to be acknowledged that this is not a randomized study. Data was studied retrospectively. While inverse propensity weighting was used to ensure that treatment and control groups were matched on key variables, it is possible that students that attended the 21st Century programs were different in some important way from students in these schools who did not attend (e.g., maternal employment or marital status). This other difference between them could be the factor that impacted any difference in academic outcomes. This could be a rival hypothesis to findings presented.

Furthermore, I did not collect data on what non-participants did after school. These children could have done nothing after school, could have attended a different, off-site, after school program, or could have participated in an array of extra-curricular activities. With the counterfactual unknown, we must be cautious in drawing conclusions. There is also not data on how many hours each day participants spent in the after school program or what activities they participated in during their after school time. Having this information would provide more nuance to dosage variables. Finally, there is limited ability to comment on program quality within or between the five after school programs in this study. The state CIPAS report provides an evaluation of the district's program, but provides no comments on individual site quality. There is not data on the quality of staff-student interactions or student engagement at any of the sites. These limitations should be considered in interpreting results.

Findings: The Impact of Participation in the 21st Century Program

The impact of participation in the 21st Century after school program on students' math and reading achievement was examined for each of the five years of this study (2008-2013) using OLS multiple regression. It was found that program participation had a statistically significant negative impact on math and reading achievement in 2009 and 2011. These impacts ranged between negative one and negative two points. The impact of participation in the other three years was not statistically significant. These are discouraging findings and indicate that the 21st Century program may not be achieving its academic program goals. These results run contrary to other studies that found positive effects on achievement test scores from after school program participation (Durlak, et al., 2010; Sheldon et al., 2010). In fact, results are more similar to those found in the Mathematica study (James-Burdumy et al., 2007) which found a negative impact of program participation.

Some differential effects were found for program participation. Mixed results were found for participation at the different grade levels. Statistically significant positive results were found for grade two and three participation on math achievement and grade two and five participation on reading achievement. Impacts ranged from 2.88 to 3.88 points. Negative impacts of program participation were found for grade two and five participation on math achievement and grade two participation on reading. Impacts ranged from -2.03 to -5.44 points. These contradicting findings from year to year make it difficult to make any policy recommendations regarding target years for after school programming.

Students who participated in the after school program at School 1 in 2009 scored 4.05 points higher in math than non-participants, controlling for demographic variables. These findings are encouraging and provide possible support for after school programs as an academic intervention.

Finally, positive impacts of participation were seen for certain demographic subgroups. Students identified as Hispanic that participated in the program were estimated to score 4.71 points higher on the math assessment, on average, based on the 2013 data. Based on this model, Hispanic participants are predicted to outscore White participants. Similarly, students identified as African American who participated in the program were estimated to score 4.68 points higher on the reading assessment, based on 2011 data. Based on this model, African American participants are predicted to outscore White participants and non-participants, holding other factors constant. Finally, a positive effect was seen for students identified with special needs who participated in the program in 2010. They scored 3.99 points higher on the reading assessment, on average, than students with special needs that did not participate. These results are promising for these subgroups. More research is needed to determine if these results hold across settings.

Findings: The Impact of Program Dosage

Next, the impact of program dosage on math and reading achievement was studied for each year of the data using OLS multiple regression. Dosage was defined as the number of days a student attended the program during each school year preceding the NECAP assessment. A statistically significant positive impact of program dosage was found on math achievement in 2011. This model predicts that a student who attended the after school program for 50 days or more would score higher on the math assessment then students that did not attend the program. Statistically significant positive impacts were seen in reading achievement in the 2011 and 2013 data. These models predict that program participants would outscore their non-participant peers after 83 and 102 days, respectively, of program attendance.

While these results were not found for each year of this study, they do correspond with the research base. Other researchers (e.g., Auger et al., 2013; Huang et al., 2009; James-Burdumy et al., 2007; Pierce et al., 2010) of after school programs have found positive academic impacts of higher dosage levels on student achievement. These results are encouraging and suggest that a higher intensity of program participation leads to higher achievement in math and reading. These results also indicate that there is a threshold point of participation needed to see results. Based on the results of this study, the dosage threshold is lower for math than reading.

Several effects were found for interaction variables between dosage and student characteristics. Grade level of participation again showed mixed results. A higher dosage of participation in grade two resulted in higher math and reading scores in 2009, but lower reading scores in 2011. Higher doses in grade three resulted in higher math and reading scores in 2010, but lower math scores in 2009. No effects were found for program dosage in grade four. Statistically significant results were seen for grade five dosage, with positive effects found on reading achievement in 2012 and negative effects for math achievement in 2011.

These results are similar to those found in Lauer et al.'s (2004) meta-analysis which also showed mixed results by grade level. Lauer et al., however, found that on average, greater program dosage in the younger grades had more positive impacts on reading scores, while higher doses in the upper elementary years had a greater impact on math scores. The results presented here are variable and make it difficult to generalize results. However, higher doses in the upper elementary years (grade five) resulted in lower math scores and higher reading scores, which was unexpected given Lauer et al.'s findings.

Several differential effects of program dosage were found by school/program of attendance. The more days a student attended the after school program at School 4, the lower they would be predicted to score on the Math NECAP, based on data from 2010 and 2011. Different effects of program attendance were also found for subsets of populations attending the different schools. Students identified with special needs who attended the after school program at School 4 had mixed results. They were predicted to score lower in reading than their non-participating peers in 2010, but higher in 2012. Students with special needs attending the after school program at School 5 scored lower than their non-participating peers based on 2010 data. Positive effects of program dosage were found on reading achievement for students in the ELL program that attended the after school program at School 2, based on 2013 data. These mixed results for the impact of program dosage on students with special learning needs suggests a need for more research in this area.

The 21st Century program is a targeted intervention for students attending high poverty schools, with a primary goal of increasing academic achievement in reading and math. The 21st Century programs in Nashua enroll a higher percentage of students that qualify for free or reduced price lunch than are found in the school population. No differential effects were found for program dosage for students that qualified for free or reduced price lunch (FRL). However, several differential effects were found for these students by school/program of attendance. A higher dosage of attendance in the after school program at School 1 predicted higher math achievement (2009 & 2013) and higher reading achievement (2010) for students qualifying for FRL. These results are encouraging and suggest an ability of the program to meet its stated goals.

Results were not so encouraging for students qualifying for FRL that attended the program at higher dosages at Schools 2, 4, or 5. A negative impact of dosage on reading achievement was predicted for this group of students at School 2 (2011), School 4 (2011 & 2012), and School 5 (2011). These results are concerning in that they suggest that the program could be having a negative impact on the students it aims to help. Results are variable and would require further study before enacting any policy changes. Miller (2003) found that the quality of the after school program explained 27-47% of the variance in outcome measures. Program quality is not assessed in this study; however, this could be an important consideration in understanding site-based differences in outcomes.

Differential impacts of program participation were found for certain demographic groups. Males that attended the after school program at higher doses were predicted to score lower in math and reading based on 2011 data. No significant results were found for males in any of the other years of the study. In the literature, Vandell et al. (2007) found mixed results of program dosage on males' social skills and behavior, however, these results did not consider academic impacts. Positive impacts of program dosage were found on math achievement for students in the ELL program (2012) and for students identified as Hispanic (2013). These results indicate that after school participation could effectively close the achievement gap that exists for these populations of students. Further research would be needed to determine if these results hold across settings.

Findings: Longitudinal Impacts of Participation & Dosage

The final question examined in this study considered the effects of participation and dosage in the after school program over time. No statistically significant results were found for math achievement. This suggests that students that attend the program over the course of three years do no better or worse than their peers that do not attend the program at all. Results were found, however, for reading achievement over time. Students who participated in the program at low doses had lower initial reading scores than high dose participants or non-participants, but had the highest rates of growth. High dose participants had slightly lower (.26 points) initial reading scores than non-participants, but had a higher rate of growth. High dose participants are predicted, in fact, to outscore non-participants after three years.

Mixed results of longitudinal analyses of the effect of after school programs on academic achievement are also seen in the literature (Cosden et al., 2001; James-Burdumy et al., 2007; Lauer et al., 2004). Similar to the results found here in reading, Scott-Little et al. (2002) also found that students with the lowest initial achievement scores experienced greater gains over time from program participation. Interestingly, the results of this study run contrary to longitudinal studies by Huang et al. (2009) and Welsh et al. (2002) that found greater gains over time in math achievement, but not reading, for program participants. The longitudinal results found for reading, but not math in this study, could indicate a program focus on reading intervention during the after school hours.

As stated above, low dosage participants had higher growth rates in their reading achievement than high dosage participants. This finding runs contrary to findings from this study looking at the effects of dosage within an academic year. This could indicate that higher doses of program participation are beneficial within a year, but that over time, high levels of participation buffer positive impacts of program participation.

No differential effects of participation or dosage were found by school, grade level, or student characteristics in the longitudinal models. What is notable, however, is the effect of school of attendance over time. School 3 had the lowest initial math scores and the second lowest initial reading scores of the five schools. Students attending School 3 had the highest rate of growth in reading and math, resulting in the highest average reading and math scores after three years. This finding points to the critical importance of school day instruction for student learning and the need for high quality after school programs to coordinate with and supplement school day instruction.

Conclusions

This study resulted in variable findings on the impact of after school program participation. While many positive impacts of program participation have been found, these vary by year and subject area. Possible reasons for the variability of results include methodological features of this study, variance in program quality, and a misalignment of program features and federal goals for 21st Century programs.

While this study's methodology overcomes many of the flaws in the existing literature, it does lack a nuanced consideration of the activities that participants and non-participants engage in during the after school hours. Information was not available on the particular activities that participating students engaged in during their time in the 21st Century program. If available, this information might account for some of the unexplained variance in student achievement across schools. Furthermore, information was not available on the after school activities of non-participants. Some students may have participated in off-site after school programs, while others may have been home with a parent, and still others were unsupervised. Having students with this variety of after school experiences considered together as the control group likely underestimates program effects of 21st Century Program participation.

Throughout this study, differences in student achievement were found by school and program of attendance. This indicates variability in program quality and school day instruction that were beyond the scope of this study. Miller (2003) found that program quality accounts for 27-47% of variance in program outcomes. Several researchers (Cross et al., 2010; Leos-Urbel, 2015; Miller, 2003; Vandell et al., 2005) found that the quality of staff-student relationships mediates student engagement and is the strongest predictor of youth outcomes in after school programs.

The five programs included in this study shared a common district-level director; however, they likely had differences that could impact program quality and student outcomes. The quality of program staff is critical to building supportive relationships with students. Each of the five programs had a different site coordinator, as well as different staff workers. Consideration of staffing ratios, staff turnover, and staff training and support would likely account for some of the variation in outcomes between schools.

Furthermore, coordination between after school program and school day staff is important to program quality and student outcomes (James-Burdumy et al., 2007; Pierce et al., 1999). Each of the five programs had a different building principal. Consideration of principal support for the after school program, as well as principal involvement in programming would be important in measures of program quality. Program offerings likely varied between sites. Some schools may have had more recreational offerings, while others focused more on academic classes, or the arts. Alignment of program offerings with program goals and outcome measures is an important consideration in examining variability in outcomes.

The goals of after school programming have shifted over time. In their origins, after school programs were designed to provide safe, supervised environments for children during the after school hours (Dynarski et al., 2004). With the federal government's funding of 21st Century Community Learning Centers in the 1990's, the program goals expanded to include bringing families and the community into the schools during after school and weekend hours. After school programs were seen as a forum for positive youth development, and a venue for boosting resilience through the offering of academic, enrichment, and recreational programming for students (James-Burdumy et al., 2007).

During this time period, after school programs were examined through the lens of building resiliency in students at risk of poor academic outcomes due to their poverty status and/or their attendance at high poverty schools. After school programs can reduce environmental risks for children living in high poverty areas, and help boost their resilience to stress by building their social competence and problem solving skills. Resilience is thought to be boosted by fostering protective factors in children (Woodland, 2016). Researchers and theorists have suggested that this is done by developing meaningful, caring relationships with adults and peers, setting high expectations, and providing opportunities for mastery (Egeland et al., 1993; Masten & Coatsworth, 1998; Woodland, 2016; Zolkowski & Bullock, 2012).

However, with the massive funding increases that came to the program with its connection to the No Child Left Behind legislation in 2001, dramatic shifts to program goals occurred. While the increased funding certainly facilitated the expansion of the program to many schools, it resulted in a narrowing of program goals. At this point in time, the specific measurable goal of student growth toward meeting state standards in reading and math was front and center in the program's charge, leading to the use of state-level standardized achievement tests as a key outcome measure (U.S. Department of Education, 2003).

While building academic competence is one important component of promoting resiliency, it is a narrow slice of a much larger pie. As policy-makers and program staff focus more on test scores, the more comprehensive vision of student supports is lost. An over-focus on teaching discreet academic skills runs the risk of decreasing student engagement, thus leading to less positive student outcomes.

This leaves policy-makers and program staff with a dilemma. One option appears to be to maintain the primary goal of increased student academic achievement and to realign program activities toward this end, knowing that this risks negatively impacting student engagement in the program. An alternative option is to widen the lens of program goals and realign with the resiliency framework. This would include a wide array of goals and outcome measures. Academic competence would be considered along with other important goals of social emotional learning. Other areas of focus might include social skills, opportunities for problem solving, character development, and boosting self-confidence (Hirsch et al., 2010).

This is work of critical importance. Children living in poverty and children attending high poverty schools are being left behind academically, limiting their future life choices and perpetuating the cycle of poverty. Promoting resiliency to the daily challenges they face by boosting their academic and social competence is a way to break this cycle. Results of this study indicate some promise of after school programs to do this for students attending high poverty schools, particularly with regard to reading achievement. Intensity and duration of attendance were found to be critical factors to realizing program effects. Specific recommendations for future research and policy consideration are outlined below.

Recommendations for Future Research

More research is needed on the academic impact of after school programs. The generalizability of results from this study would be boosted by its replication with other populations. This study was conducted using data gathered on students attending urban elementary schools in New Hampshire. The external validity (Shadish et al., 2002) of the study would be compromised by application of findings to drastically different settings. Similar studies should be conducted utilizing data from other grade levels (middle and high school) and other regions of the United States. The outcome measure used in this study (NECAP) is no longer administered in New Hampshire and has been replaced with the Smarter Balanced exam. It would be interesting to study academic effects of after school program attendance utilizing this new outcome measure.

Future research should also include more detailed information on the after school plans of students in the control group. Vandell et al. (2007) found longitudinal impacts on math achievement test scores after two years of program participation, compared to students who were unsupervised during the after school hours. That is a different comparison group than comparing participants to students who were home with a parent or who participated in a different structured after school activity. In this study, there is not data on what students in the control group did after school. This lack of nuance in the comparison group has likely minimized effects seen for program participation.

In this study, variable effects were found for the differential impact of program participation and dosage for students at different grade levels and for students of different subgroups (gender, race, special learning status). Differential impacts can provide important guidance for policy-makers in targeting programs to those who most benefit. The inconsistent findings for differential effects in this study make it challenging to make any substantive policy recommendations regarding target audiences for after school programs. Further study of this area would boost the ability to make sound policy recommendations.

School and program differences were found for all sections of this study and for most years of the data. One variable that differs by school is the school principal and the after school program coordinator. Further research is recommended to study the impact of building and program leadership on the academic achievement of students. Furthermore, the quality of after school programs should be further studied as a possible predictor of academic achievement differences for students attending different sites. Previous researchers have found that measurable indicators of site quality, such as supportive adult and peer interactions, impact outcomes (Leos-Urbel, 2015; Miller, 2003; Pierce et al., 2010). Structured environments that provide homework assistance and have regular communication between program and school staff have also been found to produce better academic outcomes (James-Burdumy et al., 2007; Pierce et al., 2010). Including measures of these site quality indicators in future research would help practitioners and policy-makers to design and support programs that make a difference for children.

Recommendations for Policy & Practice

Answering these questions has real practical and policy implications for educational leaders. The federal government has invested billions of dollars over the past 20 years on after school programs. It is critical to know if this investment is yielding results for the students it aims to help, or if this money would be better targeted in a different manner to help children at risk. Furthermore, insight into the nuances of who is best helped and by which kinds of programs will help leaders and policy-makers better design programs to get better results for the children they serve. With limited government funds and millions of children living in poverty in the United States, this is of critical importance.

The results of this study indicate variable findings for the impact of participation in the 21st Century after school program on academic achievement. There are some glimmers of success where high doses of attendance or attendance over time improved academic outcomes, especially in reading. High doses of program participation were even found to close the achievement gap for particular groups of students during some years of the study. However, other years found no results or even negative results of participation. This does not make a strong case for increased federal support of these programs for the purpose of improving academic achievement; nor does it present a strong case for decreasing support of these programs as a venue for promoting resiliency in youth.

After school program staff should educate parents about the importance of program dosage as a predictor of academic achievement. Encouraging students to attend at higher rates within a school year, as well as for multiple years would likely improve student achievement in reading and math. Improving the quality of programs is one important component of increasing students' desire to participate in the program more frequently.

School-level variables have significant effects throughout this study. Beyond further research of school and site effects, as recommended above, these findings indicate that improving the quality of after school programming may have beneficial achievement outcomes. Educating district directors and site coordinators about the elements of successful after school programs, and enforcing high quality standards at each site through state-level monitoring would likely improve student results. Broadening the CIPAS evaluation process to include site level monitoring and ratings would be an important first step. In addition, more federal funding should be targeted toward staff training with the goal of improving program quality.

That being said, after school programs provide many benefits to students and their families beyond improved academics. As program goals have shifted over the past 15 years towards a narrower focus on academics, other program components that promote student resiliency have been unmeasured and overlooked. Student safety, equity of access, and social-emotional learning are critical program components that decrease risk and promote resiliency for children living in poverty.

Twenty First Century programs provide a safe, supervised environment for students after school who may otherwise be unsupervised. These programs keep the students safe and also shelter them from possible engagement in delinquent behaviors in the after school hours. Subsidized after school programs also serve an equity function, expanding opportunity for children living in poverty. Affluent students engage in a variety of extracurricular activities that are often unattainable for low income families. 21st Century after school programs provide access to activities such as sports, cooking, or arts activities for their participants. Access to caring and competent adult mentors promotes social skills and self-confidence. The programs also provide peace of mind for working parents to know that their children are cared for after school.

These benefits are not measured by academic achievement scores, but are, nonetheless, just as critical. Policy-makers should reconsider the goals of the 21st Century program and broaden their evaluation beyond just academic indicators. Consideration of the various ways these programs promote resiliency in youth would give a more nuanced view of program benefits. Recognizing and celebrating the varied benefits of after school programs will help to ensure sustainability of high quality after school programming for the decades to come.

Bibliography

After School Alliance (2015). Evaluations Backgrounder: A summary of formal evaluations of afterschool programs' impact on academics, behavior, safety and family life. Retrieved from www.afterschoolalliance.org.

Auger, A., Pierce, K. & Vandell, D.L. (2013). Participation in out of school settings and student academic and behavioral outcomes. Paper presented at the Annual Meeting of the American Education Research Association, April 28, 2013.

Beck, E.L. (1999). Prevention and intervention programming: Lessons from an after-school program, *Urban Review (31)*, 107-124.

Boys & Girls Clubs of America (2015). *Who We Are: Our History.* Retrieved from www.bgca.org/whoweare/Pages/History.aspx.

Center for American Progress (2015). State Reports: New Hampshire. Retrieved from talkpoverty.org.

Chaput, S.S., Little, P.M.D., & Weiss, H. (2004). Understanding and measuring attendance in out-of-school time programs. *Issues and Opportunities in Out-of-School Time Evaluation, 7*, 1-12.

Committee on Appropriations (2003). *Investment in after-school programs: A hearing before a subcommittee of the Committee on Appropriations United States Senate*. Washington, D.C.: May 13, 2003. Retrieved from http://www.access.gpo.gov/congress/senate.

Condly, S.J. (2006). Resilience in children: A review of literature with implications for education. *Urban Education, 41*(3), 211-236.

Connell, J.P. & Gambone, M.A. (1999). *Youth development in community settings: A community action framework.* Unpublished manuscript. Retrieved from http://www.ydsi.org/YDSI/pdf/publication_02.pdf.

Cosden, M., Morrison, G., Albanese, A.L. & Macias, S. (2001). When homework is not home work: After-school programs for homework assistance. *Educational Psychologist, 36*(3), 211-221.

Cosden, M., Morrison, G., Gutierrez, L. & Brown, M. (2004). The effects of homework programs and after-school activities on school success. *Theory into Practice, 43*(3), 220-226.

