Attempting Education Reform through the Courts:

The Inefficacy of *Abbott V* Court-Mandated Funding Reallocation and Reforms in Underperforming School Districts in New Jersey

Undergraduate Honors Thesis
Sanford School of Public Policy
Duke University
Durham, North Carolina

Kevin Song
December 2010

I would like to thank the following people without whom this thesis would not have been possible:

Dr. Helen F. Ladd, my thesis advisor and professor, for teaching me everything I know about education policy and for challenging me to work, think, and write better than I have ever before.

Dr. Kenneth S. Rogerson, my honors thesis seminar director, for pushing me to write a thesis that October day in Glasgow last year and for his patience when I was hopelessly lost and needed support.

Austin T. Stanion, a best friend and former roommate, for his infallible positivity and for teaching me how perseverance requisites balance in 5K runs.

Jonathan E. Amgott, my thesis buddy and workshop partner, for commiserating with me for the past year and for helping me with my thesis presentation.

Samuel T. Savitz, my Glasgow friend and thesis workshop partner, for his extensive feedback that brought order to chaos.

Table of Contents

I. Abstract

This research project analyzes the effects of the 1998 *Abbott V* New Jersey Supreme Court decision on student achievement in the state's underperforming Abbott districts. The Court mandated the reallocation of the districts' state aid to best practice reforms. This project uses regression analysis to compare the Abbott districts to socioeconomically similar districts in New Jersey while controlling for the confounding factors of each district's Black student percentage, Hispanic student percentage, socioeconomic status, and expenditure per pupil. The results show that *Abbott V* had no effect on fourth grade achievement between 2004 and 2007 and lowered achievement growth for cohorts of students who entered fourth grade in 2004 and 2005. The results hold for both language arts and math achievement measured both as district proficiency rate and district average score. The project concludes that *Abbott V* was not effective in increasing student achievement and that the state was justified in eliminating the Abbott district system under the School Funding Reform Act of 2008.

II. Searching for Successful Education Reform

> "Let us think of education as the means of developing our greatest abilities, because in each of us there is a private hope and dream which, fulfilled, can be translated into benefit for everyone and greater strength for our nation."
>
> – John F. Kennedy

For decades, politicians and education specialists have searched for ways to improve underperforming school districts. This is especially true for districts in urban areas with high percentages of disadvantaged and underserved minority students. In the 1990s, New Jersey moved away from traditional legislature-driven school reforms and to a new approach of court-mandated reforms.

Using this new system, the New Jersey Supreme Court prescribed a broad set of reforms for the state's worst performing districts based on the best practices identified in education research. The court gave this ruling in 1998, and the affected districts – termed the Abbott districts – began their implementation of these practices in 1999. Since then, these Abbott districts have narrowed the achievement gap between their students and other students in New Jersey: In 1992, New Jersey had a 35 point Black-white gap in fourth grade language arts and a 38 point gap in fourth grade math, both higher than the national averages of 32 points and 35 points, respectively.[1] By 2007, the New Jersey gaps had narrowed to 26 points in language arts and 23 points in math, both lower than the then national averages of 27 points and 26 points (Khadaroo 2009).

Since these results look very promising and can potentially become a model for education reform in other states, it is important to examine whether these results can be attributed to the court-mandated reformed. Over the past decade, New Jersey districts have experienced new and

[1] The test results came from administrations of the National Assessment of Educational Progress exam.

differing demographic patterns that have affected their student ethnic and socioeconomic compositions. In addition, New Jersey Governor John Corzine restructured the state's education system with the School Funding Reform Act in 2008 and effectively eliminated the court-mandated reforms.[2] An analysis on how these court-mandated reforms influenced student achievement during the years in which they took effect can show the possible implications of the new restructuring.

This analysis uses regression models to measure the effect of the *Abbott V* court ruling that mandated that the Abbott districts reallocate their state aid to best practice measures. It isolates the effect by controlling for factors that also influence student achievement, specifically student ethnic composition, socioeconomic characteristics, and district spending. The results show that *Abbott V* had no statistically significant effect on language arts and math achievement for fourth grade students and lowered the student achievement gains in the Abbott districts between fourth and eighth grade. The results help justify New Jersey's elimination of the ineffective Abbott district system under the School Funding Reform Act.

[2] The School Funding Reform Act was actually passed in the state legislature in 2008 but was only approved by the courts in 2009.

III. History of the Abbott Districts

The history of educational inequality and demographic trends in New Jersey provides the proper context for understanding the motivations behind the creation of the court-mandated reforms and for the analysis of policy implications.

The Battle for School Reform in the New Jersey Supreme Court

The Abbott district designation and policies arose from two series of court cases focused on education finance and equity. The New Jersey Constitution requires the state to provide a "thorough and efficient education" for all students. In the two series of cases, the *Robinson* cases starting in 1973 and the *Abbott* cases starting in 1981, plaintiffs from low wealth districts sued the state for using a significantly inequitable financing scheme (Walker and Gutmore 2002, p. 62). Prior to these cases, almost all of a district's funding came from local property taxes. In the *Robinson* cases, the New Jersey Supreme Court ruled that policy unconstitutional because it created a funding disparity between the state's high wealth and low wealth districts that in turn limited the educational opportunities for students in the low wealth districts (Walker and Gutmore 2000, p. 12). The state responded in 1976 by passing a 2% state income tax to provide monetary aid for the low wealth districts (Greif 2004, p. 621).

However, the same plaintiffs continued to argue that the low wealth districts could not provide their students with a quality education and started the series of *Abbott* cases in 1981. In the first decision in 1988, the Supreme Court ruled in *Abbott I* that the state income tax was not enough and, then in 1990 with *Abbott II*, that the state needed to give extra funding to the 28 districts with the lowest socioeconomic levels (Greif 2004, p. 621).[34] The court based its

[3] The original 28 districts are Asbury Park, Bridgeton, Burlington City, Camden, East Orange, Elizabeth, Garfield, Gloucester City, Harrison, Hoboken, Irvington, Jersey City, Keansburg, Long Branch, Millville, New Brunswick, Newark, Orange, Passaic, Paterson, Pemberton, Perth Amboy, Phillipsburg, Pleasantville, Trenton, Union City,

classification using the 1970 Census data which was the most recent socioeconomic data available when the *Abbott* cases began in 1981. Under *Abbott II*, the court required that the state give these officially designated Abbott districts enough money so that their per pupil funding matched that of New Jersey's wealthiest district (Greif 2004, p. 41). Eight years later, in 1998, the court ruled with *Abbott V* that the state had not yet fully addressed the educational inequalities in the Abbott districts, that the increased funding was necessary but not sufficient. The court was concerned that the money was not spent in the right places so it ruled that state must earmark that money to fund certain educational best practices in the Abbott districts (Walker and Gutmore 2000, p. 14).

Abbott V marked a significant departure for education reform because the court itself, rather than defer the creation of policies to the legislature, enumerated specific programs that the state had to help implement in the Abbott districts. The remedies were meant to address the educational deficiencies in the districts and constituted four different categories: standards-based reform, early childhood education, social and health services, and facilities improvements. The ambitious agenda called for, among other things, school-based management teams to oversee school personnel, curriculum, and budgeting; half-day preschools and full day kindergartens; health and nutrition programs; and lower classes sizes (Walker and Gutmore 2000, p. 15).[5]

The Difficult Implementation Process and the "Promising" Results

The New Jersey Supreme Court established a strict timetable for the state and school districts to redistribute the funds and implement the best practices, but the districts had difficulty implementing the reforms. Critics argue that the remedies had "no overarching framework that

---

Vineland, and West New York. In 1998, the state added Neptune and Plainfield, and in 2004, Salem City to bring the total to 31 districts.

[4] Also, New Jersey districts tend to be smaller than those of other states. In the 2009-2010 school year there were 605 districts.

[5] Refer to Appendix 1 for a full listing of the *Abbott V* best practices.