Creswell, J.W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 3rd Edition.* Thousand Oaks, CA: Sage Publications.

Cross, A.B., Gottfredson, D.C., Wilson, D.M., Rorie, M. & Connell, N. (2010). Implementation quality and positive experiences in after-school programs. *American Journal of Community Psychology*, *45*, 370-380.

Currie, J. (2010). Early Childhood Intervention Programs: What Do We Know? Los Angeles, CA: UCLA and NBER.

D'Amico, J.J. (2001). A closer look at the minority achievement gap, *ERS Spectrum*, *19*(2), 4-10.

Darling-Hammond, L. (2000). Transforming Urban Public Schools: The Role of Standards and Accountability, Retrieved from ERIC.

DeNavas-Walt, C. & Proctor, B.D. (2014). Income and Poverty in the United States: 2013. U.S. Census Bureau, Current Population Reports, p. 60-249, Washington, DC: U.S. Government Printing Office.

Dietel, R. (2009). After-School programs: Finding the right dose. *Phi Delta Kappa, 91*(3), 62-64.

Duncan, G.J., Yeung, W.J., Brooks-Gunn, J. & Smith, J.R. (1998). How much does childhood poverty affect the life chances of children? *American Sociological Review*, *63*(3), 406-423.

Durlak, J.A., Weissberg, R.P. & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology, 45*, 294-309.

Dynarski, M. (2015). *The $1.2 billion after school program that doesn't work.* Retrieved from Brookings.edu – March 19, 2015.

Dynarski, M., Moore, M., James-Burdumy, S., Rosenberg, L., Deke, J. & Mansfield, W. (2004). When schools stay open late: The national evaluation of 21st century community learning centers program. Princeton, NJ: Mathematica Policy Research.

Egeland, B., Carlson, E. & Sroufe, A. (1993). Resilience as process. *Development and Psychopathology, 5*(4), 517-528.

Elementary & Secondary Education Act (ESEA) (2011). *21st Century Community Learning Centers: FY 2010 Program Performance Report.*

Fashola, O.S. (1998). *Review of Extended-Day and After-School Programs and their Effectiveness.* Center for Research on the Education of Students Placed at Risk, Report No. 24, October 1998.

Frankel, S. & Daley, G. (2007). *An Evaluation of After School Programs Provided by Beyond the Bell's Partner Agencies: Executive Summary and Synopsis of Methodology and Findings*. New York: Partnership for After School Education.

Freedman, D.A. & Berk, R.A. (2008). Weighting regressions by propensity scores. *Evaluation Review, 32*(4), 392-409.

Galambos, D.L. & Maggs, J.L. (1991). Out-of-school care of young adolescents and self-reported behavior, *Developmental Psychology (27)*, 644-655.

Gardner, M., Roth, J.L., & Brooks-Gunn, J. (2009). *Can After-School Programs Help Level the Playing Field for Disadvantaged Youth?* Equity Matters: Research Review (No.4). Campaign for Educational Equity, Teachers College, Columbia University.

Glazer, N.T. & Williams, S. (2001) Averting the homework crisis. *Educational Leadership, (58),* 43-45.

Gordon, K.A. (1996). Resilient Hispanic youths' self-concept and motivational patterns. *Hispanic Journal of Behavioral Sciences, 18*(1), 63-73.

Graham, S.E. & Kurleander, M. (2011). Using propensity scores in educational research: General principals and practical applications. *The Journal of Educational Research, 104*(5), 340-353.

Graham, S.E., Singer, J.D. & Willett, J.B. (2008). An introduction to the multilevel model for change. In P. Alasuutari, L. Bickman & J. Brannen (Eds.), *The Sage Handbook of Social Research Methods* (pp. 377-394). LA: Sage Publications.

Grossman J.B., Walker, K., & Raley, R. (2001). *Challenges and opportunities in after-school programs: Lessons for policymakers and funders.* Philadelphia, PA: Public/Private Ventures.

Guo, S. & Fraser, M.W. (2014). Propensity Score Analysis: Statistical Methods and Applications, 2nd Edition. LA: Sage Publications.

Hausner, M.E. (2000). *The Impact of Kindergarten Intervention Project Accelerated Literacy on Emerging Literacy Concepts and Second Grade Reading Comprehension (*Doctoral Dissertation). Retrieved from U.S. Department of Education Educational Resources Information Center (ERIC).

Hirsch, B.J., Mekinda, M.A., Stawicki, J. (2010). More than attendance: The importance of after-school program quality. *American Journal of Community Psychology*, *45*, 447-452.

Ho, D.E., Imai, K., King, G., Stuart, E.A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*(3), 199-236.

Huang, D., Gribbons, B., Kim, K.S., Lee, C. & Baker, E.L. (2000). *A Decade of Results: The Impact of the LA's Best After School Enrichment Program on Subsequent Student Achievement and Performance*. Los Angeles, CA: UCLA Center for the Study of Evaluation.

Huang, D., Leon, S., Harven, A.M., La Torre, D., & Mostafavi, S. (2009). *Exploring the relationship between LA's BEST program attendance and cognitive gains of LA's BEST students.* Los Angeles, CA: UCLA Center for the Study of Evaluation.

James-Burdumy, S., Dynarski, M. & Deke, J. (2007). When elementary schools stay open late: Results from the national evaluation of the 21st Century community learning centers program. *Educational Evaluation and Policy Analysis, 29*(4), 296-318.

James-Burdumy, S., Dynarski, M. & Deke, J. (2008). After-school program effects on behavior: Results from the 21st Century learning centers program national evaluation. *Economic Inquiry*, *46*(1), 13-18.

Johnson, L.J., Zorn, D., Williams, J., & Smith, J. (1999). *1998-1999 School Year Program Evaluation: Urban School Initiative School Age Child Care Expansion.* Cincinnati, OH: University of Cincinnati.

Katz, M. (1997). *On Playing a Poor Hand Well: Insights from the lives of those who have overcome childhood risks and adversities.* New York: Norton.

Lauer, P.A., Akiba, M., Wilkerson, S.B., Apthorp, H.S., Snow, D., Martin-Glenn, M. (2004). The effectiveness of Out-of-School-time strategies in assisting low-achieving students in reading and mathematics: A Research synthesis. Prepared for the Institute of Education Sciences Department of Education, Washington, D.C.

Laughlin, L. (2014). A Child's Day: Living Arrangements, Nativity, and Family Transitions: 2011, Current Population Reports, 70-139, Washington, DC: U.S. Census Bureau.

Leos-Urbel, J. (2015). What works after school? The relationship between after-school program quality, program attendance, and academic outcomes. *Youth & Society, 47*(5), 684-706.

Lord, H. & Mahoney, J.L. (2007). Neighborhood crime and self-care: Risks for aggression and lower academic performance. *Developmental Psychology, 43*(6), 1321-1333.

Luke, D.A. (2004). *Multilevel Modeling.* Thousand Oaks: Sage Publications.

Mahoney, J.L., Lord, H. & Carryl, E. (2005). An ecological analysis of after-school program participation and the development of academic performance and motivational attributes for disadvantaged children. *Child Development*, *76*(4), 811-825.

Marsh, H.W. (1992). Extracurricular activities: Beneficial extension of the traditional curriculum or subversion of academic goals?, *Journal of Educational Psychology (84)*, 553-562.

Marshall, N.L., Garcia Coll, C., Marx, F., McCartney, K, Keefe, N. & Ruh, J. (1997). After-School time and children's behavioral adjustment. *Merrill-Palmer Quarterly, 43*(3), 497-514.

Masten, A.S. (2007). Resilience in developing systems: Progress and promise as the fourth wave rises. *Development and Psychopathology, 19*, 921-930.

Masten, A..S. & Coatsworth, J.D. (1998). The development of competence in favorable and unfavorable environments, *American Psychologist*, *53*(2), 205-220.

Masten,A.S.,&Motti-Stefanidi,F. (2009). Understanding and promoting resilience in children: Promotive and protective processes in schools. In T. Gutkin & C. Reynolds (Eds.), *Handbook of School Psychology*. Hoboken, NJ: Wiley.

McHale, S.M., Crouter, A.C. & Tucker, C.J. (2001). Free-time activities in middle childhood: Links with adjustment in early adolescence. *Child Development*, *72*(6), 1764-1779.

Miller, B.M. (2003). *Critical Hours: Afterschool Programs and Educational Success.* Nellie Mae Education Foundation.

Moore, K.A. (2006). *Child Trends: Research to Results (12).* Washington DC: The Atlantic Philanthropies.

National Center for Education Statistics (NCES). (2011a). *The Nations' Report Card: Math 2011: National Assessment of Educational Progress at grades 4 & 8.* Institute of Educational Sciences, U.S. Department of Education. Retrieved from nces.ed.gov.

National Center for Education Statistics (NCES). (2011b). *The Nations' Report Card: Reading 2011: National Assessment of Educational Progress at grades 4 & 8.* Institute of Educational Sciences, U.S. Department of Education. Retrieved from nces.ed.gov.

New Hampshire Department of Education (NH DOE). (2011). Continuous improvement process for after school (CIPAS) Visitation team report: 21st Century Extended Day Program, Nashua School District, Nashua, NH: Belle-Isle, H., Gaspa, K., Kerr-Vanderslice, J., Nowacki, C. National Center for Educational Accountability: November 15-17, 2011.

New Hampshire Department of Education (NH DOE). (2013). *21st Century Community Learning Centers: Program Information*. Retrieved from http://www.education.nh.gov/instruction/integrated/title_iv_b_program.htm.

New Hampshire Department of Education (NH DOE). (2015). *New England Common Assessment Program.* Retrieved from http://education.nh.gov/instruction/assessment/necap/index.htm.

Pettit, G.S., Laird, R.D., Bates, J.E. & Dodge, K.A. (1997). Patterns of after-school care in middle childhood: Risk factors and developmental outcomes. *Merrill-Palmer Quarterly*, *43*(3), 515-538.

Pierce, K.M., Bolt, D.M., Vandell, D.L. (2010). Specific features of after-school program quality: Associations with children's functioning in middle childhood. *American Journal of Community Psychology*, 45, 381-393.

Policy Studies Associates. (2001). Building Quality and Supporting Expansion of After-school Projects: Evaluation results from the TASC After-school program's Second Year. Washington DC.

Posner, J.K. & Vandell, D.L. (1994). Low-income children's after-school care: Are there beneficial effects of after-school programs? *Child Development*, *65*(2), 440-456.

Posner, J.K. & Vandell, D.L. (1999). After-school activities and the development of low-income urban children: A longitudinal study. *Developmental Psychology, 35*(3), 868-879.

Roosa, M.W. (2000). Some thoughts about resilience versus positive development. *Child Development, 71*(3), 567-569.

Roth, J.L., Malone, L.M. & Brooks-Gunn, J. (2010). Does the amount of participation in afterschool programs relate to developmental outcomes? A review of the literature. *American Journal of Community Psychology, 45*, 310-324.

Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688-701.

Russell, C.A. & Woods, Y. (2012). Evaluation of the New Hampshire 21st Century community learning centers: Findings from the 2011-2012 school year. Prepared by Policy Studies Associates, Inc., Washington, D.C.: October 2012.

Rutter, M. (1979). Protective factors in children's responses to stress and disadvantage. In M.W. Kent & I.E. Rolf (Eds.) *Primary Prevention of Psychopathology: Social Competence in Children* (pp. 49-74). Hanover, NH: University Press of New England.

Sanger, C. (2011). After-school enrichment: Extending learning time. *Leadership, 40*(5), 36-38.

Schinke, S.P., Cole, K.C. & Poulin, S.R. (2000). Enhancing the educational achievement of at-risk youth. *Prevention Science, 1*(1), 51-60.

Schorr, L.B. (1989). *Within Our Reach: Breaking the Cycle of Disadvantage.* NY: DoubleDay.

Scott-Little, C., Hamann, M.S. & Jurs, S.G. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, *23*(4), 387-419.

Shadish, W.R., Clark, M.H., Steiner, P.M., Hill, J. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association, 103*, 1334-1350.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Wadsworth Cengage Learning.

Sheldon, J., Arbreton, A., Hopkins, L., Grossman, J.B. (2010). Investing in success: Key strategies for building quality in after-school programs. *American Journal of Community Psychology*, *45*, 394-404.

Shernoff, D.J. (2010). Engagement in after-school programs as a predictor of social competence and academic performance. *American Journal of Community Psychology, 45*, 325-337.

Shonkoff, J.P. & Phillips, D.A. (Eds.) (2000). *From Neurons to Neighborhoods: The Science of Early Child Development*. National Academy Press: Washington, D.C.

Shumow, L. & Posner, J. (2005). *After School Programs for Low-Income Children: Differences in Program Quality.* Psychology Press.

Sirin, S.R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453.

Smokowski, P.R. (1998). Prevention and intervention strategies for promoting resilience in disadvantaged children. *Social Service Review*, *72*(3), 337-364.

Thoemmes, F. (2012). Propensity score matching in SPSS. Eprint arXiv.1201.6385. January 2012.

Traub, J. (2000). What no school can do. *New York Times Magazine.* January 16.

U.S. Department of Education (1992). *Characteristics of At-Risk Students.* Washington, D.C.: National Center for Education Statistics, August.

U.S. Department of Education (2003). *21st Century Community Learning Centers: Non-Regulatory Guidance.* Office of Elementary and Secondary Education Academic Improvement and Teacher Quality Programs, February 2003.

U.S. Department of Education (2015). *21st Century Community Learning Centers: Funding Status*. Office of Elementary and Secondary Education, retrieved from https://www2.ed.gov/programs/21stcclc/funding.html.

Vandell, D.L. & Corasaniti, M.A. (1988). The relation between third graders' after-school care and social, academic, and emotional functioning. *Child Development*, *59*(4), 868-876.

Vandell, D.L. & Pierce, K.M. (1999). *Can after-school programs benefit children who live in high-crime neighborhoods?* Paper presented at the Biennial meeting of the Society for Research in Child Development, Albuquerque, NM.

Vandell, D.L. & Ramanan, J. (1991). Children of the National Longitudinal Survey of Youth: Choices in after-school care and child development. *Developmental Psychology, 27*(4), 637-643.

Vandell, D.L., Reisner, E.R. & Pierce, K.M. (2007). Outcomes linked to high-quality afterschool programs: Longitudinal findings from the study of promising afterschool programs. Report to the Charles Stewart Mott Foundation.

Vandell, D.L., Reisner, E.R., Pierce, K.M., Brown, B.B., Lee, D., Bolt, D. & Pechman, E.M. (2005). The study of promising after-school programs: Examination of intermediate outcomes in year two. Report to the Charles Stewart Mott Foundation.

Vandell, D.L., Reisner, E.R., Pierce, K.M., Brown, B.B., Lee, D., Bolt, D. & Pechman, E.M. (2006). The study of promising after-school programs: Examination of longer term outcomes after two years of program experiences. Report to the Charles Stewart Mott Foundation.

Weissbourd, R. (1997) *The Vulnerable Child: What Really Hurts America's Children and What We Can Do About it.* Reading, MA: Addison-Wesley Publishing Company.

Weitzman, B., Mijanovich, T., Silver, D., & Brazill, C. (2008). If you build it, will they come? Estimating unmet demand for after-school programs in America's distressed cities. *Youth & Society, 40,* 3-34.

Welsh, M.E., Russell, I.W., Reisner, E.R., White, R.N., Winter, N., & Pearson, L (2002). Promoting learning and school attendance through after-school programs: Student level changes in educational performance across TASC's first three years. New York: The After-School Corporation.

Woodland, M.H. (2016). After-school programs: A resource for young black males and other urban youth. *Urban Education, 51*(7), 770-796.

YMCA (2015). *History: The YMCA in the United States.* Retrieved from www.ymca.net/history.

Zolkowski, S.M. & Bullock, L.M. (2012). Resilience in children and youth: A review. *Children & Youth Service Review, 34*, 2295-2303.

APPENDICES

Appendix A

Approval letter from the Institutional Review Board of the University of New Hampshire

University *of* New Hampshire

Research Integrity Services, Service Building
51 College Road, Durham, NH 03824-3585
Fax: 603-862-3564

02-Jun-2016

Scarpati, Jennifer
Education, Morrill Hall
22 Bryant Rd
Nashua, NH 03062

IRB #: 6476
Study: 21st Century After School Programs: Dosage and Academic Outcomes for Elementary School Students
Approval Date: 31-May-2016

The Institutional Review Board for the Protection of Human Subjects in Research (IRB) has reviewed and approved the protocol for your study as Expedited as described in Title 45, Code of Federal Regulations (CFR), Part 46, Subsection 110.

Approval is granted to conduct your study as described in your protocol for one year from the approval date above. At the end of the approval period, you will be asked to submit a report with regard to the involvement of human subjects in this study. If your study is still active, you may request an extension of IRB approval.

Researchers who conduct studies involving human subjects have responsibilities as outlined in the document, *Responsibilities of Directors of Research Studies Involving Human Subjects.* This document is available at http://unh.edu/research/irb-application-resources. Please read this document carefully before commencing your work involving human subjects.

If you have questions or concerns about your study or this approval, please feel free to contact me at 603-862-2003 or Julie.simpson@unh.edu. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

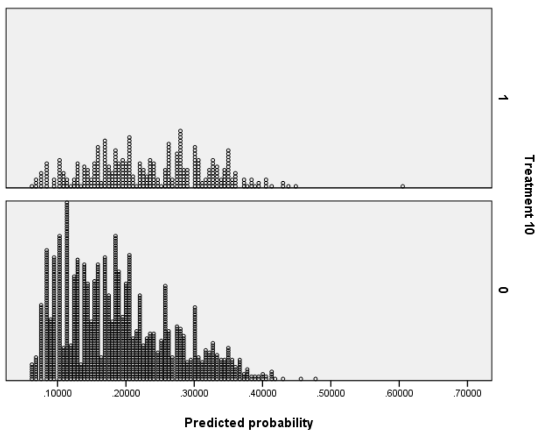
For the IRB,

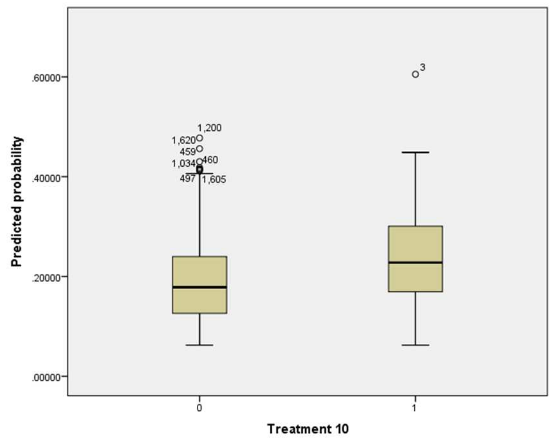
Julie F. Simpson
Director

cc: File
DeMitchell, Todd

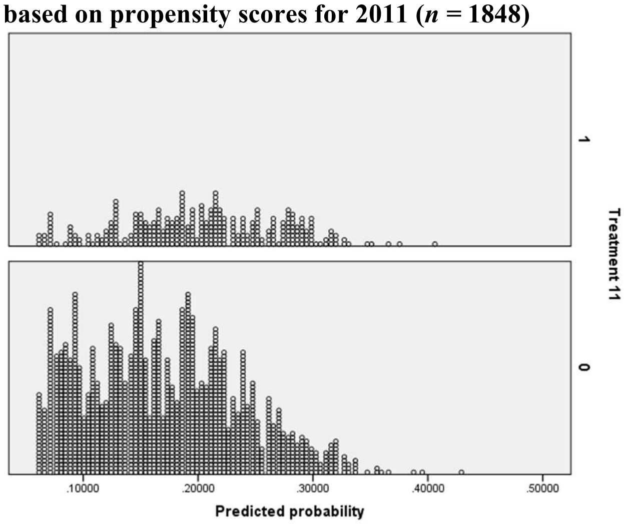
Appendix B

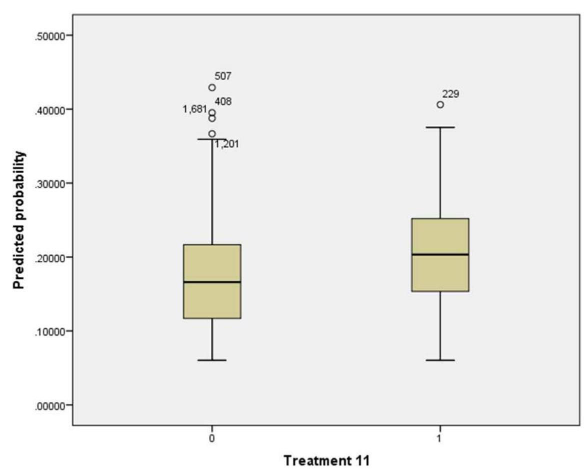
Figures 57 & 58: Dot plots and Box plots demonstrating common support for treatment and control groups based on propensity scores for 2010 (*n* = 1893)