[brought] the disparate elements together" (Walker and Gutmore 2002, p. 64) and that the Abbott districts did not have the organizational capacity to adopt them. For example, none of the districts managed to open their half-day preschools by the Fall 1999 deadline and even by 2002, they only managed to serve two-thirds of the total eligible children (Greif 2005, p. 626). Simply put, the size and scope of the reforms overwhelmed the Abbott districts.

Despite the difficulties, the standardized test scores for the Abbott districts suggested that they had moved in the right direction. While the New Jersey mean for fourth grade language arts scores decreased by two points between 2001 and 2007, the Abbott district mean increased by three points. So in this period, the Abbott districts closed the language arts gap by five points in this period or 0.40 standard deviations (Geortz and Weiss 2009, p. 23). This information indicates that the Abbott districts made progress, but the research compared the Abbott districts to the state overall when it would be more telling to compare Abbott districts to peers that also have low socioeconomic status to create a common benchmark.

New Socioeconomic and Demographic Trends in New Jersey

Further complicating the case, the Abbott districts have experienced significant socioeconomic and demographic changes since the *Abbott* cases began in 1981. Starting in 1974, New Jersey uses Census socioeconomic data[6] to index each district into District Factor Groups (DFGs) ranging from A to J, with A as the lowest and J the highest.[7][8] In the 1990 *Abbott II* case, the Supreme Court classified the 28 DFG A districts from the 1970 Census data as Abbott districts.[9] Since then, nine have become DFG B districts, one has become a DFG C district, and

---

[6] The socioeconomic indicators are: percent of adults failing to complete high school, percent of adults who attended college, occupational composition, population density, income, unemployment, and poverty.
[7] There are a total of 8 DFGs: A, B, CD, DE, FG, GH, I, and J.
[8] The state continues to use DFGs for designation purposes even after the School Funding Reform Act of 2008.
[9] The court used a 1970 index because the proceedings for *Abbott II* started in 1981 before state updated the DFGs with 1980 Census data.

one has even become a DFG FG[10] district. In other words, some of the Abbott districts that had low socioeconomic status during the original designation have undergone considerable socioeconomic improvement.

Changing demographic patterns have also changed the student ethnic composition in the Abbott districts. After World War II, white families in New Jersey tended to leave the urban centers and moved into the suburbs. By 1970, the yet-to-be-designated Abbott districts, which are mostly in these urban areas, were overwhelmingly Black and enrolled over 85% of the state's ethnic minority students. Then, New Jersey experienced a wave of immigration in the 1990s during which immigrants, especially from Latin America, accounted for half of the state's population. In addition, following the gentrification of some urban, white families began returning to the urban centers. In 2007, the Abbott districts were 44.6% Hispanic and only 40.2% Black, and collectively enrolled only half of the state's minority students (Goertz and Weiss 2009, p. 8).

The Declassification of the Abbott Districts by the School Funding Reform Act of 2008

In 2008, New Jersey implemented a new education policy that completely declassified the Abbott districts and eliminated the *Abbott V* court mandates. The New Jersey legislature passed the School Funding Reform Act (SFRA) which created sweeping changes to the state's education system. Partially due to the aforementioned demographic trends, the state was frustrated with the "fragmented system of funding under which limited resources cannot be distributed equally to all districts where at-risk children reside" (*School Funding Reform Act* 2008, p. 3). Essentially, the state believed that the *Abbott* cases, including the targeted reallocation of funding in *Abbott V*, created a system that rigidly tied up too much money. Under the new SFRA funding formulas, the state would still give extra aid to the Abbott districts

[10] For some reason, the state uses one letter for some designations and two letters for others.

through a hold harmless clause that stipulates that a district will not receive less state aid than it had before SFRA. In exchange, the state would free itself from the obligations and provisions of the *Abbott* cases. In response, the Education Law Center, a proponent of the Abbott district system, challenged the bill before the New Jersey Supreme Court on grounds that it would reverse the gains made by the Abbott districts (Sciarra, 2009 p. 2). In 2009, the court accepted the constitutionality of SFRA, mainly on grounds of the hold harmless clause, and the state enacted the provisions of the SFRA starting in the fall of 2009.

IV. Methodology and Data Selection

That the state moved away from the Abbott system when the districts were showing progress raises questions about the wisdom of the state's decision. While SFRA essentially maintains *Abbott II* – the hold harmless provision preserves the extra aid that the state gives to the now-declassified Abbott districts – the state no longer allocates the extra aid to the best practice reforms as established by *Abbott V*. If these court-mandated reforms had been successful in increasing student achievement in the Abbott districts, then SFRA could erase that progress. If they were not successful, then that would be more reason for the state to try the new approach under SFRA. Determining whether and how the *Abbott V* court-mandated reforms changed student achievement reveals the implications of abandoning the Abbott system in favor of adopting SFRA.

Statistically determining this association requires a regression analysis of student achievement. Therefore student achievement metrics are the dependent variables. To distinguish the effect of the *Abbott V* reform on these student achievement metrics from other confounding factors, it is necessary to control for the student characteristics – specifically the percentages of students who are Black and Hispanic, and participation in free and reduced lunch programs – and district expenditure per pupil.[11] Furthermore, to refine the regression analysis, it is important to compare the Abbott districts to their peer districts that are socioeconomically similar. The regression models incorporate these nuances.

[11] Since I am examining the effects of the *Abbott V* court-mandated reforms, I control for expenditure because it was the result of the *Abbott II* ruling.

Using Regression Analysis to Analyze District Test Results

When determining the correlation between one independent factor to a dependent factor, researchers typically use regression analyses to control for confounding factors or the other factors that might influence the dependent factor. A general regression formula looks like this[12]:

$$Y_i = \beta_1 X_{1_i} + \beta_2 X_{2_i} + \beta_3 X_{3_i} + \ldots + c + \varepsilon_i$$

The regression coefficients, β_n, indicate how much changes in their respective variables, X_n, correspond with changes in the dependent variable, Y, for each observation, i, in the sample after accounting for the other variables in the model. There is also a constant coefficient, c, to indicate the baseline of the dependent variable when all the explanatory variables have values of 0. The error term, ε, then captures the effect of all the unmodeled factors that influence Y.

To isolate the effect of a treatment or categorical difference, one can use a 0 or 1 indicator variable as an explanatory independent variable. A 0 represents an observation without the treatment, and a 1 represents an observation that has the treatment. The regression coefficient for the indicator variable reveals whether that treatment is correlated with the dependent variable.

It is important to note that while a regression reveals the correlation of independent and explanatory variables to dependent variables, it is not enough to prove causation. That is, a regression does not say that a change in an explanatory variable, even if it is statistically significant and highly correlated, causes a change in the dependent variable.

[12] There are many different types of regression models to take into account the many different types of statistical relationships between variables. For clarity and space, I will explain and use ordinary least squares regressions for this project because it is the standard one used in education research (McEwan 2008, p. 90).

The Dependent Variable: Student Achievement Measures

The dependent variables of the regression should be the best metrics of student achievement. There are two primary methods for measuring student achievement: the status model for examining the changes in the same grade level over time, and the growth model for examining the changes of a cohort of students as they progress through grades over time. In applying these two models, the availability of data and need for comparable data limit the researchable time frame, the content subjects, grade levels, and the test result format.

First, the availability of comparable test results limits the time frames. Prior to 2003, New Jersey administered three standardized tests: the ESPA for fourth grade, GEPA for eighth grade, and the HSPA for eleventh grade.[13] Starting in 2004, the New Jersey Department of Education changed its testing policy and replaced the ESPA with the grade-specific NJASK[14] for third and fourth graders. Then in 2006, the state started administering NJASK for grades 5-7. Finally in 2008, the state redesigned all the NJASK tests and replaced the GEPA with NJASK8 for eighth grade. The following table presents the tests that were administered to each grade:

Years[15]	1999-2003	2004-2005	2006-2007	2008-Present[16]
Grade 3	-	NJASK3	NJASK3	NJASK3*
Grade 4	ESPA	NJASK4	NJASK4	NJASK4*
Grade 5	-	-	NJASK5	NJASK5*
Grade 6	-	-	NJASK6	NJASK6*
Grade 7	-	-	NJASK7	NJASK7*
Grade 8	GEPA	GEPA	GEPA	NJASK8*
Grade 11	HSPA	HSPA	HSPA	HSPA

Figure 1. Tests Administered by Grade.

[13] The full names are the Elementary School Proficiency Assessment, the Grade Eight Proficiency Assessment, and the High School Proficiency Assessment.

[14] The New Jersey Assessment of Skills and Knowledge. The tests for different grades are specified by the grade number; for example the fourth grade version is NJASK4.

[15] Again, these are end-of-grade tests so a test administered in spring 1999 was for school year 1998-1999.

[16] The NJASK tests in the 2008-Present column have asterisks to demonstrate they are substantively different from the NJASK tests from the previous columns despite having the same name.

These frequent test changes are problematic for analyzing the test results because the results must be scaled the same to allow for direct comparison across the tests. For example, to make a direct comparison of fourth grade scores from 1999 to 2006, a score of 200 on the ESPA must mean the same as a 200 on the NJASK4. However, since the state changed the contents of the fourth grade tests within that time period, the scores are not directly comparable (*Technical Report* 2008, p. 56, 60). Education researchers can work around this problem by examining differences in standard deviation units for the tests (Reardon and Robinson 2008, p. 498). Unfortunately, the New Jersey Department of Education does not calculate standard deviations for their tests so this cross-test comparison is impossible for this analysis.

Considering this comparability limitation and the benefit of having more usable data, my regressions are based on the longest possible time frames. First, the status models examine the NJASK4 test results for fourth grades for the four years between 2004 and 2007. Also, the students who entered fourth grade in the 2003-2004 school year and took the 2004 NJASK4 started kindergarten in the 1999-2000 school year, which was the first year that the state began the *Abbott V* fund reallocation reforms. So in addition to being the longest period in which the tests did not change, the 2004-2007 time frame includes students who have been educated under the *Abbott V* reforms and best practices since they started school. In other words, this time frame isolates the effects of the *Abbott V* reforms.

Second, the growth model examines the two groups of students who entered the fourth grade in 2004 and in 2005. The growth model reveals the changes in the achievements of pseudo-cohorts[17] over time by comparing their fourth grade scores to their eighth grade scores in 2008 and 2009, respectively. Ideally, this model would use individual student achievement data

[17] These groups are pseudo-cohorts rather than true cohorts because each district has the same group students advancing through the grades year after year, there is some change as families move around and students enter and exit school districts.

from each student from when he or she was in fourth grade and when he or she was in eighth grade. Unfortunately, because New Jersey does not post the tracked test results of each student as he or she progresses through school, these regressions models can only use aggregate test results on the district level for this time period. In other words, my analysis can only examine groups of student pseudo-cohorts specifically the 2004/2008[18] cohort and the 2005/2009 cohort. Although it would be useful to use the most recent 2006/2010 cohort, the test results from the 2010 administration of the test are not yet available.

The content and reporting methods of the New Jersey tests also determined the test subjects and result formats. New Jersey has administered language arts and math subject tests in all of the tests during these time periods and science subject tests starting in 2005. Since education researchers are primarily interested in student achievement in language arts and math and since the newly crafted science tests have a very limited time span of data, this analysis only focuses on language arts and math. Also, the state reports the results in two ways: the percentage of the district's students who have scored proficient and the average score of all the students in a district. Since both proficiency rates and average scores are available for all the years in my time frames, the regression models use both of these test results.

The following table shows the different metrics:

Regression	Model	Time Frame	Subject	Test Result
1	Status	2004-2007	Language Arts	District Proficiency Rate
2	Status	2004-2007	Language Arts	District Average Score
3	Status	2004-2007	Math	District Proficiency Rate
4	Status	2004-2007	Math	District Average Score
5	Growth	2004/2008 and 2005/2009	Language Arts	District Proficiency Rate
6	Growth	2004/2008 and 2005/2009	Language Arts	District Average Score
7	Growth	2004/2008 and 2005/2009	Math	District Proficiency Rate
8	Growth	2004/2008 and 2005/2009	Math	District Average Score

Figure 2: Chart of Student Achievement Metrics

[18] This stands for the class of students who entered fourth grade in 2004 and eighth grade in 2008.

While it may seem excessive to conduct eight different regressions, these metrics measure different things and can have different policy implications. First, the two models help determine the differentiated effects of the *Abbott V* reforms on student achievement. If the status model indicates the treatment is effective for fourth grade students but the growth model indicates that the effectiveness disappears by the time they reach eighth grade the policy implications will be different than if both the status and growth models show effectiveness. Also, it is possible that the reforms have a delayed effect so that the status model could show that the treatment was not effective for students in fourth grade but that the growth model could show that the treatment was effective by the time those students reached eighth grade. Second, the differentiation between subjects is important because language arts and math are taught and learned differently and the *Abbott V* reforms may have differing impacts on them. Third, looking at both forms of test results will elucidate the effect of the *Abbott V* reforms on different segments of the student distribution: the proficiency rates show the movement of the students who are at the bottom of the distribution, and the average scores show the movement of the entire distribution of students. Using test metrics that incorporate a combination of these achievement models, subjects and test results paints a more complete picture of what happened in the Abbott districts.

The Control Variables: Removing Confounding Factors

To determine the correlation between the *Abbott V* reforms and the changes in these eight student achievement metrics, it is necessary to control for other confounding variables that may affect student achievement (Rice and Schwartz 2008, p. 139). I control for three measures of a district's student characteristics – district percentage of Black students, district percentage of Hispanic students, participation in free and reduced lunch – and one measure of district spending.

Data for all these variables came from the National Center for Education Statistics' Core of Common Data (CCD) website.

Ethnic Composition

First, there is a well-documented ethnic educational achievement gap in America so a district's ethnic composition may influence that district's test results (Reardon and Robinson 2008, p. 499). The achievement gap describes how different ethnic groups have sizably different average test results in both language arts and math. According to the 2005 administration of the National Assessment of Educational Progress (NAEP) language arts test[19], fourth grade Black students scored 0.81 standard deviations below their white peers; Hispanic students, 0.72 standard deviations below; and Asian students, statistically indistinguishable (Reardon and Robinson 2008, p. 502). The magnitudes of the achievement gaps for eighth grade language arts scores are nearly identical, so for language arts, the relative achievement rankings of ethnic groups are as follows:

$$Asians = Whites > Hispanic > Black$$

A similar pattern emerges for student achievement in math. Fourth grade Black students scored 0.92 standard deviations below their white peers; Hispanic students, 0.74 standard deviations below; and Asian students, 0.17 standard deviations above (Reardon and Robinson 2008, p. 502). Once again, the patterns in student achievement in eighth grade math mirror those in fourth grade math. The relative math rankings are as follows:

$$Asians > Whites > Hispanics > Blacks$$

To control for ethnic composition, the regression models include variables for the percentages of Black students and the percentage of Hispanic students in a district. Controlling

[19] More formally, it is the Main NAEP exam that tests grade-level cohorts. There is also a Long-Term Trend NAEP that tests age-level cohorts that does not account for grade retention.

for student composition is especially important in New Jersey because the state has experienced changing demographic patterns. Since Asian students are generally not educationally disadvantaged, the regression models will not account for the percentage of Asian students. The Core of Common Data has student enrollment broken down into ethnic groups by grade, year, and district. Unfortunately the CCD enrollment data might not completely reflect the composition of students who take the end-of-grade tests that factor into district's test results because students may be absent on the test day or have voided exams. However, because the achievement gap is so important in education and because the error caused by absentees and voided exams is likely to be small, these regressions include variables for ethnic composition.

Socioeconomic Characteristics

The regressions also control for differences in the socioeconomic characteristics of the districts as measured by the percentage of a district's student who participate in free and reduced lunch. Generally, students from lower socioeconomic backgrounds are disadvantaged – such as in terms of out-of-school enrichment, early childhood care, and health – and show lower student achievement (Reardon and Robinson 2008, p. 510). The most common metric for controlling for socioeconomic characteristics is participation in free or reduced lunch programs[20]. The lunch programs are federally funded initiatives aimed at providing low-income students with meals during the school day. Thus, they are highly correlated with poverty rates, especially at the elementary school level (Duncombe and Yinger 2008, p. 248). The CCD has data for participation in the lunch programs in each school district, but unfortunately they are not disaggregated by grade levels or by individual schools. However, this is not too concerning

[20] While the cutoff eligibility point for reduced lunch is slightly higher than the cutoff eligibility point for free lunch, the programs are both aimed are poorer students and similar enough to be grouped together.

because the socioeconomic levels of students in fourth and eighth grade should be consistent with the overall socioeconomic levels of the district.