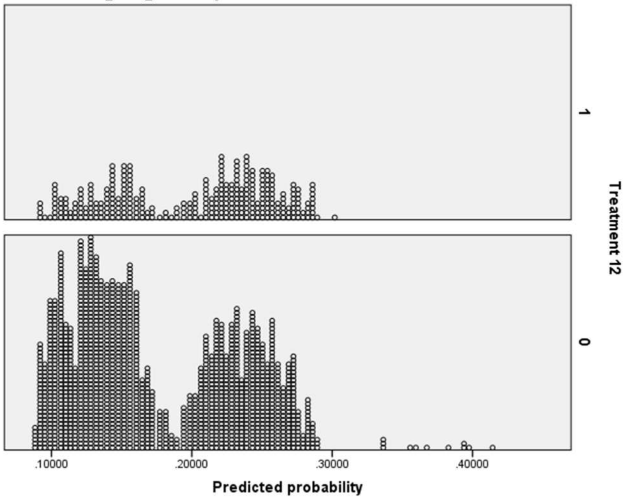


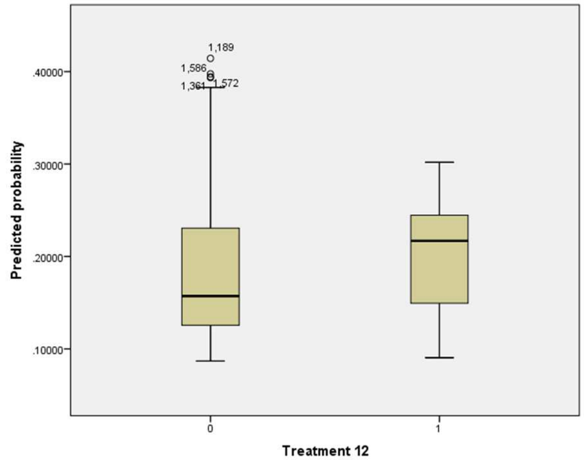
Figures 59 & 60: Dot plots and Box plots demonstrating common support for treatment and control groups based on propensity scores for 2011 (*n* = 1848)



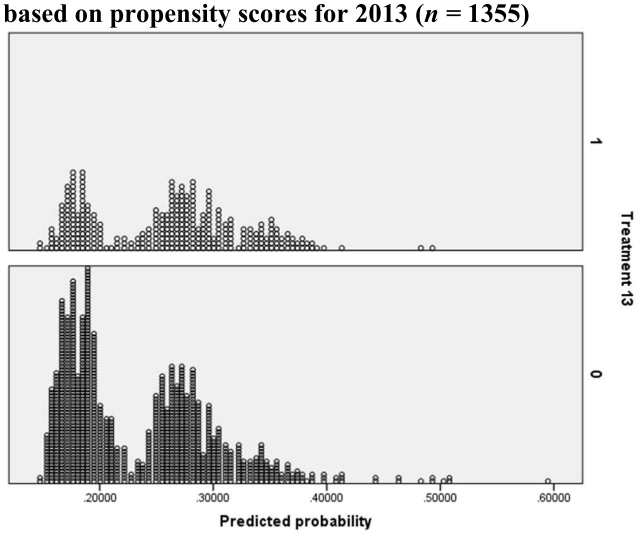


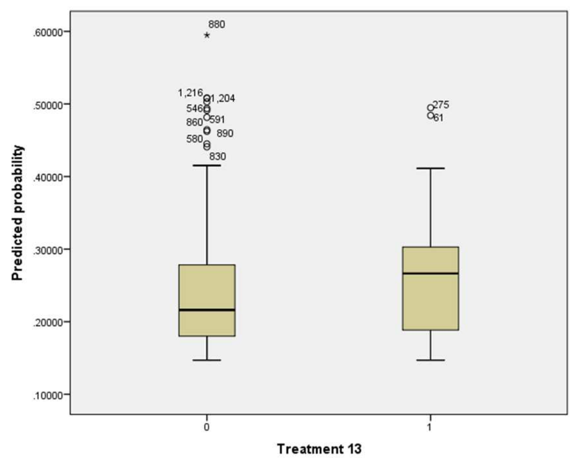
Figures 61 & 62: Dot plots and Box plots demonstrating common support for treatment and control groups based on propensity scores for 2012 (*n* = 1764)





Figures 63 & 64: Dot plots and Box plots demonstrating common support for treatment and control groups based on propensity scores for 2013 (*n* = 1355)





Appendix C

Figure 65 & 66: Box plots and Dot plots demonstrating distribution of propensity scores for the treatment and control groups after weighting for 2010 (*n* = 1893)

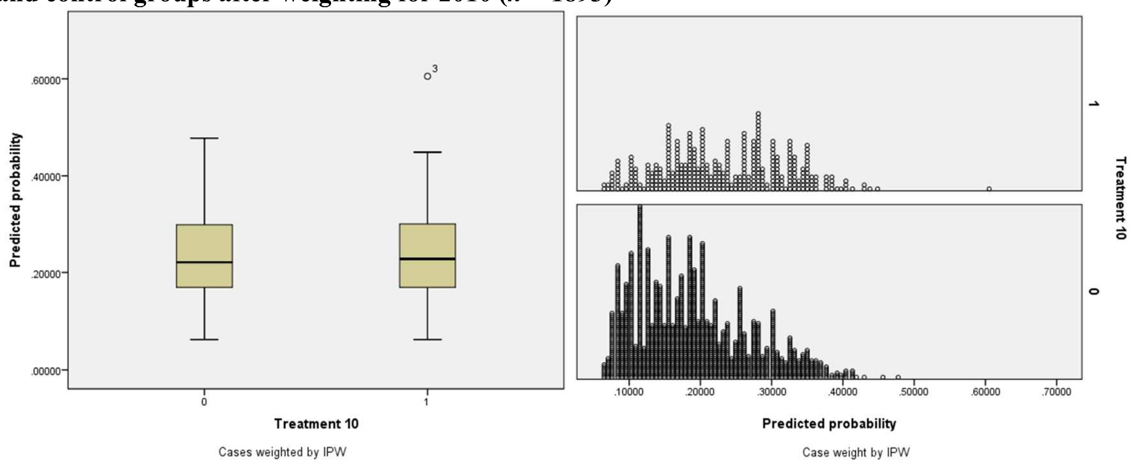


Figure 67 & 68: Box plots and Dot plots demonstrating distribution of propensity scores for the treatment and control groups after weighting for 2011 (*n* = 1848)

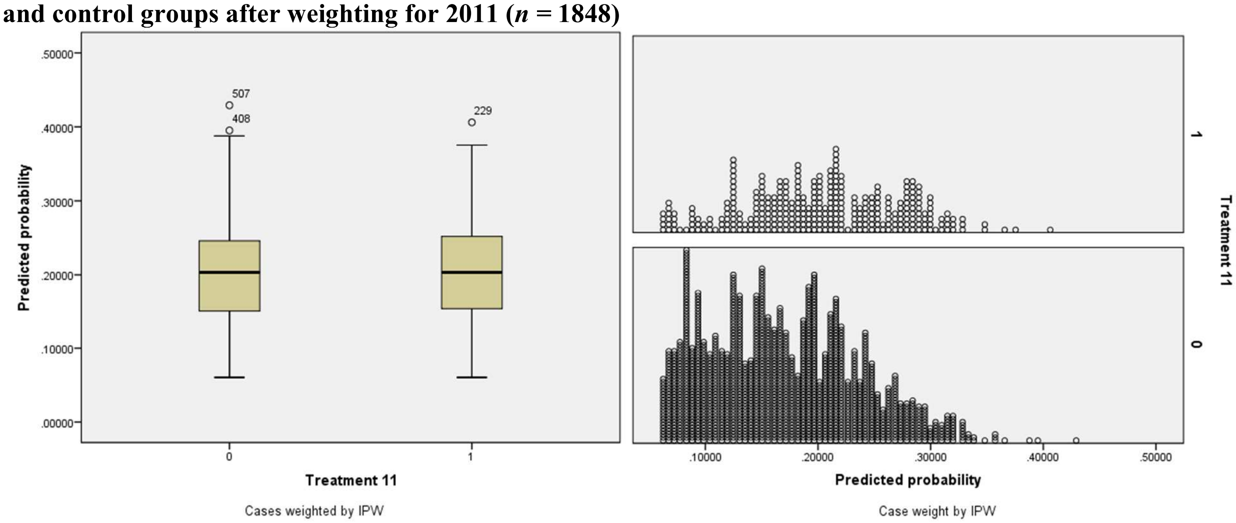


Figure 69 & 70: Box plots and Dot plots demonstrating distribution of propensity scores for the treatment and control groups after weighting for 2012 (*n* = 1764)

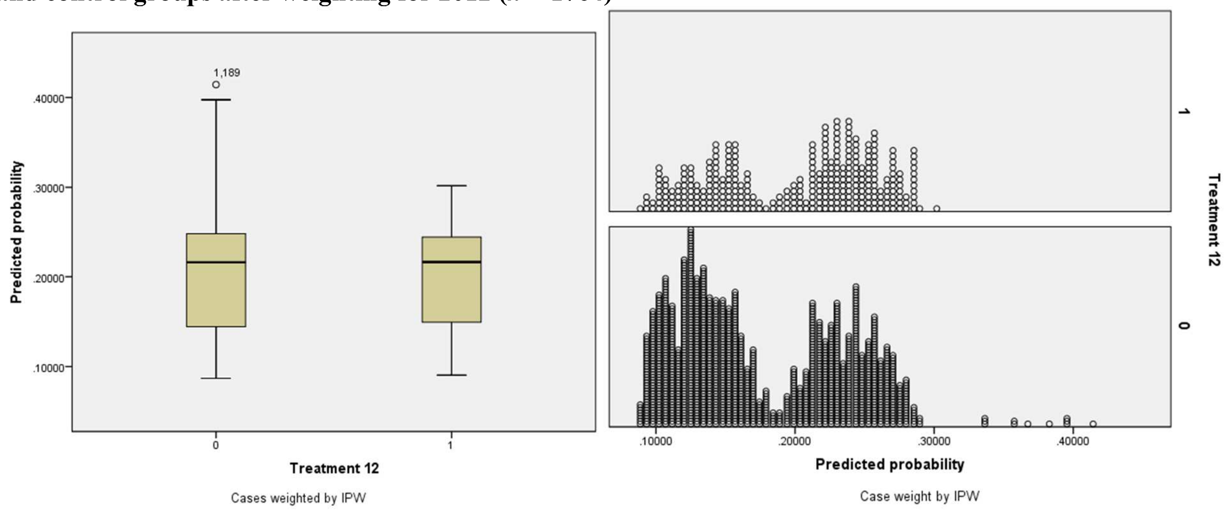
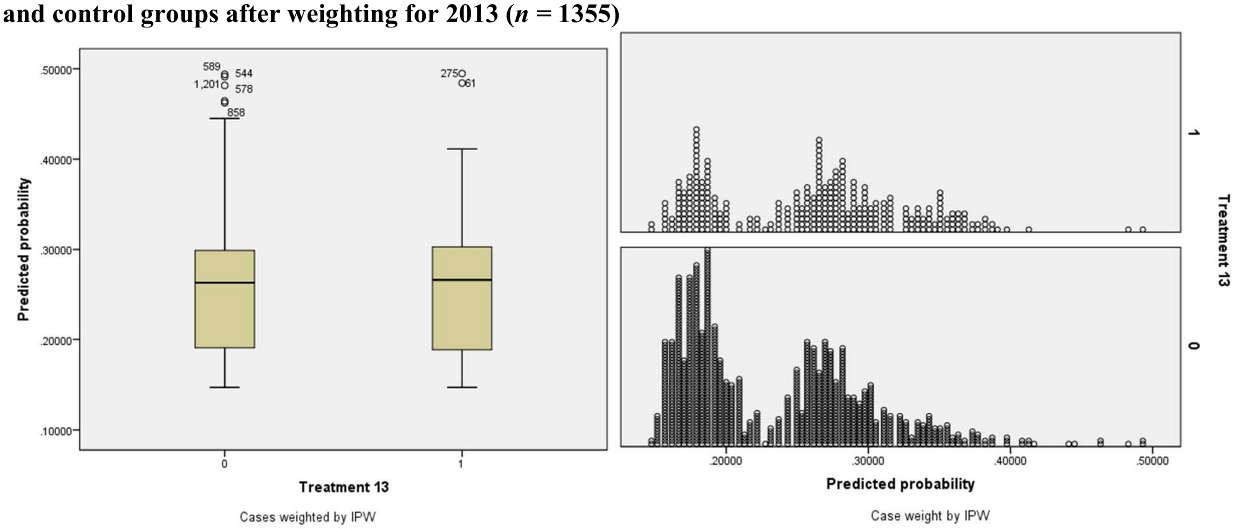


Figure 71 & 72: Box plots and Dot plots demonstrating distribution of propensity scores for the treatment and control groups after weighting for 2013 (*n* = 1355)



Appendix D

Table 51: Taxonomy of fitted models for Math Achievement predicted by student participation in the 21st Century Program – 2009 (n = 1559)

	Model 1: Particip	Model 2: School	Model 3: Grade	Model 4: Demogs	Model 5: Site Quality	Model 6: Demog Interactions	Model 7: Grade Partic	Model 8: Final
Constant	40.77***	40.27***	44.19***	48.49***	48.95***	48.95***	46.22***	46.63***
Particip 09								
est.	-.95	-1.19	-1.39~	-1.27~	-1.97	-2.38	N/A	N/A
School 1								
est.		4.19**	4.31**	3.55**	1.30	3.58**	3.68**	1.50
School 2								
est.		1.88	1.94	1.57	.90	1.57	1.70	1.72
School 4								
est.		-2.41*	-2.44*	-1.98~	-.97	-1.93~	-1.63	-1.64
School 5								
est.		.96	.87	.13	-.75	.19	-.08	-.04
Grade 09								
est.			-.99*	-1.05*	1.11*	-1.03*	-.46	-.49
Male								
est.				1.98**	1.95*	1.42	2.06**	2.01**
Hisp								
est.				-1.25	-1.23	-2.25	-3.98**	-3.93**
AFAM								
est.				-2.37*	-2.37*	-3.05~	-2.34*	-2.36*
FRL								
est.				-3.83***	-3.63***	-3.83**	-3.76***	-3.75***
ELL								
est.				-7.33***	-7.21***	-7.34**	-6.98***	-6.85***
Special Ed								
est.				-6.27***	-6.39***	-6.84***	-6.14***	-6.15***
P1								
est.					4.14~	N/A	N/A	4.05*
P2								
est.					1.39	N/A	N/A	N/A
P4								
est.					-2.25	N/A	N/A	N/A
P5								

est.					1.72	N/A	N/A	N/A
Male*Part								
est.						1.15	N/A	N/A
Hisp*Part								
est.						2.06	N/A	N/A
Afam*Part								
est.						1.29	N/A	N/A
FRL*Part								
est.						-.04	N/A	N/A
ELL*Part								
est.						.32	N/A	N/A
SpEd*Part								
est.						1.15	N/A	N/A
Gr. 2 Part								
est.							4.28**	3.71**
Gr. 3 Part								
est.							-4.02**	-4.53**
Gr. 4 Part								
est.							-1.33	-2.03
Gr. 5 Part								
est.							-3.17*	-3.87**
Summary Statistics								
R^2	0	4	4.6	16.1	17	16.3	19.6	20.1

~ $p <$ .10, * $p <$ .05, ** $p <$ .01, *** $p <$ .001

Table 52: Taxonomy of fitted models for Math Achievement predicted by student participation in the 21st Century Program – 2010 (n = 1876)

	Model 1: Particip	Model 2: School	Model 3: Grade	Model 4: Demogs	Model 5: Site Quality	Model 6: Demog Interactions	Model 7: Grade Partic	Model 8: Final
Constant	40.36***	39.34***	40.41***	49.16***	49.29***	48.32***	49.95***	49.24***
Particip 09								
est.	.50	.23	.22	.34	-.18	2.04	-2.89	-.58
School 1								
est.		3.77**	3.91**	2.88*	2.05	2.72	2.82*	2.98*
School 2								
est.		1.46	1.51	1.25	.31	1.12	1.29	1.39
School 4								
est.		-.38	-.39	.07	1.29	-.01	.13	.17
School 5								
est.		2.28~	2.37~	1.18	-.46	.98	1.13	1.13
Grade 09								
est.			-.36	-1.23***	-1.24***	-1.22***	-1.46***	-1.25***
Male								

est.				1.79*	1.78*	2.09*	1.84*	1.81*
Hisp								
est.				-3.22**	-3.17**	-2.75*	-3.59***	-3.53***
AFAM								
est.				-4.33***	-4.22***	-3.51*	-4.58***	-4.47***
FRL								
est.				-3.87***	-3.74***	-3.18*	-3.90***	-3.86***
ELL								
est.				-5.89***	-5.85***	-4.85**	-5.43***	-5.68***
Special Ed								
est.				-6.64***	-6.61***	-7.86***	-6.57***	-6.57***
P1								
est.					1.62	N/A	N/A	N/A
P2								
est.					1.95	N/A	N/A	N/A
P4								
est.					-2.49	N/A	N/A	N/A
P5								
est.					3.04	N/A	N/A	N/A
Male*Part								
est.						-.37	N/A	N/A
Hisp*Part								
est.						-.82	N/A	N/A
Afam*Part								
est.						-1.58	N/A	N/A
FRL*Part								
est.						-1.39	N/A	N/A
ELL*Part								
est.						-2.36	N/A	N/A
SpEd*Part								
est.						2.51	N/A	N/A
Gr. 2 Part								
est.							2.58	N/A
Gr. 3 Part								
est.							5.88**	3.58**
Gr. 4 Part								
est.							2.69	N/A
Gr. 5 Part								
est.							2.95	N/A
Summary Statistics								
R^2	0	1.9	2.1	17.5	18.3	18	18.8	18.6

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 53: Taxonomy of fitted models for Math Achievement predicted by student participation in the 21st Century Program – 2011 (*n* = 1627)

	Model 1: Particip	Model 2: School	Model 3: Grade	Model 4: Demogs	Model 5: Site Quality	Model 6: Demog Interactions	Model 7: Grade Partic	Model 8: Final
Constant	41.89***	41.66***	43.04***	48.87***	49.47***	47.82***	46.70***	46.98***
Particip 09								
est.	-1.66~	-1.78*	-1.79*	-1.68*	-3.37~	.02	1.39	-.63
School 1								
est.		3.04*	3.00*	1.92	.05	1.89	1.76	2.12
School 2								
est.		-.52	-.64	-.88	-1.89	-.84	-.96	-.83
School 4								
est.		-.75	-.70	-.29	-.15	-.26	-.07	-.18
School 5								
est.		.14	.28	-.38	-2.33	-.30	-.01	-.19
Grade 09								
est.			-.41	-.73*	-.76*	-.71~	-.17	-.23
Male								
est.				.59	.53	1.87	.87	.78
Hisp								
est.				-.77	-.82	-1.42	-1.45	-1.31
AFAM								
est.				-2.37	-2.12	-2.45	-2.68~	-2.66~
FRL								
est.				-3.23**	-3.00*	-2.47	-2.83*	-2.93*
ELL								
est.				-5.54***	-5.37***	-5.34**	-4.97***	-5.35***
Special Ed								
est.				-5.24***	-5.15***	-5.81**	-5.33***	-5.19***
P1								
est.					3.87	N/A	N/A	N/A
P2								
est.					2.22	N/A	N/A	N/A
P4								
est.					-.49	N/A	N/A	N/A
P5								
est.					3.71	N/A	N/A	N/A
Male*Part								
est.						-2.49	N/A	N/A
Hisp*Part								
est.						1.23	N/A	N/A
Afam*Part								
est.						-.04	N/A	N/A
FRL*Part								
est.						-1.32	N/A	N/A

ELL*Part								
est.						-.33	N/A	N/A
SpEd*Part								
est.						1.10	N/A	N/A
Gr. 2 Part								
est.							-4.17	N/A
Gr. 3 Part								
est.							-1.18	N/A
Gr. 4 Part								
est.							-2.56	N/A
Gr. 5 Part								
est.							-6.79*	-4.63**
Summary Statistics								
R^2	.6	2.1	2.4	11.3	12	11.8	13	12.5