District Spending

Finally, district spending also influences student achievement so my regressions will have a variable for district expenditure per student. Generally, money matters in education because increased spending creates a higher quality of education (Krueger 2002, p. 21). Spending provides for benefits such as decreased class sizes, increased teacher salaries, better professional development, and more support services. For simplicity, researchers often use education expenditure data to aggregate these inputs (Rice and Schwartz 2008, p. 136).

Spending is especially important for this analysis because New Jersey increased the state aid to Abbott districts as part of the *Abbott* court cases. The state aid allowed the Abbott districts to spend on average $14,294 per pupil while New Jersey's highest wealth districts in DFG J spent $12,999 per pupil in 2007 (Goertz and Weiss 2009, p.41).[21] Since the increased spending that resulted from *Abbott II* could increase student achievement, the regression models control for spending as to not confound the effects of *Abbott II* and *Abbott V*.

The CCD provides data for district spending, but it is not separated by grade levels and values need to be adjusted for inflation. Since this analysis examines achievement metrics at specific grade levels, it would be ideal to have grade-disaggregated per pupil spending. It generally costs more to educate students at the secondary school level – due to athletics programs, extracurricular activities, higher certification for teachers, and science labs – so district per pupil spending statistics may not reflect the per pupil spending in the district's elementary schools. However, this type of information is not available and controlling for spending is essential to isolate the effects of *Abbott V*. Finally, adjusting for inflation in the

[21] The figures are in 2007 dollars.

expenditure statistics is necessary to compare the real, rather than nominal, value of the spending so I use the Consumer Price Index to level all the spending into 2010 dollars.[22]

The Observations: Abbott Districts and Their Socioeconomic Peers

The observations and data points for my regressions include 29 out of the 31 Abbott districts and all the districts classified as DFG A or B in 2004. As explained previously, New Jersey indexes its districts based on socioeconomic characteristics. In 2004, Abbott districts constituted 19 out of the 31 DFG A districts and 10 out of the 72 DFG B. Since the purpose of this analysis to research the policy implications of court-mandated reforms for underperforming districts, the data set does not include districts in the higher DFGs because they do not have low socioeconomic characteristics nor are they likely to face court-mandated reforms.

To ensure that the data set only includes comparable districts, some wealthier Abbott districts are removed from this analysis. First, two of the Abbott districts, Pemberton and Neptune, are excluded because both were classified as DFG CD districts in 2004. While these two districts may have had low socioeconomic levels during their original classification in *Abbott II*, they had progressed enough by 2004 to enter into a higher DFG group.[23] These two districts are not in the same starting position as the other 29 Abbott districts at the beginning of the analysis time frame. Second, between 2004 and the present, some DFG A and B districts moved up in classification. For example, one of the Abbott districts, Hoboken, was in DFG B in 2004 and moved up three classifications to DFG FG by 2008. This district and all other 2004 DFG A and B districts that have subsequently improved their classifications are included in this analysis because the regression models control for changes in socioeconomic characteristics.

[22] More specifically, I will use the average annual CPI statistics. To covert year y dollars to 2010 dollars, I will use the following numbers: y=2004, 15.57%; y=2005, 11.78%; y=2006, 8.29%; y=2007, 5.29%; y=2008, 1.40%; y=2009, 1.76%.

[23] No Abbott district has even been striped of the designation, probably because it is politically difficult.

Accounting for these considerations, there are 29 Abbott and 73 socioeconomically similar non-Abbott districts, and thus there are 102 comparable districts.[24] Since there are four years between 2004 and 2007, there are 407[25] observations for the status model regressions. Since there are two pseudo-cohorts of students for which there is comparable fourth and eighth grade data, there are 204 observations for my growth model regressions.

To distinguish between the Abbott and non-Abbott districts, there is an Abbott indicator variable in the regression models. It takes on a value of 1 for an Abbott district and 0 for a non-Abbott district. While ideally this analysis would determine the particular impact of each specific mandated reform remedies on student achievement, this nuance is difficult to capture. First, since the Abbott districts could not keep to the court's implementation timetable, the reforms started at different times with different levels of quality for different districts. Thus it is difficult to definitively say when a particular district fully implemented a specific reform. Second, many of these remedies were designed to address the same goal, primary school literacy for example, so it hard to disentangle the effects of one remedy from another (MacInnes 2005). This analysis aggregates the Abbott districts together into an intent-to-treat group so that the Abbott indicator variable reveals whether *Abbott V* changed the student achievement in Abbott districts relative to that in similar non-Abbott districts.

Finally, it is also important to consider the effects of charter schools when analyzing student achievement.[26] The district test results in the data set include only students in traditional schools. Charter schools are often created in districts with low socioeconomic characteristics to target especially low performing or especially high achieving students (Bifulco and Bulkley

[24] See Appendix 2 for a list of the 73 non-Abbott districts.
[25] There are 407 districts rather than 408 [102 x 4] districts because 2005 data for one of the districts is unavailable.
[26] Appendix 2 has a list of all the charter schools in New Jersey with data on their name, grade(s), host districts, DFG, Abbott status, year opened, and enrollment.

2008, p. 430). Students exit traditional public schools and enter into charter schools so their test scores count on behalf of the charter school, not their original school district. However, this selection process does not pose a problem for this analysis. There are 77 charter schools in the state and 58 of them are in districts that were classified as DFG A or B in 2004. Out of those, 39 charter schools are potentially relevant to this analysis because they enrolled students up to and including the eighth grade. Also, all but one of those 39 schools are in an Abbott district. Looking at the fourth grade 2008 test results from two Abbott districts – Camden City which has highest percentage of students are enrolled in charter schools, and Newark which is the largest district in the state and that has the highest number of charter schools – illustrates the non-issue of charter schools to this analysis:

Level	Charter School	Total Enrollment (students)	District Percent Enrollment (%)	Language Arts Score	Math Score
Charter		112	9.3	**195.1**	**193.0**
	D.U.E. Season	60	5.0	192.2	189.5
	LEAP Academy University	52	4.3	198.5	197.0
Non-Charter		1096	90.7	**190.0**	**200.6**
Combined		1208	100	**190.5**	**199.9**

Figure 3: Effect of Charter Schools in Camden City

Level	Charter School	Total Enrollment (students)	District Percent Enrollment (%)	Language Arts Score	Math Score
Charter		271	8.2	**211.0**	**222.5**
	Discover	16	0.5	227.5	230.3
	Gray	32	1.0	223.7	245.5
	Maria L. Varisco-Rogers	14	0.4	210.5	220.7
	Marion P. Thomas	41	1.2	208.4	221.1
	New Horizons	80	2.4	203.4	209.3
	Robert Treat Academy	49	1.5	229.8	257.4
	University Heights	39	1.2	188.6	185.6
Non-Charter		3045	91.8	**203.9**	**221.8**
Combined		3587	100	**204.4**	**221.9**

Figure 4: Effect of Charter Schools in Newark

In these two districts, while the average scores of the charter schools were different from that of the non-charter schools, they were not substantially different enough nor was relative enrollment high enough to significantly alter the district averages. There are two charter schools in Camden City that enroll fourth grade students. The charter weighted average language score of 195.1 was higher than that of the non-charters of 190, and the charter average math score of 193.0 was lower than that of the non-charter of 200.6. While these differences may seem large, charter schools enrolled only 9.3% of the district's students, so Camden City's combined weighted averages was 190.5 for language arts and 199.6 for math. That is, Camden City's charter schools increased the district's average language arts score by 0.5 and decreased the average math score by 1.0.[27] The charter schools in Newark also have very small effects on district averages. The seven charter schools there have a weighted average language arts score of 211 and a weighted average math score of 222.5 which are higher than the non-charter averages of 203.9 and 221.8, respectively. Since the charter schools enrolled only 8.2% of the district's fourth grade students, the combined weighted averages increased by 0.5 to 204.4 for language arts and increased by 0.1 to 221.9 for math.[28] These small changes are unlikely to affect the findings from analyzing the data set.

The Regression Models

The first regression formula for the status models takes the following form:

$$Y_{dt} = \beta_1 A_d + \beta_2 T_t + \beta_3 A_d T_t + \beta_4 B_{dt} + \beta_5 H_{dt} + \beta_6 L_{dt} + \beta_7 E_{dt} + c + \varepsilon_{dt}$$

The dependent variable, Y_{dt}, represents the four metrics for student achievement – proficiency rates and average scores for language arts and math – for each district, d, between the years, t, 2004 to 2007. The sample includes Abbott districts and their socioeconomic peer

[27] This is shown in the four highlighted cells in Figure 3.
[28] This is shown in the four highlighted cells in Figure 4.

districts in DFGs A and B as determined at the start of my time frame in 2004. The coefficient for the Abbott indicator variable, A_d, shows whether student achievement was different between the Abbot and non-Abbott district groups.[29] The coefficient for the time variable, T_t, reveals whether achievement changed in the time frame.[30] Then, the coefficient for the Abbott-time interaction variable, A_dT_t, shows whether changes in student achievement over time in the Abbott district group were different from that in the non-Abbott district group.[31] Finally there are variables to control for the confounding factors: the percentage of Black students, B; the percentage of Hispanic students, H; and the percentage of students on free or reduced lunch, L; and the inflation-adjusted expenditure per pupil, E.

The second regression formula for the growth models takes the following form:

$$\Delta Y_{dp} = \beta_1 A_{dp} + \beta_2 \Delta B_{dp} + \beta_3 \Delta H_{dp} + \beta_4 \Delta L_{dp} + \beta_5 \Delta E_{dp} + c + \varepsilon_{dp}$$

The dependent variable, ΔY_{dp}, represents the difference between the fourth grade and eight grade student achievement metrics for each district, d, for each pseudo-cohort, p. For example, one observation for this dependent variable is the difference between Newark's fourth grade math proficiency in 2004 and its eighth grade math proficiency in 2008. The coefficient for the Abbott indicator variable, A_{dp}, shows whether student achievement growth was different between the two district groups. The values for the four control variables are also the difference between the fourth grade and eighth grade statistics: change in Black student percentage, ΔB_{dp}; change in Hispanic student percentage, ΔH_{dp}; change in free or reduced lunch participation, ΔL_{dp}; and change in expenditure per pupil, ΔE_{dp}.

---

[29] The Abbott indicator variable takes a value of 0 for non-Abbott districts and a value of 1 for Abbott districts.
[30] The time variable takes a value of 0 for year 2004, 1 for 2005, 2 for 2006, and 3 for 2007.
[31] The Abbott-time interaction variable is a multiplication of the values of the Abbott indicator variable and the time variable for each district and year. It creates a common baseline by taking on a value of 0 for the two district groups in the year 2004.

V. Analyses Yield Mixed Results

Descriptive Data Reveal Disparate Trends

Before conducting the regressions, looking at the control variables reveals the confounding trends in the Abbott and in the socioeconomically similar non-Abbott. These changes highlight the different trends in these two district groups and underscore the need to account for them. The descriptive statistics here are based on fourth grade data from 2004 to 2007, which is the data for the status regression models.

Language Arts

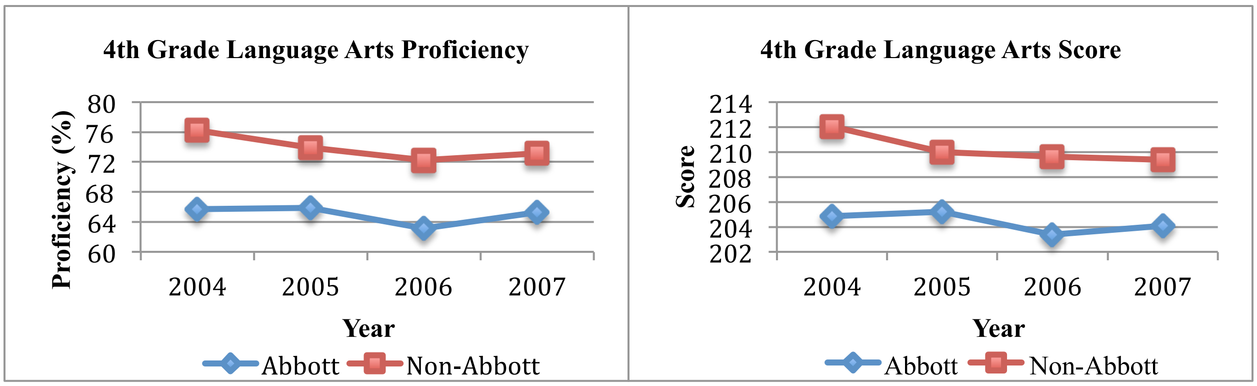

Figure 5: Descriptive Language Arts Data

For language arts, fourth graders in Abbott districts scored lower than their counterparts in non-Abbott districts as measured by both proficiency rates and average score, and, in addition, both district groups saw declining student achievement. However, the decrease in non-Abbott districts – proficiency fell by 3.2 percentage points and average score fell by 2.7 points – is greater that the decrease in Abbott districts – proficiency fell by 0.5 percentage points and average score fell by 0.8 points. So although language arts achievement decreased in both district groups, the Abbott districts suffered relatively less than the non-Abbott districts.

Math

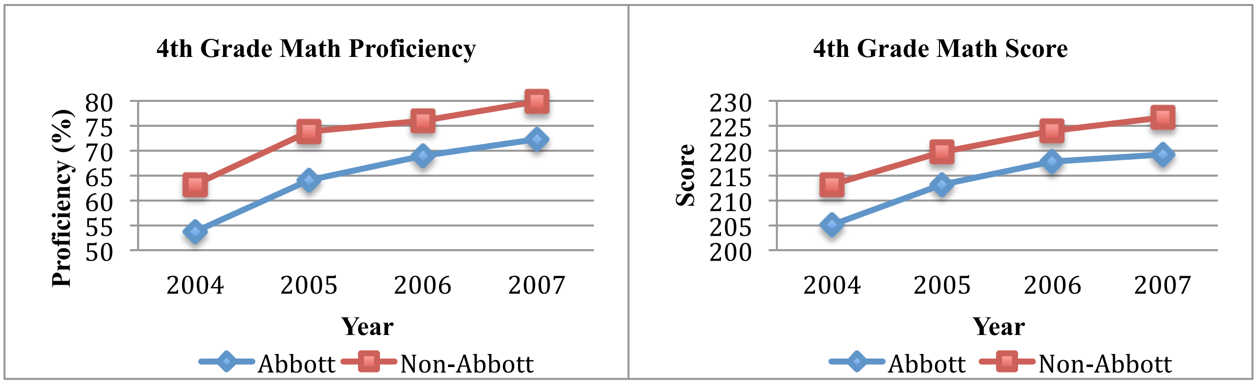

Figure 6: Descriptive Math Data

As with language arts, non-Abbott districts performed better in math as measured by both proficiency rates and average score. However in contrast to the decline in language arts achievement, both district groups exhibited an increase in math achievement during the same time period. In Abbott districts, proficiency increased by 18.5 percentage points and average scores increased by 14.1 points, which are higher than the respective gains of 13.7 percentage points and 13.5 in non-Abbott districts.

Together, these descriptive statistics show that in Abbott districts, language arts achievement decreased at a lower rate and math achievement increased at a higher rate than in non-Abbott districts. While the students in the Abbott districts still exhibited lower achievement, they demonstrated that they were making progress relative to the their peers in non-Abbott districts.

As with the test results, the Abbott and non-Abbott district groups also experienced trends of different magnitudes and directions in their student composition, free and reduced lunch participation, and expenditure per pupil.