~ $p <$ .10, * $p <$ .05, ** $p <$ .01, *** $p <$ .001

Table 54: Taxonomy of fitted models for Math Achievement predicted by student participation in the 21st Century Program – 2012 (n = 1499)

	Model 1: Particip	Model 2: School	Model 3: Grade	Model 4: Demogs	Model 5: Site Quality	Model 6: Demog Interactions	Model 7: Grade Partic	Model 8: Final
Constant	40.49***	38.85***	40.05***	46.14***	46.40***	47.10***	46.23***	46.14***
Particip 09								
est.	1.27	1.09	1.10	1.22	.09	-.67	N/A	1.22
School 1								
est.		4.41*	4.44*	3.12~	.44	3.20~	3.20~	3.12~
School 2								
est.		2.87*	2.70~	2.26~	1.94	2.25~	2.40~	2.26~
School 4								
est.		1.32	1.32	2.26~	1.99	2.30~	2.25~	2.26~
School 5								
est.		1.41	1.37	.64	.16	.67	.70	.64
Grade 09								
est.			-.35	-.61	-.61	-.62	-.55	-.61
Male								
est.				2.54**	2.60**	1.58	2.53**	2.54**
Hisp								
est.				-1.81~	-1.88~	-2.95~	-1.78~	-1.81~
AFAM								
est.				-3.65*	-3.62*	-2.57	-3.62*	-3.65*
FRL								
est.				-3.54**	-3.28**	-3.33~	-3.54**	-3.54**
ELL								
est.				-6.43***	-6.35***	-7.50***	-6.48***	-6.43***
Special Ed								

est.				-6.80***	-6.75***	-6.42**	-6.76***	-6.80***
P1								
est.					5.16~	N/A	N/A	N/A
P2								
est.					.79	N/A	N/A	N/A
P4								
est.					.58	N/A	N/A	N/A
P5								
est.					1.15	N/A	N/A	N/A
Male*Part								
est.						1.93	N/A	N/A
Hisp*Part								
est.						2.22	N/A	N/A
Afam*Part								
est.						-2.04	N/A	N/A
FRL*Part								
est.						-.44	N/A	N/A
ELL*Part								
est.						2.37	N/A	N/A
SpEd*Part								
est.						-1.31	N/A	N/A
Gr. 2 Part								
est.							.89	N/A
Gr. 3 Part								
est.							.74	N/A
Gr. 4 Part								
est.							.14	N/A
Gr. 5 Part								
est.							.57	N/A
Summary Statistics								
R^2	0	1.8	1.9	16.1	16.6	17.1	15.9	16.1

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 55: Taxonomy of fitted models for Math Achievement predicted by student participation in the 21st Century Program – 2013 (n = 1316)

	Model 1: Particip	Model 2: School	Model 3: Grade	Model 4: Demogs	Model 5: Site Quality	Model 6: Demog Interactions	Model 7: Grade Partic	Model 8: Final
Constant	40.29***	38.06***	35.33***	41.52***	41.55***	42.66***	41.26***	42.63***
Particip 09								
est.	.68	.48	.48	.56	.44	-1.70	N/A	-1.79
School 1								
est.		5.76***	5.75***	4.88***	5.49**	4.76**	4.81**	4.82**
School 2								
est.		2.67*	2.77*	2.52*	2.74	2.55*	2.57*	2.56*

School 4								
est.		3.05*	3.08*	3.46**	3.27*	3.45**	3.49**	3.54**
School 5								
est.		1.91	2.04	.97	.11	1.02	.97	1.03
Grade 09								
est.			.77*	.61~	.62~	.62~	.69	.63~
Male								
est.				1.59~	1.58~	1.84	1.57~	1.51~
Hisp								
est.				-1.41	-1.40	-3.38*	-1.41	-3.75**
AFAM								
est.				-3.73**	-3.81**	-3.51*	-3.71**	-3.53**
FRL								
est.				-3.63**	-3.60**	-4.04**	-3.65**	-3.68**
ELL								
est.				-5.28***	-5.31***	-5.79***	-5.23***	-5.17***
Special Ed								
est.				-9.73***	-9.72***	-8.71***	-9.74***	-9.75***
P1								
est.					-1.08	N/A	N/A	N/A
P2								
est.					-.42	N/A	N/A	N/A
P4								
est.					.42	N/A	N/A	N/A
P5								
est.					1.51	N/A	N/A	N/A
Male*Part								
est.						-.68	N/A	N/A
Hisp*Part								
est.						4.08*	N/A	4.71**
Afam*Part								
est.						-.01	N/A	N/A
FRL*Part								
est.						.62	N/A	N/A
ELL*Part								
est.						1.18	N/A	N/A
SpEd*Part								
est.						-2.64	N/A	N/A
Gr. 2 Part								
est.							.50	N/A
Gr. 3 Part								
est.							1.23	N/A
Gr. 4 Part								
est.							-.10	N/A
Gr. 5 Part								
est.							.39	N/A

Summary Statistics								
R^2	0	2.5	3.1	18	18.1	19.2	18	19

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 56: Taxonomy of fitted models for Reading Achievement predicted by student participation in the 21st Century Program – 2009 (n = 1559)

	Model 1: Particip	Model 2: School	Model 3: Grade	Model 4: Demogs	Model 5: Site Quality	Model 6: Demog Interactions	Model 7: Grade Partic	Model 8: Final
Constant	42.76***	43.58***	47.35***	53.56***	53.80***	54.96***	51.83***	54.52***
Particip 09								
est.	-1.40~	-1.55~	-1.54~	-1.30~	-1.79	-4.14*	N/A	-.45
School 1								
est.		.93	1.10	.50	-1.18	.62	.55	.29
School 2								
est.		-.74	-.63	-.71	-1.98	-.67	-.68	-.82
School 4								
est.		-3.31**	-3.33**	-2.99**	-1.59	-2.92**	-2.78*	-2.90**
School 5								
est.		-.16	-.32	-.74	-1.32	-.49	-.83	-.75
Grade 09								
est.			-1.07**	-1.42***	-1.46***	-1.43***	-.93~	-1.68***
Male								
est.				-1.63*	-1.64*	-1.79	-1.60*	-1.61*
Hisp								
est.				-2.69~	-2.72~	-3.64~	-4.04**	-3.56*
AFAM								
est.				.38	.31	-1.16	.46	.44
FRL								
est.				-2.69**	-2.47*	-3.79**	-2.61**	-2.61**
ELL								
est.				-8.42***	-8.27***	-9.42***	-8.32***	-8.59***
Special Ed								
est.				-8.95***	-9.11***	-9.76***	-8.87***	-8.98***
P1								
est.					3.06	N/A	N/A	N/A
P2								
est.					2.64	N/A	N/A	N/A
P4								
est.					-3.01	N/A	N/A	N/A
P5								
est.					1.13	N/A	N/A	N/A
Male*Part								
est.						.31	N/A	N/A

Hisp*Part								
est.						1.83	N/A	N/A
Afam*Part								
est.						3.01	N/A	N/A
FRL*Part								
est.						2.21	N/A	N/A
ELL*Part								
est.						1.56	N/A	N/A
SpEd*Part								
est.						1.69	N/A	N/A
Gr. 2 Part								
est.							2.26	N/A
Gr. 3 Part								
est.							-3.32**	-3.24*
Gr. 4 Part								
est.							-1.27	N/A
Gr. 5 Part								
est.							-2.55~	N/A
Summary Statistics								
R^2	0	2.1	3.3	19	20	19.6	20.4	19.8

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 57: Taxonomy of fitted models for Reading Achievement predicted by student participation in the 21st Century Program – 2010 (*n* = 1876)

	Model 1: Particip	Model 2: School	Model 3: Grade	Model 4: Demogs	Model 5: Site Quality	Model 6: Demog Interactions	Model 7: Grade Partic	Model 8: Final
Constant	42.23***	40.97***	39.54***	49.10***	49.37***	48.80***	49.97***	49.47***
Particip 09								
est.	-.04	-.31	-.29	.05	-.88	.83	N/A	-1.34
School 1								
est.		2.93*	2.72*	1.26	-.25	1.25	1.21	1.45
School 2								
est.		1.67	1.61	1.21	-.13	1.15	1.20	1.30
School 4								
est.		.45	.46	.59	1.57	.55	.64	.71
School 5								
est.		3.18*	3.05*	1.76	.67	1.66	1.71	1.80
Grade 09								
est.			.48	-.37	-.37	-.41	-.63	-.41
Male								
est.				-1.08	-1.09	-.46	-1.04	-.95
Hisp								
est.				-3.03**	-2.99**	-1.87	-3.33**	-3.23**
AFAM								

est.				-1.38	-1.27	-1.76	-1.61	-1.46
FRL								
est.				-3.45**	-3.35**	-3.31*	-3.48**	-3.44**
ELL								
est.				-6.62***	-6.47***	-6.62**	-6.22***	-6.70***
Special Ed								
est.				-7.77***	-7.75***	-9.75***	-7.72***	-9.75***
P1								
est.					2.87	N/A	N/A	N/A
P2								
est.					2.81	N/A	N/A	N/A
P4								
est.					-1.83	N/A	N/A	N/A
P5								
est.					2.17	N/A	N/A	N/A
Male*Part								
est.						-1.01	N/A	N/A
Hisp*Part								
est.						-2.30	N/A	N/A
Afam*Part								
est.						.79	N/A	N/A
FRL*Part								
est.						-.19	N/A	N/A
ELL*Part								
est.						-.24	N/A	N/A
SpEd*Part								
est.						3.77~	N/A	3.99~
Gr. 2 Part								
est.							2.23	N/A
Gr. 3 Part								
est.							4.86*	2.81*
Gr. 4 Part								
est.							2.34	N/A
Gr. 5 Part								
est.							3.10	N/A
Summary Statistics								
R^2	0	1.2	1.6	17.2	17.9	17.9	17.9	18.2

~ $p <$ .10, * $p <$ .05, ** $p <$ .01, *** $p <$ .001

Table 58: Taxonomy of fitted models for Reading Achievement predicted by student participation in the 21st Century Program – 2011 (*n* = 1627)

	Model 1: Particip	Model 2: School	Model 3: Grade	Model 4: Demogs	Model 5: Site Quality	Model 6: Demog Interactions	Model 7: Grade Partic	Model 8: Final
Constant	43.43***	44.46***	44.54***	51.92***	51.59***	50.20***	53.31***	53.31***
Particip 09								
est.	-1.55~	-1.60~	-1.60~	-1.49~	-.81	1.11	-1.17	-1.08
School 1								
est.		1.54	1.54	.25	-.11	.10	-.27	-.31
School 2								
est.		-2.77*	-2.77*	-3.09**	-2.20	-3.08**	-3.43**	-3.39**
School 4								
est.		-2.37~	-2.36~	-2.12~	-1.50	-2.26~	-1.97	-2.09~
School 5								
est.		-.96	-.95	-1.57	-1.52	-1.54	-1.62	-1.49
Grade 09								
est.			-.02	-.35	-.35	-.26	-.72~	-.69~
Male								
est.				-1.00	-1.00	-.11	.80	-.77
Hisp								
est.				-1.48	-1.52	-2.37~	-1.84~	-1.71~
AFAM								
est.				-.75	-.68	-3.43~	-.85	-3.05
FRL								
est.				-3.76**	-3.71**	-1.74	-3.75**	-3.71**
ELL								
est.				-4.46**	-4.38**	-4.06*	-3.93**	-3.90**
Special Ed								
est.				-7.28***	-7.30***	-7.90***	-7.49***	-7.48***
P1								
est.					.53	N/A	N/A	N/A
P2								
est.					-1.80	N/A	N/A	N/A
P4								
est.					-1.51	N/A	N/A	N/A
P5								
est.					-.29	N/A	N/A	N/A
Male*Part								
est.						-1.41	N/A	N/A
Hisp*Part								
est.						1.93	N/A	N/A
Afam*Part								
est.						5.28~	N/A	4.68~
FRL*Part								
est.						-3.88~	N/A	N/A

ELL*Part								
est.						-.88	N/A	N/A
SpEd*Part								
est.						1.27	N/A	N/A
Gr. 2 Part								
est.							-4.92*	-5.44**
Gr. 3 Part								
est.							.58	N/A
Gr. 4 Part								
est.							1.21	N/A
Gr. 5 Part								
est.								
Summary Statistics								
R^2	.6	3	3	16.3	16.5	17.3	17.8	18.2

~ $p <$ .10, * $p <$ .05, ** $p <$ .01, *** $p <$ .001

Table 59: Taxonomy of fitted models for Reading Achievement predicted by student participation in the 21st Century Program – 2012 (n = 1499)

	Model 1: Particip	Model 2: School	Model 3: Grade	Model 4: Demogs	Model 5: Site Quality	Model 6: Demog Interactions	Model 7: Grade Partic	Model 8: Final
Constant	41.98***	40.88***	41.94***	49.31***	49.07***	50.19***	51.13***	50.75***
Particip 09								
est.	.68	.58	.59	.76	1.18	-.75	N/A	.04
School 1								
est.		2.76	2.79	.92	.42	.98	.93	.74
School 2								
est.		2.00	1.85	1.38	1.95	1.32	1.17	1.19
School 4								
est.		1.06	1.05	2.03	2.32	2.07	1.96	1.94
School 5								
est.		.70	.67	-.29	-.05	-.26	-.42	-.53
Grade 09								
est.			-.31	-.64~	-.63~	-.67~	-1.11*	-1.06*
Male								
est.				-.10	-.09	-1.13	-.17	-.17
Hisp								
est.				-1.91~	-1.96~	-2.45	-1.93~	-1.95~
AFAM								
est.				-1.84	-1.88	-.96	-1.89	-1.92
FRL								
est.				-3.67*	-3.03*	-2.92~	-2.88*	-2.94*
ELL								
est.				-7.56***	-7.51***	-7.67***	-7.53***	-7.48***
Special Ed								

est.				-7.67***	-7.65***	-9.67***	-7.69***	-7.71***
P1								
est.					.83	N/A	N/A	N/A
P2								
est.					-1.14	N/A	N/A	N/A
P4								
est.					-.67	N/A	N/A	N/A
P5								
est.					-.49	N/A	N/A	N/A
Male*Part								
est.						2.03	N/A	N/A
Hisp*Part								
est.						1.06	N/A	N/A
Afam*Part								
est.						-1.52	N/A	N/A
FRL*Part								
est.						-.32	N/A	N/A
ELL*Part								
est.						-.07	N/A	N/A
SpEd*Part								
est.						3.87	N/A	N/A
Gr. 2 Part								
est.							-1.05	N/A
Gr. 3 Part								
est.							.03	N/A
Gr. 4 Part								
est.							-.92	N/A
Gr. 5 Part								
est.							3.66*	3.88*
Summary Statistics								
R^2	.1	.7	.8	16.9	17	17.6	17.7	17.6

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 60: Taxonomy of fitted models for Reading Achievement predicted by student participation in the 21st Century Program – 2013 (n = 1316)

	Model 1: Particip	Model 2: School	Model 3: Grade	Model 4: Demogs	Model 5: Site Quality	Model 6: Demog Interactions	Model 7: Grade Partic	Model 8: Final
Constant	42.84***	40.06***	39.60***	47.33***	47.47***	47.72***	47.26***	47.33***
Particip 09								
est.	-.61	-.93	-.93	-.79	-1.17	-1.36	N/A	-.79
School 1								
est.		6.34***	6.34***	4.81***	5.52**	4.67**	4.81***	4.81***
School 2								
est.		4.46***	4.48***	4.15***	3.50*	4.15***	4.16***	4.15***

School 4								
est.		2.71*	2.72*	2.91*	3.05*	2.93*	2.92*	2.91*
School 5								
est.		3.48*	3.50*	2.42~	1.28	2.45~	2.40~	2.42~
Grade 09								
est.			.13	-.12	-.12	-.13	-.10	-.12
Male								
est.				-1.87*	-1.91*	-1.34	-1.88*	-1.87*
Hisp								
est.				-.88	-.89	-2.25~	-.86	-.88
AFAM								
est.				-1.26	-1.36	-.11	-1.24	-1.26
FRL								
est.				-3.88***	-3.79***	-3.78*	-3.89***	-3.88***
ELL								
est.				-4.72***	-4.76***	-5.47**	-4.74***	-4.72***
Special Ed								
est.				-10.12***	-10.18***	-10.35***	-10.12***	-10.12***
P1								
est.					-1.24	N/A	N/A	N/A
P2								
est.					1.32	N/A	N/A	N/A
P4								
est.					-.28	N/A	N/A	N/A
P5								
est.					2.05	N/A	N/A	N/A
Male*Part								
est.						-1.20	N/A	N/A
Hisp*Part								
est.						2.74	N/A	N/A
Afam*Part								
est.						-2.14	N/A	N/A
FRL*Part								
est.						-.35	N/A	N/A
ELL*Part								
est.						1.58	N/A	N/A
SpEd*Part								
est.						.54	N/A	N/A
Gr. 2 Part								
est.							-.92	N/A
Gr. 3 Part								
est.							-.52	N/A
Gr. 4 Part								
est.							-.76	N/A
Gr. 5 Part								

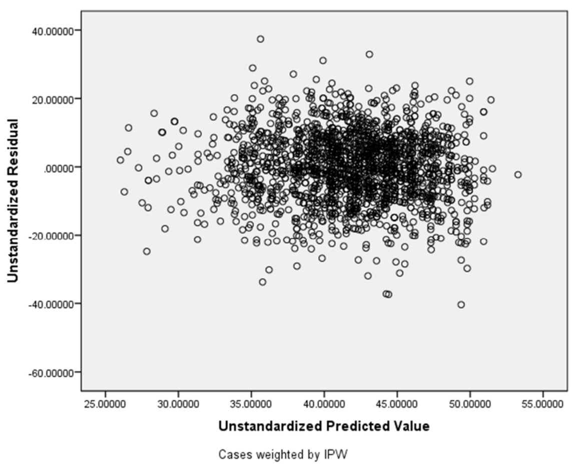
est.							-.96	N/A
Summary Statistics								
R^2	.1	3.6	3.7	19.3	19..6	20.3	19.3	19.3

~ $p <$ .10, * $p <$ .05, ** $p <$ .01, *** $p <$ .001

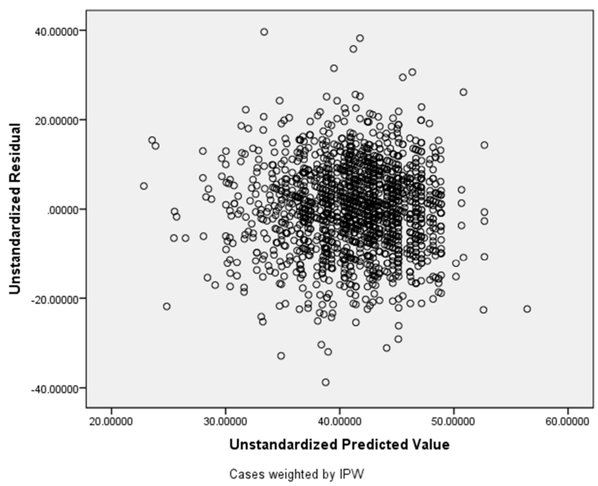
Appendix E

Figures 73 - 80: Bivariate scatterplots of unstandardized residuals versus unstandardized predicted values for effects of participation on Math and Reading achievement (2009-2013)

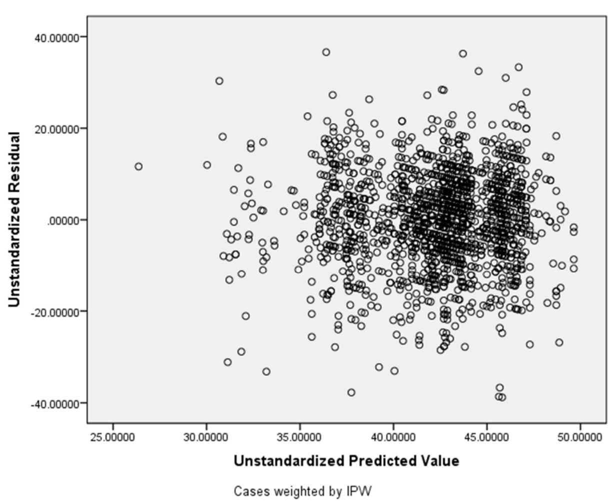
2009 Math



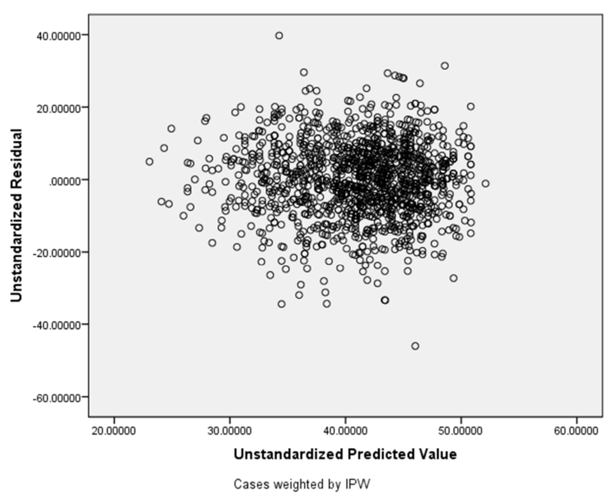
2010 Math



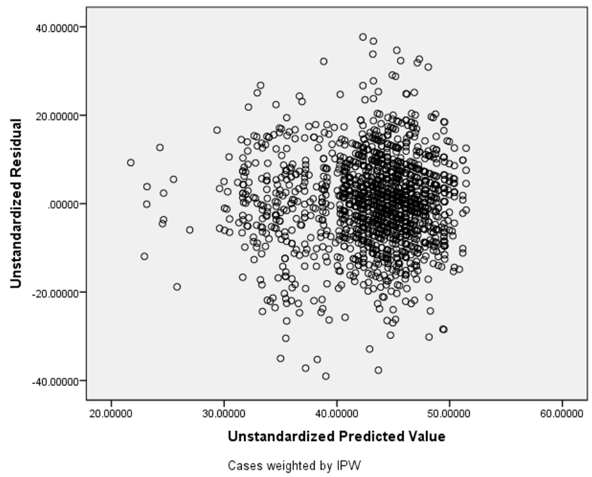
2011 Math



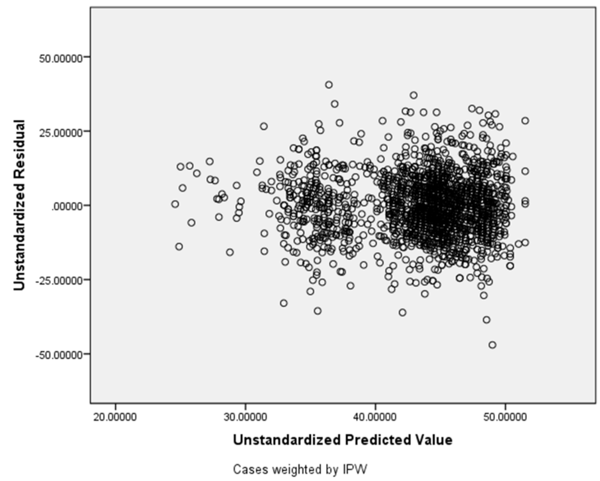
2013 Math



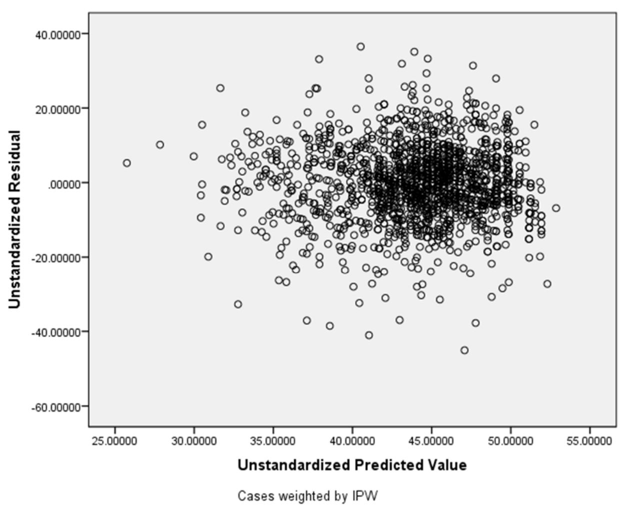
09 Reading



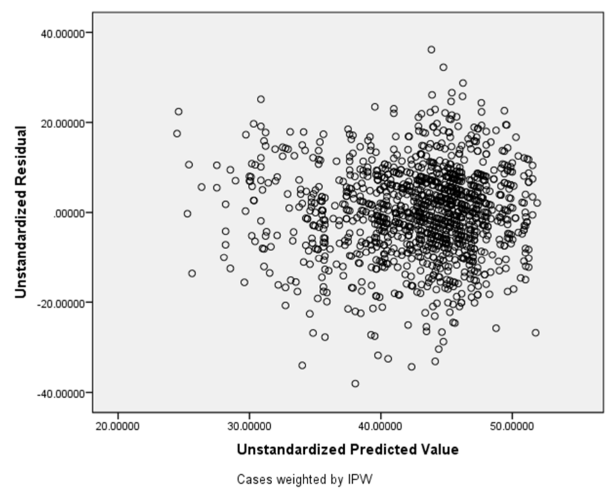
10 Reading



11 Reading



13 Reading



Appendix F

Table 61: Taxonomy of fitted regression models demonstrating impact of 21st Century Program dosage on Math Achievement - 2009 (*n* = 1559)

	Model 1: Particip & Days	**Model 2: Demogr Controls**	**Model 3: Site Qual**	**Model 4: Demog interact**	**Model 5: Grade Interact**	**Model 6: Site Qual* FRL**	**Model 7: Site Qual* ELL**	**Model 8: Site Qual*SP**	**Model 9: Final**
Constant	40.71***	50.16***	50.27***	50.29***	47.77***	48.66***	49.06***	49.37***	49.33***
Particip 09									
est.	-2.34~	-2.32~	-2.11	-2.42~	-2.15~	-2.03	-1.88	-2.15~	-2.04
09 Days									
est.	.02	.02	.01	.01	N/A	.00	.00	.00	-.01
09 Grade									
est.		-1.60***	-1.63***	-1.58***	-.97*	-1.15**	-1.17**	-1.19**	-1.17**
School 1									
est.		3.56**	1.69	3.57**	3.70**	1.40	1.27	1.20	1.22
School 2									
est.		1.65	1.32	1.62	1.75	2.15	1.99~	1.53	1.91~
School 4									
est.		-1.84~	-.59	-1.72	-1.52	-.95	-1.42	-1.39	-1.41
School 5									
est.		-.04	-.25	-.01	-.12	.27	.06	.13	-.03
FRL									
est.		-3.75*	-3.81***	-3.39**	-3.79***	-3.95***	-4.20***	-4.38***	-4.32***
Poverty									
est.		-.25	N/A	N/A	N/A	N/A	N/A	N/A	N/A
African Am									
est.		-2.61*	-2.71*	-3.14*	-2.55*	-2.74*	-2.62*	-2.75*	-2.96**
Hispanic									
est.		-2.99*	-3.18*	-4.70*	-4.38**	-4.55**	-4.81**	-4.48**	-4.36**
Male									
est.		2.01**	1.99**	1.48	2.20**	2.19**	2.16**	2.13**	2.12**
ELL									
est.		-7.43***	-7.17***	-7.01**	-7.15***	-6.89***	-5.67**	-6.78***	-7.50***
Special Ed									
est.		-6.23***	-6.30***	-7.45***	-6.28***	-6.43***	-6.48***	-6.10***	-6.58***
Days Sch1									
est.			.04	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch2									
est.			.01	N/A	N/A	N/A	N/A	N/A	N/A

Days Sch4									
est.			-.04	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch5									
est.			.01	N/A	N/A	N/A	N/A	N/A	N/A
FRL*Days									
est.				-.02	N/A	N/A	N/A	N/A	N/A
HISP*DAYS									
est.				.05	N/A	N/A	N/A	N/A	N/A
AFAM* Days									
est.				.01	N/A	N/A	N/A	N/A	N/A
MALE* DAYS									
est.				.02	N/A	N/A	N/A	N/A	N/A
SPED*DAYS									
est.				.03	N/A	N/A	N/A	N/A	N/A
ELL*Days									
est.				-.01	N/A	N/A	N/A	N/A	N/A
GR2*Days									
est.					.07***	.07**	.06**	.06**	.06**
Gr 3*Days									
est.					-.02	-.02	-.03~	-.03~	-.03~
Gr 4*Days									
est.					.01	N/A	N/A	N/A	N/A
Gr 5*Days									
est.					.00	N/A	N/A	N/A	N/A
FRL*DAYS1									
est.						.06**	.07**	.07**	.07**
FRL*DAYS2									
est.						-.01	N/A	N/A	N/A
FRL*DAYS4									
est.						-.02	N/A	N/A	N/A
FRL* DAYS5									
est.						-.02	N/A	N/A	N/A
ELL* DAYS1									
est.							.08	N/A	N/A
ELL* DAYS2									
est.							-.10	N/A	N/A

ELL* DAYS4									
est.							-.03	N/A	N/A
ELL*DAYS5									
est.							-.07	N/A	N/A
Spd* DAYS1									
est.								-.02	N/A
Spd*DAYS2									
est.								.07	N/A
Spd*DAYS4									
est.								-.03	N/A
Spd*DAYS5									
est.								-.06	N/A
Summary Statistics									
R^2	.4	17.9	18.8	18.5	20.6	21.8	22.1	22.1	22.1

~ $p <$ .10, * $p <$ .05, ** $p <$ .01, *** $p <$ .001

Table 62: Taxonomy of fitted regression models demonstrating impact of 21st Century Program dosage on Math Achievement - 2010 (n = 1876)

	Model 1: Particip & Days	**Model 2: Demogr Controls**	**Model 3: Site Qual**	**Model 4: Demog interact**	**Model 5: Grade Interact**	**Model 6: Site Qual* FRL**	**Model 7: Site Qual* ELL**	**Model 8: Site Qual*SP**	**Model 9: Final**
Constant	40.37***	49.12***	49.06***	48.52***	49.23***	48.60***	48.50***	48.80***	48.80***
Particip 09									
est.	.59	-.01	.34	.26	.36	.32	.28	.48	.31
09 Days									
est.	.00	.00	.00	.02	-.02	.00	.01	.00	.00
09 Grade									
est.		-1.24***	-1.24***	-1.26***	-1.36***	-1.23***	-1.25***	-1.27***	-1.25***
School 1									
est.		2.90*	1.75	2.71*	2.66*	1.95	3.00*	2.58*	2.81*
School 2									
est.		1.23	.90	1.07	1.11	1.88	1.49	1.54	1.22
School 4									
est.		.08	1.53	1.89	1.92	2.06	2.09~	2.05	2.00
School 5									
est.		1.17	.01	.97	.85	1.28	1.21	1.21	.88
FRL									
est.		-4.12**	-3.71***	-3.64**	-3.79***	-3.65**	-3.75***	-3.86***	-3.74***
Poverty									
est.		.27	N/A	N/A	N/A	N/A	N/A	N/A	N/A
African Am									
est.		-4.38***	-4.12**	-4.04*	-4.44***	-4.43***	-4.41***	-4.29***	-4.36***
Hispanic									

est.		-3.26**	-3.25**	-3.06*	-3.58***	-3.52***	-3.54***	-3.47***	-3.55***
Male									
est.		1.79*	1.83*	2.23*	1.90*	1.88*	1.86*	1.84*	1.86*
ELL									
est.		-5.88***	-5.95***	-5.48**	-5.66***	-5.83***	-4.74**	-5.90***	-5.87***
Special Ed									
est.		-6.62***	-6.60***	-7.68***	-6.69***	-6.70***	-6.53***	-6.25***	-6.65***
Days Sch1									
est.			.02	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch2									
est.			.01	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch4									
est.			-.04~	-.05**	-.05**	-.04	-.05*	-.05*	-.05**
Days Sch5									
est.			.02	N/A	N/A	N/A	N/A	N/A	N/A
FRL*Days									
est.				.00	N/A	N/A	N/A	N/A	N/A
HISP*DAYS									
est.				-.01	N/A	N/A	N/A	N/A	N/A
AFAM*Days									
est.				.00	N/A	N/A	N/A	N/A	N/A
MALE*DAYS									
est.				-.01	N/A	N/A	N/A	N/A	N/A
SPED*DAYS									
est.				.03	N/A	N/A	N/A	N/A	N/A
ELL*Days									
est.				-.01	N/A	N/A	N/A	N/A	N/A
GR2*Days									
est.					.03	N/A	N/A	N/A	N/A
Gr 3*Days									
est.					.05*	.03*	.03*	.03**	.03*
Gr 4*Days									
est.					.02	N/A	N/A	N/A	N/A
Gr 5*Days									
est.					.02	N/A	N/A	N/A	N/A
FRL*DAYS1									
est.						.02	N/A	N/A	N/A

FRL*DAYS2									
est.						-.02	N/A	N/A	N/A
FRL*DAYS4									
est.						-.01	N/A	N/A	N/A
FRL* DAYS5									
est.						-.01	N/A	N/A	N/A
ELL* DAYS1									
est.							-.O4	N/A	N/A
ELL* DAYS2									
est.							-.05	N/A	N/A
ELL* DAYS4									
est.							-.02	N/A	N/A
ELL*DAYS5									
est.							-.06	N/A	N/A
Spd* DAYS1									
est.								.05	N/A
Spd*DAYS2									
est.								-.04	N/A
Spd*DAYS4									
est.								.00	N/A
Spd*DAYS5									
est.								-.05	N/A
Summary Statistics									
R^2	.1	17.6	18.5	18.6	19.2	19.4	19.5	19.6	19.0

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 63: Taxonomy of fitted regression models demonstrating impact of 21st Century Program dosage on Math Achievement - 2011 (n = 1627)

	Model 1: Particip & Days	Model 2: Demogr Controls	Model 3: Site Qual	Model 4: Demog interact	Model 5: Grade Interact	Model 6: Site Qual* FRL	Model 7: Site Qual* ELL	Model 8: Site Qual*SP	Model 9: Final
Constant	41.88***	49.00***	49.09***	47.66***	45.82***	46.62***	46.33***	46.77***	46.16***
Particip 09									
est.	-2.05	-1.51	-1.35	-1.37	-1.41	-1.30	-1.34	-1.56	-1.22
09 Days									
est.	.00	.00	-.01	.02	.04~	.03	.03~	.03~	.04*
09 Grade									
est.		-.76*	-.76*	-.75*	-.24	-.42	-.39	-.40	-.41
School 1									
est.		1.93	1.10	1.91	1.92	1.14	2.08	1.95	1.94

School 2									
est.		-.96	-2.19	-.87	-.91	-1.67	-.86	-1.25	-.95
School 4									
est.		-.27	.83	-.13	-.13	.89	-.09	-.57	1.37
School 5									
est.		-.39	-1.02	-.14	.32	.61	.21	.15	-.20
FRL									
est.		-2.94	-3.05~	-2.65~	-2.70*	-2.91*	-2.84*	-2.88*	-2.72*
Poverty									
est.		-.38	N/A	N/A	N/A	N/A	N/A	N/A	N/A
African Am									
est.		-2.30	-1.99	-1.58	-2.64~	-2.22	-2.56~	-2.34~	-2.24
Hispanic									
est.		-.72	-.87	-1.67	-1.40	-1.46	-1.33	-1.28	-1.43
Male									
est.		-.57	.51	2.86*	2.83**	2.75*	2.75*	2.88**	2.75*
ELL									
est.		-5.51***	-5.32***	-5.67**	-4.85***	-4.85***	-4.62**	-5.27***	-5.01***
Special Ed									
est.		-5.40***	-5.33***	-5.88***	-5.68***	-5.45***	-5.43***	-6.50***	-5.56***
Days Sch1									
est.			.02	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch2									
est.			.03	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch4									
est.			-.03	N/A	N/A	N/A	N/A	N/A	-.04*
Days Sch5									
est.			.02	N/A	N/A	N/A	N/A	N/A	N/A
FRL*Days									
est.				-.01	N/A	N/A	N/A	N/A	N/A
HISP*DAYS									
est.				.02	N/A	N/A	N/A	N/A	N/A
AFAM* Days									
est.				-.02	N/A	N/A	N/A	N/A	N/A
MALE* DAYS									
est.				-.05**	-.05**	-.05**	-.05**	-.05**	-.05**
SPED*DAYS									
est.				.01	N/A	N/A	N/A	N/A	N/A
ELL*Days									
est.				.01	N/A	N/A	N/A	N/A	N/A
GR2*Days									

est.					-.02	N/A	N/A	N/A	N/A
Gr 3*Days									
est.					.00	N/A	N/A	N/A	N/A
Gr 4*Days									
est.					-.03	N/A	N/A	N/A	N/A
Gr 5*Days									
est.					-.06*	-.04*	-.04*	-.05*	-.04*
FRL*DAYS1									
est.						.03	N/A	N/A	N/A
FRL*DAYS2									
est.						.02	N/A	N/A	N/A
FRL*DAYS4									
est.						-.03	N/A	N/A	N/A
FRL* DAYS5									
est.						-.02	N/A	N/A	N/A
ELL* DAYS1									
est.							.02	N/A	N/A
ELL* DAYS2									
est.							-.01	N/A	N/A
ELL* DAYS4									
est.							-.02	N/A	N/A
ELL*DAYS5									
est.							-.04	N/A	N/A
Spd* DAYS1									
est.								.03	N/A
Spd*DAYS2									
est.								-.12~	N/A
Spd*DAYS4									
est.								.07	N/A
Spd*DAYS5									
est.								-.01	N/A
Summary Statistics									
R^2	.6	11.5	12.5	13.6	14.5	15.0	14.1	14.7	14.6

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 64: Taxonomy of fitted regression models demonstrating impact of 21st Century Program dosage on Math Achievement - 2012 (*n* = 1499)

	Model 1: Particip & Days	Model 2: Demogr Controls	Model 3: Site Qual	Model 4: Demog interact	Model 5: Grade Interact	Model 6: Site Qual* FRL	Model 7: Site Qual* ELL	Model 8: Site Qual*SP	Model 9: Final
Constant	40.43***	45.33***	45.84***	45.35***	44.85***	45.01***	45.39***	44.87***	45.77***
Particip 09									
est.	-1.50	-1.58	-1.38	-1.63	-1.41	-1.30	-1.66	-1.56	-1.59
09 Days									
est.	.03*	.03*	.01	.03	N/A	.03	.03*	.03*	.02
09 Grade									
est.		-.46	-.47	-.49	-.29	-.30	-.40	-.30	-.49
School 1									
est.		3.07~	1.11	3.16~	3.04~	2.26	3.37*	3.24*	3.27*
School 2									
est.		2.61*	1.32	2.69*	1.99	2.05	2.18	2.13	2.74*
School 4									
est.		2.63*	2.41	2.77*	2.58~	3.12*	2.49~	2.27~	2.65*
School 5									
est.		.95	-.27	.98	.99	.56	1.08	1.29	1.03
FRL									
est.		-4.03*	-3.07*	-2.37	-3.21**	-3.41*	-3.43**	-3.29**	-3.36**
Poverty									
est.		.79	N/A	N/A	N/A	N/A	N/A	N/A	N/A
African Am									
est.		-3.68*	-3.57*	-2.92	-3.75*	-3.53*	-3.71*	-3.57*	-3.70*
Hispanic									
est.		-2.13*	-2.12*	-2,81*	-2.13*	-2.16*	-2.07~	-2.04~	-2.08~
Male									
est.		2.71**	2.75**	2.19~	2.60**	2.59**	2.72**	2.58**	2.70**
ELL									
est.		-6.38***	-6.20***	-7.71***	-6.10***	-5.99***	-6.84***	-6.27***	-8.05***
Special Ed									
est.		-7.32***	-7.33***	-6.90***	-7.30***	-7.26***	-7.44***	-6.98***	-7.52***
Days Sch1									
est.			.04	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch2									
est.			.03	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch4									
est.			.00	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch5									
est.			.03	N/A	N/A	N/A	N/A	N/A	N/A
FRL*Days									
est.				-.02	N/A	N/A	N/A	N/A	N/A

HISP*DAYS									
est.				.01	N/A	N/A	N/A	N/A	N/A
AFAM*Days									
est.				-.02	N/A	N/A	N/A	N/A	N/A
MALE*DAYS									
est.				.01	N/A	N/A	N/A	N/A	N/A
SPED*DAYS									
est.				-.02	N/A	N/A	N/A	N/A	N/A
ELL*Days									
est.				.03	N/A	N/A	N/A	N/A	.04~
GR2*Days									
est.					.01	N/A	N/A	N/A	N/A
Gr 3*Days									
est.					.00	N/A	N/A	N/A	N/A
Gr 4*Days									
est.					.01	N/A	N/A	N/A	N/A
Gr 5*Days									
est.					.02	N/A	N/A	N/A	N/A
FRL*DAYS1									
est.						.02	N/A	N/A	N/A
FRL*DAYS2									
est.						.00	N/A	N/A	N/A
FRL*DAYS4									
est.						-.02	N/A	N/A	N/A
FRL* DAYS5									
est.						.01	N/A	N/A	N/A
ELL* DAYS1									
est.							-.08	N/A	N/A
ELL* DAYS2									
est.							.06~	N/A	N/A
ELL* DAYS4									
est.							.02	N/A	N/A
ELL*DAYS5									
est.							-.01	N/A	N/A
Spd* DAYS1									
est.								-.10	N/A