Ethnic Composition

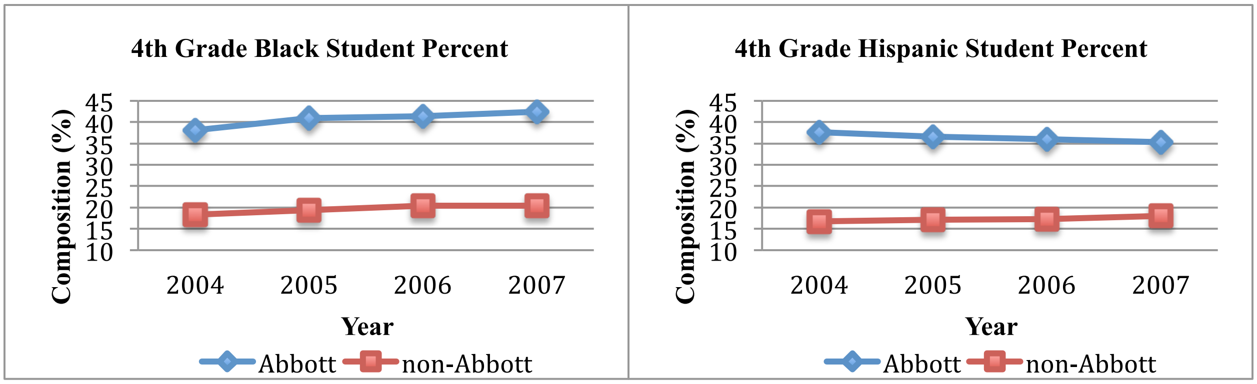

Figure 7: Descriptive Ethnic Composition Data

In the Abbott districts, the Black student composition fell 2.4 percentage points from 37.7% to 35.3% and the Hispanic student composition increased 4.4 percentage points from 38.1% to 42.5%. In non-Abbott districts, the Black student composition increased 2.2 percentage points from 16.8% to 18.0% and the Hispanic student composition increased 2.1 percentage points from 18.3% to 20.4%.

Relatively, Abbott districts lost Black students and non-Abbott districts gained Black students. Also, both district groups gained Hispanic students. Thus, the percentage of underprivileged minority students increased by 2.0 percentage points in Abbott districts and by 4.3 percentage points in non-Abbott districts. Since underprivileged minority students score lower on test than their white and Asian counterparts, this can help account for why achievement in the non-Abbott districts is decreasing faster in language arts and increasing slower in math compared to the Abbott districts.

Free and Reduced Lunch Participation

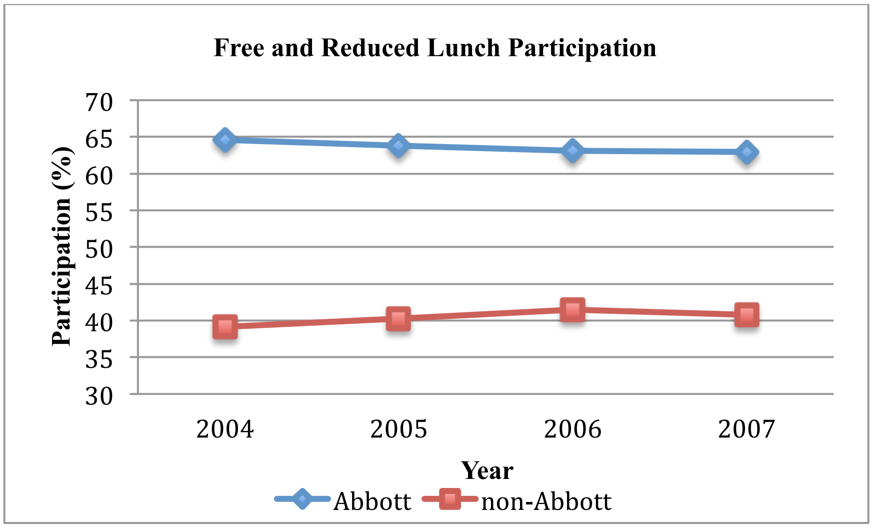

Figure 8: Descriptive Free and Reduced Lunch Participation Graph

The free and reduced lunch participation data reveals that that Abbott districts are on average socioeconomically poorer that the non-Abbott districts, but that the gap was closing. Between 2004 and 2007, the participation rate in Abbott districts decreased 1.6 percentage points from 64.6% to 63.0% and that of the non-Abbott districts increased 1.7 percentage points from 39.1% to 40.8%. Since socioeconomic characteristics are positively correlated with student achievement, these changes might also help explain the Abbott districts' relatively improved test results and thus it is important to control for them.

Expenditure

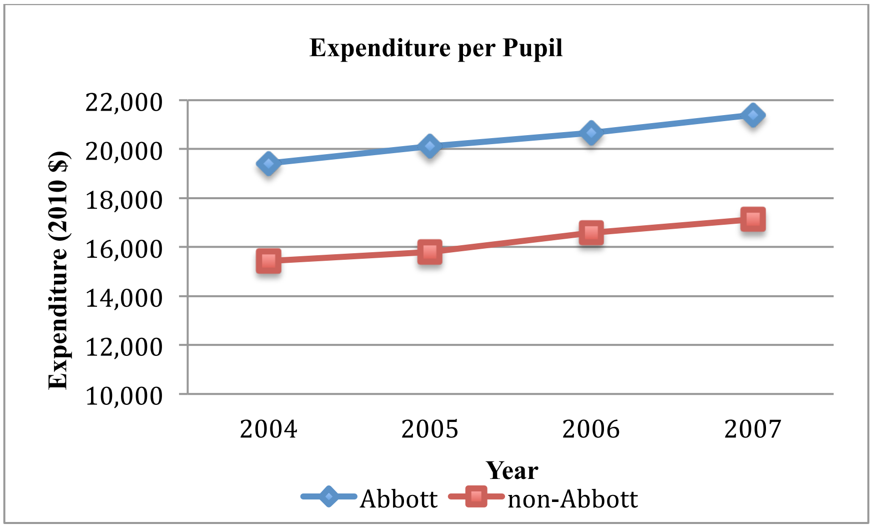

Figure 9: Expenditure per Pupil Graph

The expenditure per pupil data show that Abbott districts continually spent more on education than the non-Abbott districts. In Abbott districts, spending increased by $1,987 from $19,422 to $21,409 and in non-Abbott districts, spending increased by $1,708 from $15,424 to $17,132.[32] In percentage terms, Abbott district spending increased by 10.2% and non-Abbott district spending increased by 11.1%, and in 2007, Abbott district spending was 25.0% higher than non-Abbott district spending. So although the non-Abbott districts increased their spending by a proportionally higher amount, the Abbott districts had a higher absolute spending and absolute increase. Since more money generally means more educational opportunities, the increased spending by both the Abbott and non-Abbott districts should have positively affected student achievement if all other factors stayed constant.

[32] All the monetary values are in 2010 inflation-adjusted dollars.

Regression Analysis Disentangle Effects

Given these disparate changes in student characteristics and district spending, the effect of the *Abbott V* funding reallocation on student achievement is not directly apparent. The two regression models control for the differences and changes in these variables to reveal the effect on successive classes of fourth graders with the status model and the long-term effect on classes of students as they move through the grades with the growth model.