Spd*DA YS2									
est.								-.02	N/A
Spd*DA YS4									
est.								.07	N/A
Spd*DA YS5									
est.								-.04	N/A
Summary Statistics									
R^2	1.1	17.6	18.1	18.5	17	17.8	18.4	18.5	18

~ $p <$ .10, * $p <$ .05, ** $p <$ .01, *** $p <$ .001

Table 65: Taxonomy of fitted regression models demonstrating impact of 21st Century Program dosage on Math Achievement - 2013 (n = 1316)

	Model 1: Particip & Days	**Model 2: Demogr Controls**	**Model 3: Site Qual**	**Model 4: Demog interact**	**Model 5: Grade Interact**	**Model 6: Site Qual* FRL**	**Model 7: Site Qual* ELL**	**Model 8: Site Qual*SP**	**Model 9: Final**
Constant	40.21***	41.13***	41.60***	42.60***	40.76***	42.42***	42.26***	42.62***	42.79***
Particip 09									
est.	-3.33*	-3.32*	-3.32*	-3.11*	-3.30*	-3.38*	-2.96*	-2.85*	-2.90*
09 Days									
est.	.04**	.04**	.03~	.01	N/A	.03*	N/A	.01	.01
09 Grade									
est.		.69~	.68~	.67~	.79~	.67~	.79~	.64~	.62~
School 1									
est.		5.16***	3.14~	5.18***	5.05***	3.44*	3.80*	4.02*	4.05*
School 2									
est.		2.33~	1.99	2.39*	2.39*	1.32	2.28~	2.64*	2.43*
School 4									
est.		3.60**	2.84~	3.64**	3.63**	2.72*	3.27**	3.74**	3.70**
School 5									
est.		.69	.60	.79	.73	.95	1.41	.83	.82
FRL									
est.		-3.05~	-3.37**	-4.16**	-3.44**	-4.28***	-3.80***	-3.85***	-3.86***
Poverty									
est.		-.43	N/A	N/A	N/A	N/A	N/A	N/A	N/A
African Am									
est.		-3.68**	-3.74**	-3.49*	-3.90**	-3.73**	-3.65**	-3.72**	-3.70**
Hispanic									
est.		-1.64~	-1.70~	-3.29*	-1.70~	-1.66~	-3.76**	-3.76**	-3.78**
Male									
est.		1.57~	1.58~	1.78~	1.58~	1.59~	1.48~	1.35~	1.35~
ELL									
est.		-5.30***	-5.23***	-5.95***	-5.14***	-5.24***	-4.92***	-4.99***	-4.93***
Special Ed									
est.		-9.24***	-9.12***	-9.62***	-9.23***	-9.12***	-9.32***	-8.21***	-8.66***
Days Sch1									

est.			.04~	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch2									
est.			.01	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch4									
est.			.02	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch5									
est.			.01	N/A	N/A	N/A	N/A	N/A	N/A
FRL*Days									
est.				.01	N/A	N/A	N/A	N/A	N/A
HISP*DAYS									
est.				.03*	N/A	N/A	.04**	.04**	.04**
AFAM*Days									
est.				-.01	N/A	N/A	N/A	N/A	N/A
MALE*DAYS									
est.				-.01	N/A	N/A	N/A	N/A	N/A
SPED*DAYS									
est.				.01	N/A	N/A	N/A	N/A	N/A
ELL*Days									
est.				.01	N/A	N/A	N/A	N/A	N/A
GR2*Days									
est.					.04**	N/A	.01	N/A	N/A
Gr 3*Days									
est.					.05**	N/A	.02	N/A	N/A
Gr 4*Days									
est.					.03~	N/A	.00	N/A	N/A
Gr 5*Days									
est.					.04*	N/A	.01	N/A	N/A
FRL*DAYS1									
est.						.05*	.04	.04~	.04~
FRL*DAYS2									
est.						.03	N/A	N/A	N/A
FRL*DAYS4									
est.						.02	N/A	N/A	N/A
FRL* DAYS5									
est.						.00	N/A	N/A	N/A
ELL* DAYS1									

est.							-.05	N/A	N/A
ELL* DAYS2									
est.							.02	N/A	N/A
ELL* DAYS4									
est.							.02	N/A	N/A
ELL*DAYS5									
est.							-.05	N/A	N/A
Spd* DAYS1									
est.								-.32**	N/A
Spd*DAYS2									
est.								-.05	N/A
Spd*DAYS4									
est.								.00	N/A
Spd*DAYS5									
est.								.00	N/A
Summary Statistics									
R^2	1.8	19.8	20.3	21.2	20	20.7	22.3	22.3	22.2

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 66: Taxonomy of fitted regression models demonstrating impact of 21st Century Program dosage on Reading Achievement - 2009 (*n* = 1559)

	Model 1: Particip & Days	**Model 2: Demogr Controls**	**Model 3: Site Qual**	**Model 4: Demog interact**	**Model 5: Grade Interact**	**Model 6: Site Qual* FRL**	**Model 7: Site Qual* ELL**	**Model 8: Site Qual*SP**	**Model 9: Final**
Constant	42.70***	53.58***	53.69***	54.36***	50.96***	51.88***	51.65***	51.71***	51.12***
Particip 09									
est.	-2.63~	-2.65*	-2.43~	-2.72*	-2.56~	-2.51~	-2.36~	-2.48~	-2.64*
09 Days									
est.	.02	.02	.01	.00	N/A	-.01	-.01	-.01	.00
09 Grade									
est.		-1.43***	-1.43***	-1.43***	-.73	-.82*	-.76~	-.77~	-.76~
School 1									
est.		.48	-1.20	.47	.64	-1.48	-1.47	-1.46	.66
School 2									
est.		-.71	-1.63	-.77	-.63	-1.04	-.37	-.75	-.58
School 4									
est.		-2.90**	-1.73	-2.83*	-2.68*	-2.00	-2.71*	-2.54*	-2.69*
School 5									
est.		-.72	-1.91	-.59	-.85	-.76	-.71	-.41	-.90
FRL									
est.		-1.98	-2.41*	-2.77*	-2.60**	-3.00**	-3.03**	-3.12**	-2.65**
Poverty									
est.		-.82	N/A	N/A	N/A	N/A	N/A	N/A	N/A

African Am									
est.		.31	.25	-1.05	.40	.25	.33	.32	.39
Hispanic									
est.		-2.61~	-2.76~	-4.47*	-3.73*	-3.87**	-3.86**	-3.56*	-3.75**
Male									
est.		-1.70*	-1.70*	-2.39*	-1.52*	-1.55*	-1.57*	-1.58*	-1.54*
ELL									
est.		-8.45***	-8.16***	-7.97**	-8.26***	-7.86***	-7.09**	-7.76***	-8.17**
Special Ed									
est.		-8.92***	-8.92***	-9.62***	8.97***	-9.08***	-9.08***	-8.72***	-8.99***
Days Sch1									
est.			.04	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch2									
est.			.02	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch4									
est.			-.04	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch5									
est.			.03	N/A	N/A	N/A	N/A	N/A	N/A
FRL*Days									
est.				.00	N/A	N/A	N/A	N/A	N/A
HISP*DAYS									
est.				.05	N/A	N/A	N/A	N/A	N/A
AFAM* Days									
est.				.03	N/A	N/A	N/A	N/A	N/A
MALE* DAYS									
est.				.02	N/A	N/A	N/A	N/A	N/A
SPED*DAYS									
est.				.02	N/A	N/A	N/A	N/A	N/A
ELL*Days									
est.				-.01	N/A	N/A	N/A	N/A	N/A
GR2*Days									
est.					.07**	.07***	.07***	.07***	.07***
Gr 3*Days									
est.					.00	N/A	N/A	N/A	N/A
Gr 4*Days									
est.					.01	N/A	N/A	N/A	N/A
Gr 5*Days									
est.					.01	N/A	N/A	N/A	N/A

FRL*DAYS1									
est.						.06*	.06**	.06*	N/A
FRL*DAYS2									
est.						.02	N/A	N/A	N/A
FRL*DAYS4									
est.						-.02	N/A	N/A	N/A
FRL*DAYS5									
est.						-.01	N/A	N/A	N/A
ELL*DAYS1									
est.							.05	N/A	N/A
ELL*DAYS2									
est.							-.10	N/A	N/A
ELL*DAYS4									
est.							.00	N/A	N/A
ELL*DAYS5									
est.							-.07	N/A	N/A
Spd*DAYS1									
est.								.03	N/A
Spd*DAYS2									
est.								.05	N/A
Spd*DAYS4									
est.								-.04	N/A
Spd*DAYS5									
est.								-.10~	N/A
Summary Statistics									
R^2	.6	19.2	20.3	19.8	21	21.9	22	22.4	20.8

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 67: Taxonomy of fitted regression models demonstrating impact of 21st Century Program dosage on Reading Achievement - 2010 (n = 1876)

	Model 1: Particip & Days	**Model 2: Demogr Controls**	**Model 3: Site Qual**	**Model 4: Demog interact**	**Model 5: Grade Interact**	**Model 6: Site Qual* FRL**	**Model 7: Site Qual* ELL**	**Model 8: Site Qual*SP**	**Model 9: Final**
Constant	42.22***	49.12***	49.49***	48.88***	49.33***	49.36***	49.26***	49.34***	49.30***
Particip 09									
est.	-.32	-.36	-.15	-.26	-.92	-.24	-.14	-.42	-.38
09 Days									
est.	.00	.00	-.01	.01	N/A	-.01	-.01	-.01	.00
09 Grade									

est.		-.35	-.35	-.36	-.31	-.35	-.39	-.37	-.34
School 1									
est.		1.69	-.20	.02	.16	.27	.33	.29	.31
School 2									
est.		1.35	.41	1.38	1.43	1.24	1.79	1.21	1.51
School 4									
est.		.82	1.42	.76	.87	1.58	1.05	1.22	1.30
School 5									
est.		1.87	1.12	1.96	1.81	2.06	2.07	2.33~	2.34~
FRL									
est.		-3.08~	-3.61***	-3.57**	-3.75***	-3.85**	-3.73***	-3.82***	-4.11***
Poverty									
est.		-.74	N/A	N/A	N/A	N/A	N/A	N/A	N/A
African Am									
est.		-1.74	-1.63	-1.65	-1.94	-1.86	-1.89	-1.86	-1.86
Hispanic									
est.		-2.83**	-2.86**	-2.65*	-3.19**	-3.14**	-3.00**	-3.11**	-3.06**
Male									
est.		-1.24	-1.19	-.20	-1.15	-1.14	-1.17	-1.13	-1.20
ELL									
est.		-6.64***	-6.59***	-6.65***	-6.38***	-6.39***	-5.35***	-6.21***	-6.14***
Special Ed									
est.		-7.96***	-7.99***	-8.96***	-7.98***	-8.05***	-8.01***	-7.57***	-6.81***
Days Sch1									
est.			.04~	.03~	.03~	.01	.03	.03	N/A
Days Sch2									
est.			.02	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch4									
est.			-.01	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch5									
est.			.02	N/A	N/A	N/A	N/A	N/A	N/A
FRL*Days									
est.				.00	N/A	N/A	N/A	N/A	N/A
HISP*DAYS									
est.				-.01	N/A	N/A	N/A	N/A	N/A
AFAM*Days									
est.				.00	N/A	N/A	N/A	N/A	N/A
MALE*DAYS									
est.				-.02	N/A	N/A	N/A	N/A	N/A
SPED*DAYS									
est.				.02	N/A	N/A	N/A	N/A	N/A
ELL*Days									

est.				.00	N/A	N/A	N/A	N/A	N/A
GR2*Days									
est.					.00	N/A	N/A	N/A	N/A
Gr 3*Days									
est.					.03~	.03*	.03*	.03*	.03*
Gr 4*Days									
est.					.00	N/A	N/A	N/A	N/A
Gr 5*Days									
est.					-.01	N/A	N/A	N/A	N/A
FRL*DAYS1									
est.						.05	N/A	N/A	.04~
FRL*DAYS2									
est.						.00	N/A	N/A	N/A
FRL*DAYS4									
est.						-.02	N/A	N/A	N/A
FRL* DAYS5									
est.						-.01	N/A	N/A	N/A
ELL* DAYS1									
est.							.09	N/A	N/A
ELL* DAYS2									
est.							-.08	N/A	N/A
ELL* DAYS4									
est.							-.04	N/A	N/A
ELL*DAYS5									
est.							-.04	N/A	N/A
Spd* DAYS1									
est.								.05	N/A
Spd*DAYS2									
est.								.04	N/A
Spd*DAYS4									
est.								-.06~	-.06*
Spd*DAYS5									
est.								-.07~	-.08*
Summary Statistics									
R^2	0	18.1	18.7	18.7	19.0	19.3	19.7	20.2	20

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 68: Taxonomy of fitted regression models demonstrating impact of 21st Century Program dosage on Reading Achievement - 2011 (*n* = 1627)

	Model 1: Particip & Days	Model 2: Demogr Controls	Model 3: Site Qual	Model 4: Demog interact	Model 5: Grade Interact	Model 6: Site Qual* FRL	Model 7: Site Qual* ELL	Model 8: Site Qual*SP	Model 9: Final
Constant	43.48***	51.56***	50.97***	49.88***	50.70***	49.89***	50.07***	50.02***	50.06***
Particip 09									
est.	-2.23	-1.73	-1.65	-1.42	-1.72	-1.75	-1.88	-1.54	-1.73
09 Days									
est.	.01	.00	.02	.03	.05*	.05**	.05**	.05**	.05**
09 Grade									
est.		-.31	-.30	-.26	-.26	-.56	-.55	-.58	-.56
School 1									
est.		.61	-.06	.49	-.13	-.17	-.40	-.12	-.34
School 2									
est.		-3.04**	-2.14	-3.05**	-3.53**	-2.45~	-2.53~	-2.53~	-2.50~
School 4									
est.		-2.12~	-.69	-2.15~	-2.23~	-.67	-.76	-.71	-.73
School 5									
est.		-1.45	-.73	-1.30	-1.20	.32	.23	.40	.28
FRL									
est.		-2.80~	-3.60**	-2.14	-3.32**	-2.22~	-2.23~	-2.42*	-2.34*
Poverty									
est.		-1.02	N/A	N/A	N/A	N/A	N/A	N/A	N/A
African Am									
est.		-.73	-.64	-2.30	-1.04	-.79	-.68	-.68	-.78
Hispanic									
est.		-1.27	-1.39	-2.35*	-1.81~	-1.57~	-1.59~	-1.41	-1.57~
Male									
est.		-.77	-.76	1.05	.83	.84	.81	.81	.84
ELL									
est.		-4.44**	-4.39**	-4.30**	-3.78**	-3.81**	-4.39**	-3.82**	-3.81**
Special Ed									
est.		-7.40***	-7.43***	-8.33***	-7.57***	-7.73***	-7.75***	-7.25***	-7.70***
Days Sch1									
est.			.01	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch2									
est.			-.03	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch4									
est.			-.04~	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch5									
est.			-.02	N/A	N/A	N/A	N/A	N/A	N/A
FRL*Days									
est.				-.03	N/A	N/A	N/A	N/A	N/A

HISP*DAYS									
est.				.03	N/A	N/A	N/A	N/A	N/A
AFAM* Days									
est.				.03	N/A	N/A	N/A	N/A	N/A
MALE* DAYS									
est.				-.04**	-.03*	-.03*	-.03*	-.03*	-.03*
SPED*DAYS									
est.				.02	N/A	N/A	N/A	N/A	N/A
ELL*Days									
est.				.00	N/A	N/A	N/A	N/A	N/A
GR2*Days									
est.					-.07**	-.05**	-.05**	-.05**	-.05**
Gr 3*Days									
est.					-.02	N/A	N/A	N/A	N/A
Gr 4*Days									
est.					-.03	N/A	N/A	N/A	N/A
Gr 5*Days									
est.					-.05~	N/A	N/A	N/A	N/A
FRL*DAYS1									
est.						-.01	N/A	N/A	N/A
FRL*DAYS2									
est.						-.03~	-.04~	-.03~	-.03~
FRL*DAYS4									
est.						-.05*	-.06**	-.05*	-.05*
FRL* DAYS5									
est.						-.05*	-.05*	-.05*	-.05*
ELL* DAYS1									
est.							-.03	N/A	N/A
ELL* DAYS2									
est.							.02	N/A	N/A
ELL* DAYS4									
est.							.05	N/A	N/A
ELL*DAYS5									
est.							.00	N/A	N/A
Spd* DAYS1									
est.								-.14	N/A

Spd*DA YS2									
est.								.02	N/A
Spd*DA YS4									
est.								-.01	N/A
Spd*DA YS5									
est.								-.03	N/A
Summary Statistics									
R^2	.6	17	17.7	18.7	19.7	20.7	20.9	21	20.7

~ $p <$ .10, * $p <$ .05, ** $p <$ .01, *** $p <$ .001

Table 69: Taxonomy of fitted regression models demonstrating impact of 21st Century Program dosage on Reading Achievement - 2012 (*n* = 1449)

	Model 1: Particip & Days	**Model 2: Demogr Controls**	**Model 3: Site Qual**	**Model 4: Demog interact**	**Model 5: Grade Interact**	**Model 6: Site Qual* FRL**	**Model 7: Site Qual* ELL**	**Model 8: Site Qual*SP**	**Model 9: Final**
Constant	41.96***	49.94***	49.33***	50.01***	51.15***	48.61***	49.75***	50.20***	51.17***
Particip 09									
est.	.04	.30	.39	.20	1.67	.45	.37	.15	.32
09 Days									
est.	.01	.01	.02	.01	N/A	.02	.01	.01	.01
09 Grade									
est.		-.69~	-.67~	-.73~	-1.01*	-.67~	-.68~	-.71~	-1.12**
School 1									
est.		1.20	1.05	1.35	1.21	1.14	1.25	1.24	.90
School 2									
est.		.93	1.90	.88	.62	2.12	.93	.91	.63
School 4									
est.		1.97	3.00~	1.94	1.64	3.28*	1.98	1.51	2.80~
School 5									
est.		-1.14	-.63	-1.20	-1.58	-.77	-.82	-.76	-1.53
FRL									
est.		-3.44~	-3.09*	-2.50	-3.01*	-2.35~	-3.05*	-3.17**	-2.69*
Poverty									
est.		.39	N/A	N/A	N/A	N/A	N/A	N/A	N/A
African Am									
est.		-1.89	-1.75	-1.93	-1.80	-1.70	-1.86	-1.63	-1.34
Hispanic									
est.		-2.36*	-2.43*	-2.07	-2.18*	-2.42*	-2.35*	-2.32*	-2.41*
Male									
est.		-.35	-.30	-.65	-.50	-.34	-.32	-.38	-.55
ELL									
est.		-7.31***	-7.21***	-8.36***	-7.24***	-7.20***	-7.00***	-7.47***	-7.29***
Special Ed									
est.		-7.38***	-7.35***	-9.14***	-7.31***	-7.36***	-7.35***	-7.82***	-8.55***

Days Sch1									
est.			.00	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch2									
est.			-.02	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch4									
est.			-.03	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch5									
est.			-.01	N/A	N/A	N/A	N/A	N/A	N/A
FRL*Days									
est.				-.01	N/A	N/A	N/A	N/A	N/A
HISP*DAYS									
est.				-.01	N/A	N/A	N/A	N/A	N/A
AFAM* Days									
est.				.00	N/A	N/A	N/A	N/A	N/A
MALE* DAYS									
est.				.01	N/A	N/A	N/A	N/A	N/A
SPED*DAYS									
est.				.04	N/A	N/A	N/A	N/A	N/A
ELL*Days									
est.				.02	N/A	N/A	N/A	N/A	N/A
GR2*Days									
est.					-.01	N/A	N/A	N/A	N/A
Gr 3*Days									
est.					-.01	N/A	N/A	N/A	N/A
Gr 4*Days									
est.					-.02	N/A	N/A	N/A	N/A
Gr 5*Days									
est.					.03	N/A	N/A	N/A	.04*
FRL*DAYS1									
est.						.00	N/A	N/A	N/A
FRL*DAYS2									
est.						-.03	N/A	N/A	N/A
FRL*DAYS4									
est.						-.04	N/A	N/A	-.05*
FRL* DAYS5									
est.						-.01	N/A	N/A	N/A

ELL* DAYS1									
est.							N/A	N/A	N/A
ELL* DAYS2									
est.							.00	N/A	N/A
ELL* DAYS4									
est.							.00	N/A	N/A
ELL*DAYS5									
est.							-.01	N/A	N/A
Spd* DAYS1									
est.								-4.20	N/A
Spd*DAYS2									
est.								.01	N/A
Spd*DAYS4									
est.								.10*	.13**
Spd*DAYS5									
est.								-.05	N/A
Summary Statistics									
R^2	.3	18.4	18.7	18.9	19.4	19	18.5	19.7	20.7

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 70: Taxonomy of fitted regression models demonstrating impact of 21st Century Program dosage on Reading Achievement - 2013 (*n* = 1316)

	Model 1: Particip & Days	**Model 2: Demogr Controls**	**Model 3: Site Qual**	**Model 4: Demog interact**	**Model 5: Grade Interact**	**Model 6: Site Qual* FRL**	**Model 7: Site Qual* ELL**	**Model 8: Site Qual*SP**	**Model 9: Final**
Constant	42.66***	46.99***	47.17***	47.56***	47.83***	47.75***	47.50***	46.77***	47.23***
Particip 09									
est.	-4.07**	-3.93**	-3.86**	-3.81**	-3.95**	-3.86**	-3.77**	-3.94**	-4.01**
09 Days									
est.	.04**	.03**	.03~	.02	N/A	.02	.02	.04**	.03**
09 Grade									
est.		-.02	-.05	-.04	-.13	-.09	-.06	-.03	-.06
School 1									
est.		4.74***	4.07*	4.69**	4.70**	4.24**	5.10***	5.10***	4.90***
School 2									
est.		4.04**	3.40*	3.99**	4.02**	3.41*	3.71**	4.30***	3.48**
School 4									
est.		3.05**	3.52*	2.98**	2.99**	3.55**	2.89*	3.08**	3.15**
School 5									
est.		2.29~	2.14	2.26~	2.30~	2.54~	2.73*	2.65*	2.29~
FRL									
est.		-3.85*	-3.64***	-3.83**	-3.83**	-3.84**	-3.85**	-3.74***	-3.72***

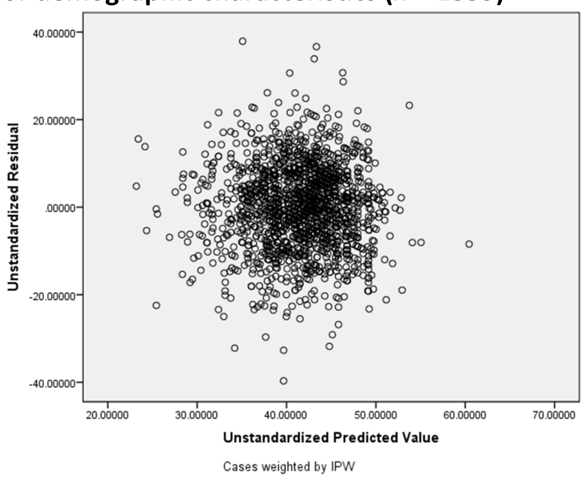
Poverty									
est.		.19	N/A	N/A	N/A	N/A	N/A	N/A	N/A
African Am									
est.		-1.33	-1.25	-.49	-1.38	-1.22	-1.32	-1.34	-1.31
Hispanic									
est.		-1.22	-1.29	-1.56	-1.21	-1.21	-1.16	-1.14	-1.12
Male									
est.		-1.99*	-2.02*	-2.18*	-1.95*	-2.07*	-1.99*	-1.99*	-1.91*
ELL									
est.		-4.70***	-4.56***	-5.94***	-6.36***	-6.37***	-6.11***	-4.93***	-5.35***
Special Ed									
est.		-10.08***	-10.08***	-9.78***	-10.02***	-10.06***	-10.07***	-8.71***	-9.71***
Days Sch1									
est.			.01	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch2									
est.			.01	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch4									
est.			-.01	N/A	N/A	N/A	N/A	N/A	N/A
Days Sch5									
est.			.00	N/A	N/A	N/A	N/A	N/A	N/A
FRL*Days									
est.				.00	N/A	N/A	N/A	N/A	N/A
HISP*DAYS									
est.				.01	N/A	N/A	N/A	N/A	N/A
AFAM* Days									
est.				-.02	N/A	N/A	N/A	N/A	N/A
MALE* DAYS									
est.				.00	N/A	N/A	N/A	N/A	N/A
SPED*DAYS									
est.				-.01	N/A	N/A	N/A	N/A	N/A
ELL*Days									
est.				.00	N/A	N/A	N/A	N/A	N/A
GR2*Days									
est.					.03*	N/A	N/A	N/A	N/A
Gr 3*Days									
est.					.03*	N/A	N/A	N/A	N/A
Gr 4*Days									
est.					.03~	N/A	N/A	N/A	N/A

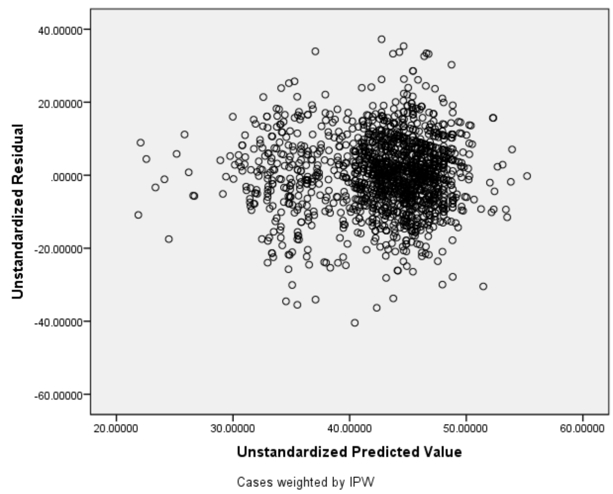
Gr 5*Days									
est.					.04*	N/A	N/A	N/A	N/A
FRL*DAYS1									
est.						.02	N/A	N/A	N/A
FRL*DAYS2									
est.						.01	N/A	N/A	N/A
FRL*DAYS4									
est.						-.02	N/A	N/A	N/A
FRL*DAYS5									
est.						-.01	N/A	N/A	N/A
ELL*DAYS1									
est.							-.08	N/A	N/A
ELL*DAYS2									
est.							.05*	N/A	.05*
ELL*DAYS4									
est.							.03	N/A	N/A
ELL*DAYS5									
est.							-.02	N/A	N/A
Spd*DAYS1									
est.								-.12*	N/A
Spd*DAYS2									
est.								-.06	N/A
Spd*DAYS4									
est.								.06	N/A
Spd*DAYS5									
est.								-.06	N/A
Summary Statistics									
R^2	1.4	20.7	20.9	21.3	20.7	20.9	21.6	21.7	21

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

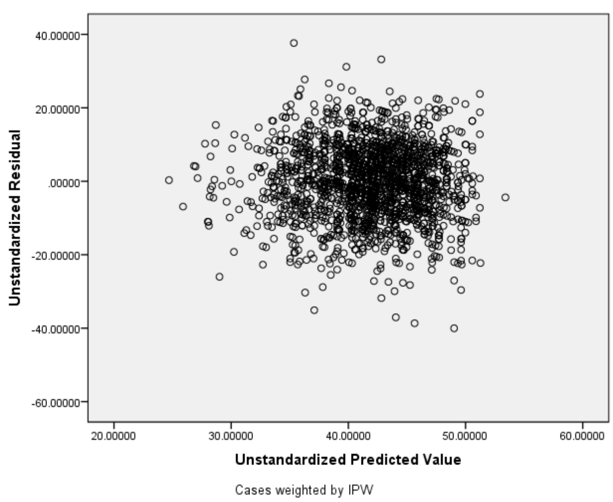
Appendix G

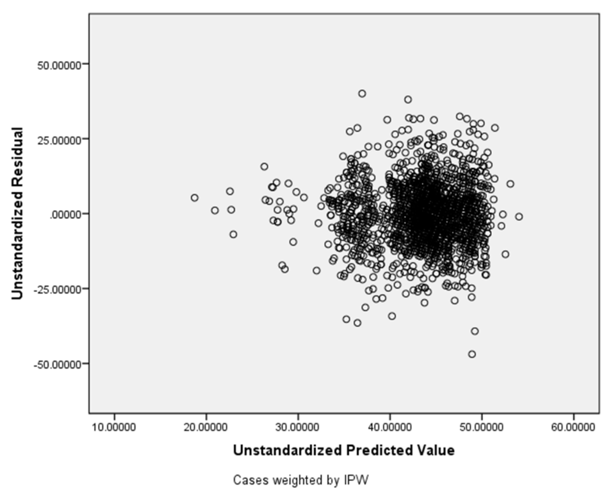
Figures 81 - 82: Raw residuals from final fitted regression models (2009) in which Math(81) and Reading (82) achievement is predicted by dosage in the 21st Century after school program, controlling for demographic characteristics (*n* = 1559)



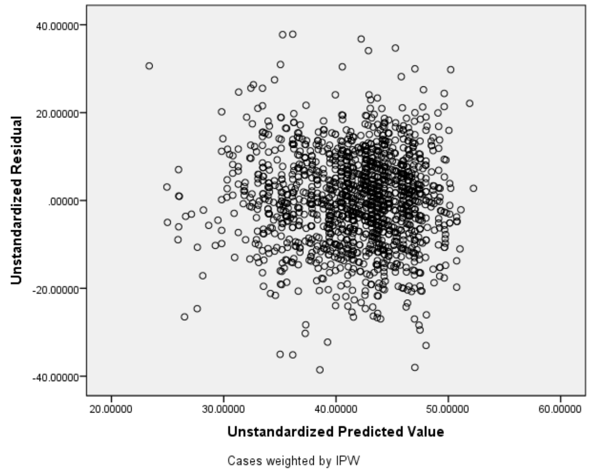


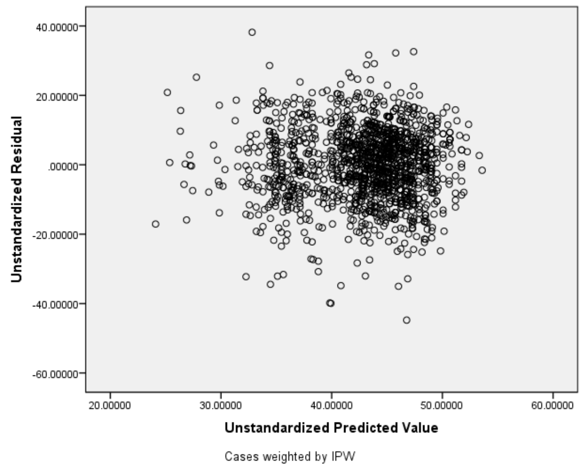
Figures 83 & 84: Raw residuals from final fitted regression models (2010) in which Math (83) and Reading (84) achievement is predicted by dosage in the 21st Century after school program, controlling for demographic characteristics (*n* = 1876)



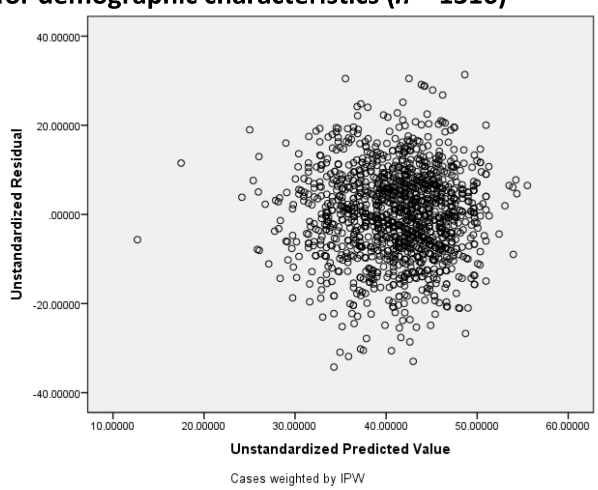


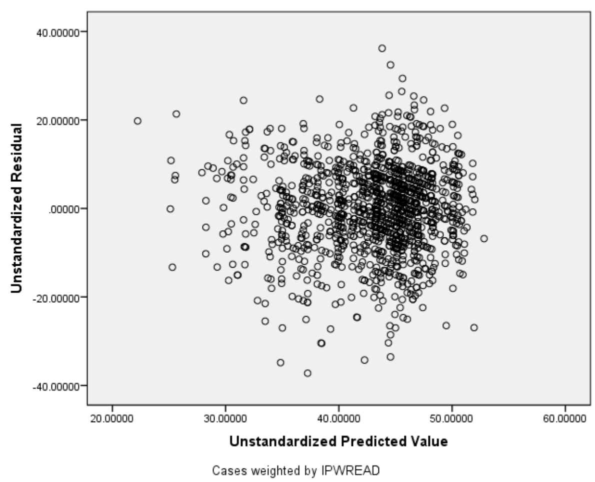
Figures 85 & 86: Raw residuals from final fitted regression models (2012) in which Math (85) and Reading (86) achievement is predicted by dosage in the 21st Century after school program, controlling for demographic characteristics (*n* = 1499)





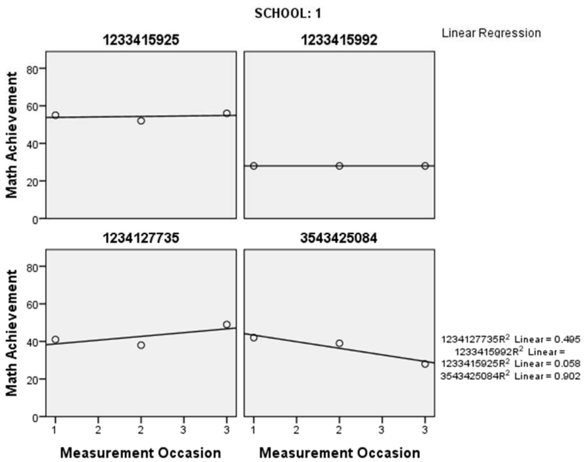
Figures 87 & 88: Raw residuals from final fitted regression models (2013) in which Math (87) and Reading (88) achievement is predicted by dosage in the 21st Century after school program, controlling for demographic characteristics (*n* = 1316)

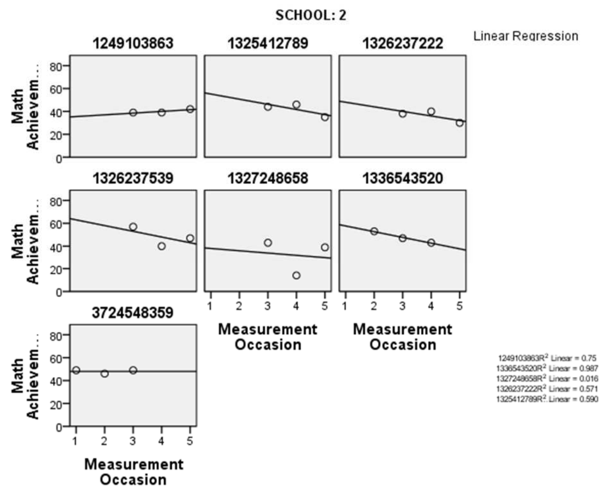


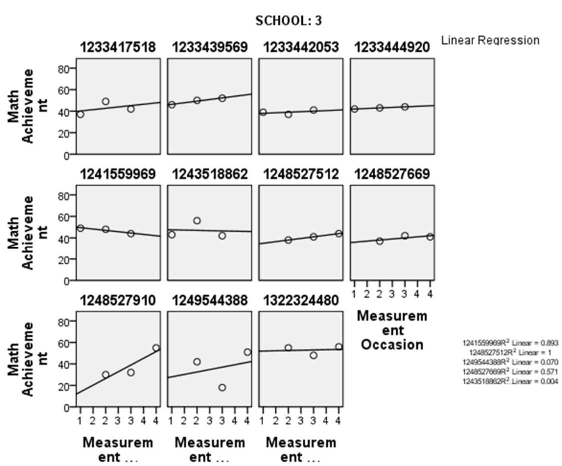


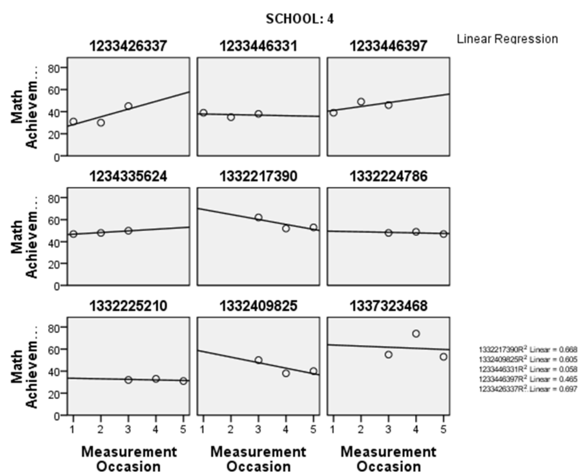
Appendix H

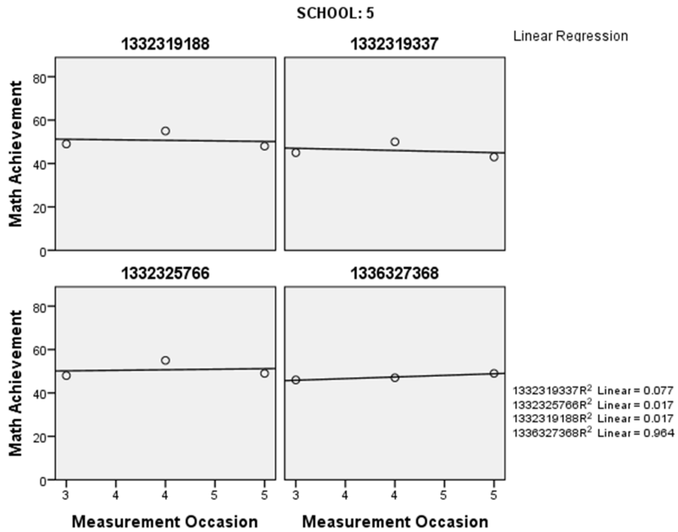
Figures 89 - 93: Fitted growth trajectories by school for Math achievement (*n* = 367)