Status Models Fail to Demonstrate Significance

Regression Number	(1^)	(1)	(2^)	(2)	(3^)	(3)	(4^)	(4)
Dependent Variable	**Language Arts Proficiency**	**Language Arts Proficiency**	**Language Arts Average**	**Language Arts Average**	**Math Proficiency**	**Math Proficiency**	**Math Average**	**Math Average**
Abbott Indicator	**-9.971*****	-0.651	**6.486*****	-0.690	**-9.671*****	-0.531	**-7.367*****	0.849
	(2.070)	(1.849)	(1.272)	(1.122)	(2.359)	(2.179)	(2.162)	(1.996)
Year	**-1.095***	**-0.899***	**-0.830***	**-0.738****	**5.230*****	**5.373*****	**4.499*****	**4.578*****
	(0.601)	(0.489)	(0.369)	(0.297)	(0.685)	(0.239)	(0.628)	(0.527)
Abbott-Time Interaction	0.698	0.156	0.416	0.0706	0.806	0.239	0.189	-0.358
	(1.108)	(0.895)	(0.681)	(0.543)	(1.263)	(1.055)	(1.157)	(0.966)
Control Variables								
Black Students (%)		**-0.171****		**-0.0972*****		**-0.203*****		**-0.169*****
		(0.0282)		(0.0171)		(0.0333)		(0.0305)
Hispanic Students (%)		0.0400		0.0232		0.0489		0.0522
		(0.0303)		(0.0184)		(0.0357)		(0.0327)
Free/Reduced Lunch (%)		**-0.264*****		**-0.181*****		**-0.251*****		**-0.260*****
		(0.0434)		(0.0263)		(0.0511)		(0.0468)
Expenditure per Pupil		7.64E-6		6.92e-5		9.35e-05		**1.86e-4***
		(1.02e-4)		(6.17e-5)		(1.20e-4)		(1.10e-4)
Constant	**75.55*****	**87.96*****	**211.5*****	**218.8*****	**65.36*****	**76.33*****	**214.1*****	**223.4*****
	(1.119)	(2.054)	(0.688)	(1.247)	(1.275)	(2.421)	(1.169)	(2.217)
Observations	407	407	407	407	407	407	407	407
R-squared	0.121	0.435	0.139	0.459	0.238	0.475	0.205	0.453

Regression coefficients shown with standard error in parentheses.
^ indicates uncontrolled regression
***** p<0.01; ** p<0.05, * p<0.1**

Figure 10: Status Model Regression Results

I first conducted the status model regressions for the four student achievement metrics without the student characteristics and spending control variables.[33] The Abbot indicator variable has significant negative coefficients in all four regressions which confirms the earlier descriptive results that show that Abbott districts perform worse in all measures relative to their peer districts. Also, the time variable has significant negative coefficients for the two language arts metrics and significant positive coefficients for the two math metrics. This is also consistent with the descriptive data that shows that language arts achievement is decreasing while math achievement is increasing in both district groups. Finally, since the Abbott-time interaction variable is not significant in any of the regressions, the Abbott districts did not exhibit any changes over time that were different from that of non-Abbott districts when the control variables are not incorporated.

After accounting for the control variables in the regressions, the Abbott indicator variable is no longer significant for any of the metrics, but both free and reduced lunch participation and Black student composition are significant.[34] That both are negatively correlated with student achievement means that students in socioeconomically poorer districts and in districts with relatively more Black students exhibit lower achievement. These three factors show that student composition can explain why student achievement is lower in Abbott districts: relative to the non-Abbott districts, the Abbott districts have more challenging students who exhibit lower achievement.

[33] These are regressions 1^, 2^, 3^, and 4^.
[34] These are regressions 1, 2, 3, and 4.

Furthermore, district spending is marginally significant and positively correlated for the math average score regression.[35] That expenditure is important for math achievement but not language arts achievement is consistent with the education research theory that students develop math and language arts skills differently. Elementary math skills are more linked to school factors than language arts skills because it is possible to drill elementary math skills into students while language arts skills are more dependent on home and family environment (Figlio and Ladd 2008, p. 177). Therefore increased funding can create improvements in math achievement more easily than in language arts achievement. Since expenditure is significant in math average score but not math proficiency rates, students who scored above the proficiency cutoff point exhibit this improvement.

Finally, the Abbott-time interaction variable is not significant in the controlled regressions. Even after accounting for the Abbott districts having more challenging students, student achievement did not increase in the Abbott districts relative to the non-Abbott districts. That means that the *Abbott V* reallocation of funds did not increase student achievement for fourth grade students in the Abbott districts.

[35] This is regression 4. It is only marginally significant at the 90% confidence level while all the other significant variables are significant at least at the 95% level.

Growth Models Show Lower Gains for Students Over Time

Regression Number	(5^)	(5)	(6^)	(6)	(7^)	(7)	(8^)	(8)
Dependent Variable	**Language Art Proficiency Change**	**Language Art Proficiency Change**	**Language Arts Average Change**	**Language Arts Average Change**	**Math Proficient Change**	**Math Proficient Change**	**Math Average Change**	**Math Average Change**
Abbott Indicator	**-6.830*** **	**-7.051*** **	**-3.574*** **	**-3.660*** **	**-8.435*** **	**-9.079*** **	**-9.688*** **	**-10.38*** **
	(-1.549)	(-1.585)	(-0.898)	(-0.917)	(-2.033)	(-2.092)	(-1.895)	(-1.926)
Control Variables								
Black Student Change (PP)[36]		-0.0869		-0.0911		-0.202		**-0.294***
		(-0.137)		(-0.079)		(-0.18)		(-0.166)
Hispanic Student Change (PP)		-0.052		-0.0797		0.0177		-0.0396
		(-0.126)		(-0.0727)		(-0.166)		(-0.153)
Free and Reduced Lunch Change (PP)		-0.08		-0.0413		-0.148		-0.13
		(-0.0817)		(-0.0473)		(-0.108)		(-0.0993)
Expenditure per Student Change ($)		**-5.74e-4****		**-2.90e-4****		-5.03e-4		**-7.12e-4****
		(-2.35e-4)		(-1.36e-4)		(-3.1e-4)		-(2.86e-4)
Constant	**2.198*** **	**3.361*** **	**3.371*** **	**4.040*** **	**-4.932*** **	**-3.576*** **	**-5.231*** **	**-3.512*** **
	(-0.84)	(-0.956)	(-0.487)	(-0.553)	(-1.102)	(-1.261)	(-1.028)	(-1.161)
Observations	204	204	204	204	204	204	204	204
R-squared	0.088	0.12	0.073	0.108	0.079	0.101	0.115	0.158

Regression coefficients shown with standard error in parentheses.
^ indicates uncontrolled regression
***** p<0.01; ** p<0.05, * p<0.1**

Figure 11: Growth Model Regression Results

I also first conducted the growth model regressions without the control variables.[37] The significant and negative correlation of Abbott indicator variable in all four regressions shows that students in Abbott districts demonstrated lower growth between fourth grade and eighth grade. The significance remains even when the regressions account for changes in student characteristics and district spending.[38] This lower growth trend manifested in both subjects of language arts and math and measured both as proficiency rate and average score.

[36] PP stands for percentage point
[37] These are regressions 5^, 6^, 7^, and 8^.
[38] These are regressions 5, 6, 7, and 8.

While the descriptive results and the status models show that students in the Abbott districts were not worse off than their peers in non-Abbott districts at the fourth grade level, the growth models reveal this was not the case as the students moved up to the eighth grade. Indeed, the pseudo-cohorts of fourth graders were not making relative progress as they grew older.

There are two other significant variables in these regressions. First, expenditure per student has a significant negative coefficient for both measures of changes in language arts test results and for math average change. This seems to suggest that increases in per student spending led to lower growth in those three test metrics. However, this may be misleading: the expenditure per student data came from the district level with all the grades aggregated. Therefore, an increase in the value of this variable does not mean that all students were the recipients of more spending. For example, a district might increase spending mainly on the high school level, yet since the data are aggregated, the value of the expenditure per student variable would increase for all grades. With New Jersey's focus on early childhood education, increased spending could have gone to the early elementary grades of PreK-3rd and not to the students in this analysis. Since the expenditure data is only available on the aggregate level, the statistically significant and negative coefficients of the expenditure per student changes do not necessarily mean that higher spending led to lower growth.

Second, the change in black student composition has a marginally significant negative coefficient for math average change that suggests lower growth in the math average scores in districts in which more Black students enrolled. This is consistent with the achievement gap trends, but interestingly, that same variable is not significant for the math proficient rate. That means that there was lower math growth in the overall distribution of students but not around the proficient and not proficient cutoff level.