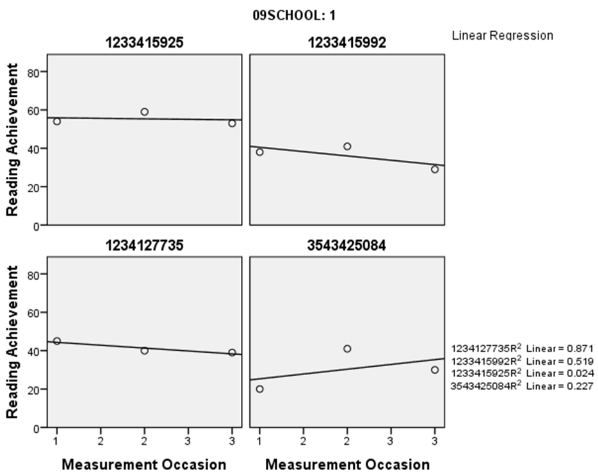


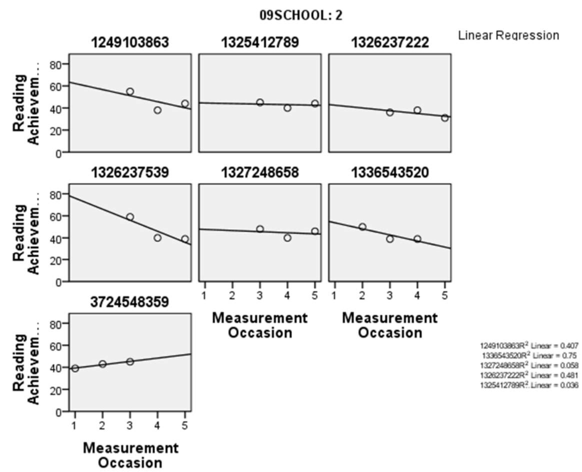


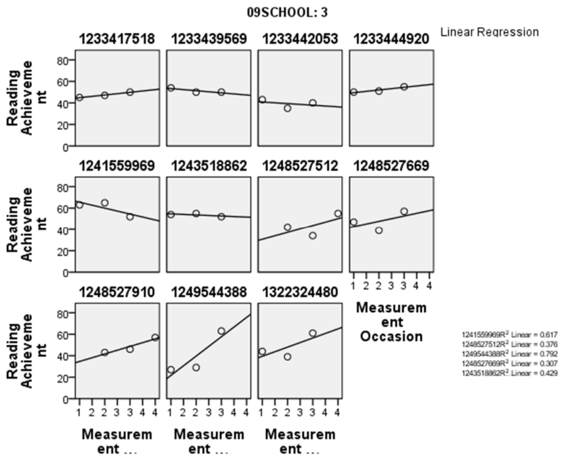


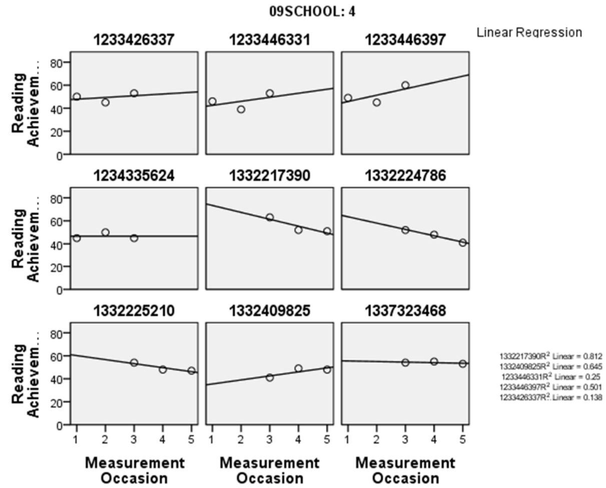


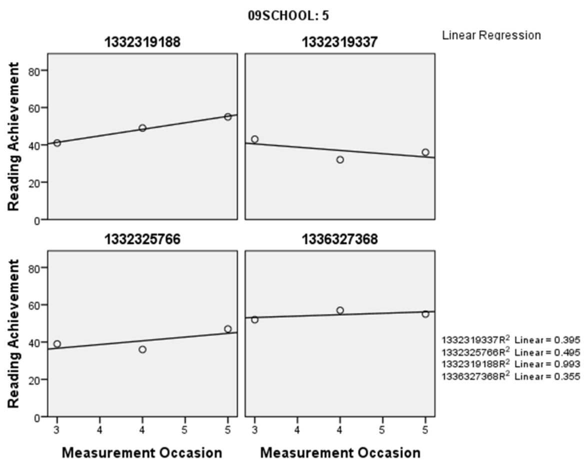
Figures 94 - 98: Fitted growth trajectories by school for Reading achievement (*n* = 353)






09SCHOOL: 3
Linear Regression
1233417518
1233439569
1233442053
1233444920
1241559969
1243518862
1248527512
1248527669
1248527910
1249544388
1322324480
Reading Achievement
Measurement Occasion
1241559969R2 Linear = 0.617
1248527512R2 Linear = 0.376
1249544388R2 Linear = 0.792
1248527669R2 Linear = 0.307
1243518862R2 Linear = 0.429


09SCHOOL: 4
Linear Regression
1233426337
1233446331
1233446397
1234335624
1332217390
1332224786
1332225210
1332409825
1337323468
Reading Achievem...
Measurement Occasion
1332217390R2 Linear = 0.812
1332409825R2 Linear = 0.645
1233446331R2 Linear = 0.25
1233446397R2 Linear = 0.501
1233426337R2 Linear = 0.138


09SCHOOL: 5
Linear Regression
1332319188
1332319337
1332325766
1336327368
Reading Achievement
Measurement Occasion
1332319337R2 Linear = 0.395
1332325766R2 Linear = 0.495
1332319188R2 Linear = 0.993
1336327368R2 Linear = 0.355

Appendix I

Table 71: Taxonomy of fitted regression models for Math achievement over time, 2009-2013 (*n* = 367)

	Model 1: Uncondit	Model 2: Partic	Model 3: School	Model 4: Demogs	Model 5: Grade	Model 6: Dosage	Model 7: School* Days	Model 8: FRL days	Model 9: Male days	Model 10: Final
Intercept	39.26***	40.02**	35.69***	38.65***	38.89***	38.58***	42.37***	40.51***	40.30***	39.39***
Wave_C	.53~	.54	3.09***	3.94***	3.76***	3.95***	.63*	3.14***	3.08***	3.08***
Partic		-1.40	-1.44	-1.62	-2.22	-1.95	N/A	1.83	-1.82	N/A
Wave_C*Partic		-.02	.02	.03	.45	.20	N/A	.01	.23	N/A
School 1			5.67*	5.55*	5.47*	5.26*	.70	5.47*	5.39*	4.72*
School 2			6.27**	5.44**	5.42**	5.65**	-1.05	5.55**	5.55**	5.91**
School 4			6.70**	6.17**	6.24**	5.98**	1.40	5.98**	5.86**	5.99**
School 5			8.26~	8.06~	8.71~	8.03~	.63	7.84~	7.66~	7.29~
Wave_C*Sch1			-3.11**	-3.20**	-3.15**	-3.11**	N/A	-3.18**	-3.08**	-2.66*
Wave_C*Sch2			-4.01***	-4.02***	-3.80***	-4.05***	N/A	-4.13***	-4.08***	-4.26***
Wave_C*Sch4			-3.11***	-3.45***	-3.37***	-3.38***	N/A	-3.35***	-3.23***	-3.21***
Wave_C*Sch5			-3.77**	-4.02**	-4.34**	-3.96**	N/A	-3.98**	-3.79**	-3.76**
Male				3.03*	2.93~	2.98~	2.93**	2.91**	3.20**	2.83**
Hisp				-.70	-1.45	-.73	.01	.34	.34	.34
Afam				-1.42	-1.92	-1.76	-1.37	-1.33	-1.40	-1.43
Poverty				-3.42~	N/A	N/A	N/A	N/A	N/A	N/A
FRL				.50	-3.23	-2.96	3.96*	-4.88**	-4.77**	-4.55**
ELL				.18	.27	.20	-.88	-.73	-.66	-.81
SpEd				-6.28**	-6.26**	-6.27**	-7.26***	-7.38***	-7.35***	-7.55***
Wave_C*Male				-.16	-.16	-.14	N/A	N/A	N/A	N/A
Wave_C*Hisp				.40	.77	.41	N/A	N/A	N/A	N/A
Wave_C*AFAM				.03	.18	.14	N/A	N/A	N/A	N/A

Wave_C*FRL				-2.33**	-1.63	-1.55	N/A	N/A	N/A	N/A
Wave_C*POV				1.75*	1.07~	.84	N/A	N/A	N/A	N/A
Wave_C*ELL				-.30	-.30	-.29	N/A	N/A	N/A	N/A
Wave_C*SpEd				-.59	-.61	-.58	N/A	N/A	N/A	N/A
Gr1 Partic					-1.71	N/A	N/A	N/A	N/A	N/A
Gr2 Partic					2.8.6	N/A	N/A	N/A	N/A	N/A
Gr3 Partic					.20	N/A	N/A	N/A	N/A	N/A
Gr4 Partic					.19	N/A	N/A	N/A	N/A	N/A
Gr5 Partic					-3.03	N/A	N/A	N/A	N/A	N/A
Wave_C*Gr1					-.70	N/A	N/A	N/A	N/A	N/A
Wave_C*Gr2					-1.04	N/A	N/A	N/A	N/A	N/A
Wave_C*Gr3					.67	N/A	N/A	N/A	N/A	N/A
Wave_C*Gr4					.13	N/A	N/A	N/A	N/A	N/A
Wave_C*Gr5					.94	N/A	N/A	N/A	N/A	N/A
Days						.01	N/A	N/A	N/A	N/A
Wave_C*Days						.00	N/A	N/A	N/A	N/A
Wave_c*Sch1Days							.00	N/A	N/A	N/A
Wave_c*Sch2Days							.00	N/A	N/A	N/A
Wave_c*Sch4Days							-.01~	N/A	N/A	N/A
Wave_c*Sch4Days							.00	N/A	N/A	N/A
Wave_c*FRL*DAYS								.00	N/A	N/A
Wave_c*Male*Days									-.01	N/A
Variance Components										
σ_{ε}^{2}	48.57***	48.62***	47.93***	47.77***	47.71***	48.08***	48.59***	48.09***	48.07***	45.30***

σ_0^2	76.08***	75.38***	70.21***	62.33***	62.85***	68.24***	69.03***	67.06***	66.69***	65.94***
σ_1^2	3.60~	3.52~	1.58	1.18	1.20	2.08	3.67~	1.88	1.94	1.15
σ_{01}	-7.93~	-7.79~	-4.51	-4.64	-5.09	-6.29	-8.33~	-5.95	-6.03	-4.95
Goodness-of-fit										
Deviance	6290.120	6288.279	6254.617	6207.337	6201.402	6182.102	6250.308	6187.765	6186.249	5995.506
AIC	6302.120	6304.279	6286.617	6265.337	6279.402	6190.102	6258.308	6195.765	6194.249	6035.506
BIC	6330.689	6342.371	6362.803	6403.423	6465.103	6209.022	6277.280	6214.722	6213.206	6130.102

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 72: Taxonomy of fitted regression models for Reading achievement over time, 2009-2013 (*n* = 353)

	Model 1: Uncondit	**Model 2: Partic**	**Model 3: School**	**Model 4: Demogs**	**Model 5: Grade**	**Model 6: Dosage**	**Model 7: School * Days**	**Model 8: FRL days**	**Model 9: Male days**	**Model 10: Final**
Intercept	41.81***	42.65***	40.40***	43.29***	43.50***	43.80***	46.25***	43.77***	43.64***	44.36***
Wave_C	.37	.21	2.47***	2.56*	2.42***	2.31***	.14	2.30***	2.28**	2.03**
Partic		-1.58	-1.37	-1.46	-2.32	-2.92~	-1.64	-1.44	-1.44	-3.26*
Wave_C*Partic		.31	.23	.29	.82	.94	.54	.21	.32	1.16~
School 1			2.26	2.17	2.61	2.49	1.22	2.33	2.30	2.28
School 2			4.28*	3.32~	3.98*	3.78~	-2.52~	3.78~	3.77~	4.22*
School 4			3.24	2.86	3.32~	2.62	-.19	2.87	2.83	2.44
School 5			.23	-2.07	-1.08	-2.34	-3.81	-1.53	-1.67	-2.44
Wave_C*Sch1			-2.77*	-2.58*	-2.52*	-2.73*	N/A	-2.59*	-2.54*	-2.54*
Wave_C*Sch2			-4.08***	-3.77***	-4.20***	-4.04***	N/A	-4.06***	-4.02***	-4.24***
Wave_C*Sch4			-2.50**	-2.40**	-2.59**	-2.30**	N/A	-2.40**	-2.36**	-2.15**
Wave_C*Sch5			-2.09	-1.67	-2.23	-1.64	N/A	-2.00	-1.90	-1.55
Male				.91	-.41	-.53	-.29	-.45	-.38	-.69
Hisp				4.16~	2.40	2.62~	2.51	2.70~	2.71~	2.59~
Afam				2.50	1.60	1.71	1.81	1.81	1.80	1.73
Poverty				-.06	N/A	N/A	N/A	N/A	N/A	N/A
FRL				-1.59	-1.74	-1.84	-1.49	-1.83	-1.71	-2.00
ELL				-7.19*	-4.37*	-4.30*	-4.34*	-4.41*	-4.40*	-4.25*
SpEd				-12.07***	-9.80***	-9.80***	-9.97***	-9.85***	-9.83***	-9.67***

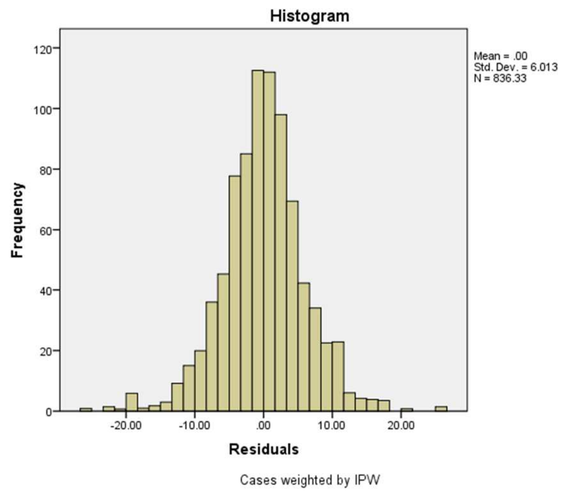
Wave_C*Male				-.82	N/A	N/A	N/A	N/A	N/A	N/A
Wave_C*Hisp				-.75	N/A	N/A	N/A	N/A	N/A	N/A
Wave_C*AFAM				-.46	N/A	N/A	N/A	N/A	N/A	N/A
Wave_C*FRL				-.41	N/A	N/A	N/A	N/A	N/A	N/A
Wave_C*POV				.40	N/A	N/A	N/A	N/A	N/A	N/A
Wave_C*ELL				1.30	N/A	N/A	N/A	N/A	N/A	N/A
Wave_C*SpEd				1.41	N/A	N/A	N/A	N/A	N/A	N/A
Gr1 Partic					N/A	N/A	N/A	N/A	N/A	N/A
Gr2 Partic					8.47~	N/A	N/A	N/A	N/A	N/A
Gr3 Partic					1.19	N/A	N/A	N/A	N/A	N/A
Gr4 Partic					.82	N/A	N/A	N/A	N/A	N/A
Gr5 Partic					-6.26	N/A	N/A	N/A	N/A	N/A
Wave_C*Gr1					N/A	N/A	N/A	N/A	N/A	N/A
Wave_C*Gr2					-5.18*	N/A	N/A	N/A	N/A	N/A
Wave_C*Gr3					.04	N/A	N/A	N/A	N/A	N/A
Wave_C*Gr4					-.57	N/A	N/A	N/A	N/A	N/A
Wave_C*Gr5					2.18	N/A	N/A	N/A	N/A	N/A
Days						.03~	N/A	N/A	N/A	.03~
Wave_C*Days						-.01~	N/A	N/A	N/A	-.01~
Wave_c*Sch1Days							.00	N/A	N/A	N/A
Wave_c*Sch2Days							.00	N/A	N/A	N/A
Wave_c*Sch4Days							-.01	N/A	N/A	N/A
Wave_c*Sch5Days							.00	N/A	N/A	N/A
Wave_c*FRL*DAYS								.00	N/A	N/A

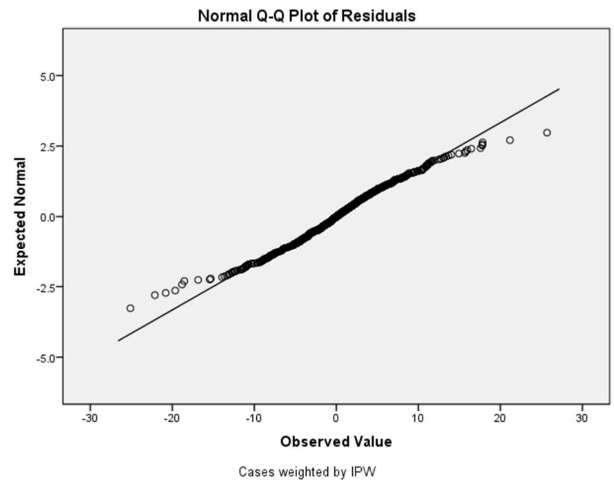
Wave_c *Male* Days									.00	N/A
Variance Components										
σ_{ε}^{2}	40.42***	40.40***	40.18***	40.38***	39.52***	39.84***	40.72***	40.14***	40.20***	41.76***
σ_{0}^{2}	93.07***	92.54***	92.72***	69.48***	70.95***	70.46***	71.81***	71.06***	70.82***	67.63***
σ_{1}^{2}	9.35***	9.36***	7.51***	7.20***	7.72***	7.43***	9.08***	7.53	7.53***	5.44**
σ_{01}	-16.71***	-16.64***	-15.38***	-12.71***	-13.42***	-12.74***	-14.98***	-13.15***	-13.12***	-10.22**
Goodness-of-fit										
Deviance	6034.698	6033.585	6007.082	5952.506	5948.896	5953.902	5978.178	5957.242	5957.387	5752.400
AIC	6046.698	6049.585	6039.082	6012.506	6008.896	6001.902	6022.178	6003.242	6003.387	5800.400
BIC	6075.056	6087.395	6114.702	6154.293	6150.683	6115.331	6126.155	6111.946	6112.090	5912.950

~ $p <$ .10, * $p <$ .05, ** $p <$ .01, *** $p <$ .001

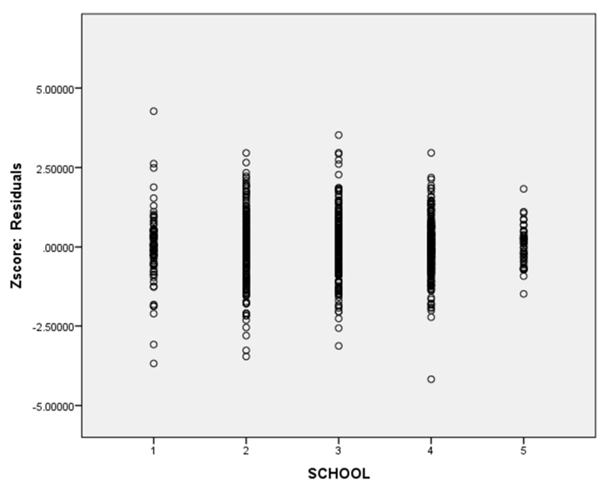
Appendix J

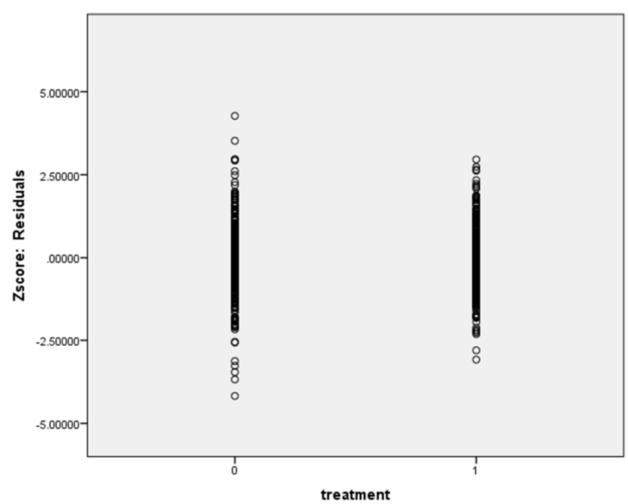
Figures 99 & 100: Histogram of unstandardized residuals for math achievement over time and Normal Z-Z plot of residuals (*n* = 367)



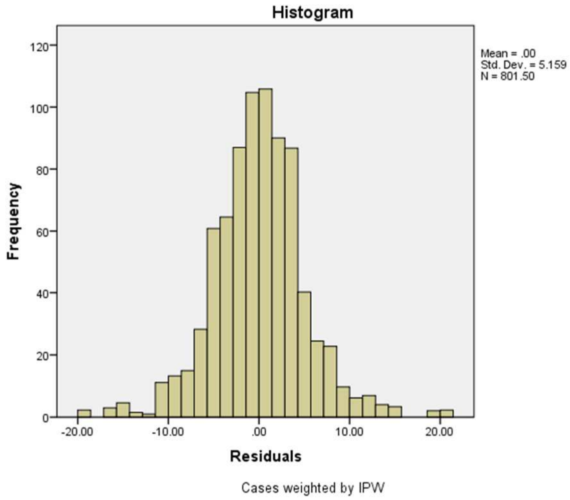


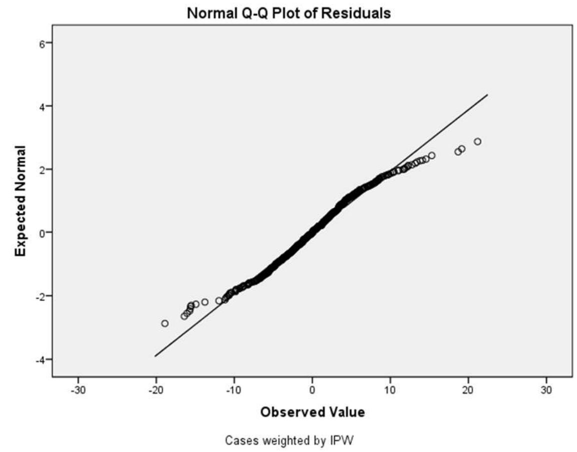
Figures 101 & 102: Dot plots of unstandardized residuals for math achievement over time by school and by treatment status (*n* = 367)





Figures 103 & 104: Histogram of unstandardized residuals for reading achievement over time and Normal Z-Z plot of residuals (n = 353)





Figures 105 & 106: Dot plots of unstandardized residuals for reading achievement over time by school and by treatment status (n = 353)