The non-significance of the free and reduced lunch variable in these growth models indicates that changes in the socioeconomic characteristics of the districts did not lead to differing growth rates in any of the test metrics. Although the data for this variable came from the district level, unlike the expenditure per student variable, the aggregate data is representative of each grade because it is unlikely that within the same year students in different grades would have dramatically different socioeconomic characteristics. While the status models reveal that socioeconomic characteristics are significant for the snapshots of test results in four grade – specifically that districts with lower socioeconomic status had lower results on all test metrics – the growth models show that the change in test results from four grade to eighth grade was not affected by socioeconomic status. In other words, districts with differing socioeconomic statuses had statistically lower absolute test results but did not have statistically different growth rates.

VI. Conclusion and Policy Implications

The results cast doubt on the efficacy of the *Abbott V* court-mandated state aid redistribution. They contradict previous data that concludes that the Abbott districts had significantly improved since the 1998 decision. Rather than starting in 1998, this analysis starts in 2004 because the fourth grade class that year was the first class of students to have started kindergarten after the implementation of the *Abbott V* reforms in 1999. The fourth grade test result trends between 2004 and 2008, the longest period of comparable tests, showed that the Abbott districts had made progress relative to their peer districts. However, after controlling for the differences in student characteristics and district spending, there is no significant relationship between the *Abbott V* treatment and student achievement. That is, the analysis implies no significant effect, either positive or negative, on language arts and math achievement in the fourth grade. Moreover, tracking the growth of groups of student pseudo-cohorts over time – rather than examining snapshots of groups of students as they entered fourth grade – reveal that students in Abbott districts had lower relative gains in achievement as they advanced from fourth grade to eighth grade. This is evident in both language arts and math achievement measured as both the percentage of students scoring proficient and average district score.

The results yield several implications for future approaches to education policy. They primarily lend support for New Jersey's decision to eliminate the Abbott district system under the School Funding Reform Act because it removes the court-mandated funding reallocation within the individual districts of *Abbott V* yet preserves the state aid funding increases of *Abbott II*. That the reallocation of state aid to best practice programs in the Abbott districts were ineffective in increasing student achievement warrant their removal. In addition, the hold harmless clause that stipulates that no district will receive less state aid under the new system allows the former Abbott districts to maintain their spending levels. The status model regressions and prevailing education research agree that increased spending creates better student achievement, especially in math.[39] In addition, by moving away from the Abbott district system, the state now has more flexibility with the discretionary distribution of aid to all districts. This is helpful given the state's changing demographic patterns, specifically the increasing enrollment of disadvantaged Black and Hispanic students in non-Abbott districts. In short, the School Funding Reform Act eliminates the ineffective funding for best practice reforms in the Abbott districts while maintaining their increased state aid and creating the flexibility for the state to adapt to demographic changes.

The results also have implications for other policymakers considering funding best practice reforms. First, the decision to reallocate funds needs to consider the organizational capacity of the districts and schools. The Abbott districts were unprepared to implement the array of 17 reforms mandated by the court and failed to launch according to the implementation timetable. That some of the reforms like full-day kindergartens were not operational in some Abbott districts even three years after the decision show that policymakers need to evaluate the capacity and expertise of the districts that they want to reform. Second, the best practice reforms themselves may not be effective for increasing student achievement. Reforms that make sense in theory may not work properly when applied or in all districts and environments. Mandating a couple of suggested reforms from a menu may be more manageable and executable than

---

[39] That the growth model regressions reveal a marginally significant negative coefficient for spending is discounted because of issues with the data availability and funding distribution across grade levels. See page 34.

mandating a broad array of reforms. In addition, policymakers can use the experiences of successful districts and schools to emulate what works.

Although this analysis attempts to be robust, there are methodological limitations imposed by the availability of data. First, while the New Jersey Supreme Court decided the *Abbott V* ruling over 12 years ago in 1998, there was only a limited time span with comparable test data – four years for the status models and two pseudo-cohorts for the growth models. Ideally, the state would have used the same tests year after year and given tests to each grade. Older data from before the 1998 *Abbott V* ruling and more recent data after the 2008 School Funding Reform Act would allow for a comparison of the student achievement trends in the two district groups and a determination of whether the trend was different during the years in which *Abbott V* was in effect. This type of analysis would also be possible if test scores were available at the individual student level so that the results could be normalized to create standard deviation comparisons.[40]

Also, expenditure per pupil data on the grade level rather than district level would have been a more precise metric for school inputs because districts do not spend equally across grade levels. This posed a problem for the growth model regressions because the Abbott district's state aid was allocated to reforms that primarily affect early childhood education. Therefore an increase in the value of the expenditure per pupil data for a district may not reflect an actual increase in school inputs for that district's eighth grade students.

The difficulty of distinguishing which particular district implemented which specific reform when also limited the breadth of this research to an analysis of the effect of the *Abbott V* intent-to-treat on student achievement rather than a more detailed analysis of the specific effects of the individual mandated best practices. Owing to the organizational inability of the Abbott

[40] See page 14 for this method.

districts to follow the court's implementation timetable and the lack of published histories, it is impossible to determine the start year for each reform for each district. In addition, there would not even be a clear start year for districts that implemented a reform incrementally – for example, gradually expanding enrollment at a pre-school instead of starting at fully capacity. The most useful analysis would identify which best practice reforms had the greatest effect, but the messy implementation prevent this degree of specificity.

Finally, a regression analysis determines correlation not causation between variables. The growth model regressions showed a significant and negative correlation between the Abbott indicator variable and student achievement, but that correlation does not imply causation. However, by controlling for four important variables that influence student achievement and by comparing the Abbott districts to their socioeconomic peer districts, the regressions account for confounding factors and strengthen the likelihood of causation.

Although these limitations prevent a quantitative assessment of the effects of *Abbott V*, qualitative research can yield more information to craft further policy implications. Interviews with district and school administrators, teachers, parents, and even the students in the former Abbott districts would provide insight about what was and was not effective about the reforms and about issues with the implementation process. In addition, their feedback would help in understanding the details and nuances this policy, specifically why something worked or did not. Albeit not as firm and objective as statistical data, qualitative information would help elucidate the effects.

While the results from this analysis show that the *Abbott V* court-mandated reallocation of state aid to an array of best practice reforms did not increase student achievement, the pursuit of effective approaches for helping disadvantaged students in socioeconomically poor districts is not over. New Jersey has enacted the School Funding Reform Act that keeps the beneficial aspects of the court-established Abbott district system and eliminates the ineffective ones. Tracking student achievement under this new legislation and conducting further research in education are essential to providing the quality of education that every student deserves in New Jersey.

VII. References

Bifulco, R., & Bulkley, K. (2008). Charter Schools. In H. F. Ladd & E. B. Fiske. (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 425-446). New York: Routledge.

Duncombe, W. D., & Yinger, J. (2008). Measurement of Cost Differentials. In H. F. Ladd & E. B. Fisk (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 238-256). New York: Routledge.

Figlio, D.N., & Ladd, H.F. (2008). School Accountability and Student Achievement. In H. F. Ladd & E. B. Fisk (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 166-182). New York: Routledge.

Goertz, M. E., & Weiss, M. (2009). *Assessing Success in School Finance Litigation: The Case of New Jersey*. New York: Teacher's College, Columbia University.

Grades 3 and 4 New Jersey Assessment of Skills and Knowledge Technical Report. (2008). Retrieved from http://www.state.nj.us/education/assessment/es/njask_tech_report08.pdf.

Greif, A. (2004). Politics, Practicalities, and Priorities: New Jersey's Experience Implementing the Abbott V Mandate. *Yale Law & Policy Review, 22*(2), 615-657.

Khadaroo, S.T. (2009). Blacks' Test Scores Lag, But New Jersey is a Bright Spot. *The Christian Science Monitor*, Retrieved from http://www.csmonitor.com/USA/2009/0714/p02s19-usgn.html

Kruger, A. B. (2002). Understanding the magnitude and effect of class size on student achievement. In L. Mishel & R. Rothstein (Eds.), *The Class Size Debate* (pp. 7-35). Washington D.C.: Economic Policy Institute.

MacInnes, G. (2005). *Student Achievement in the Abott Districts*. Trenton, NJ: New Jersey Senate Education Committee.

Mead, S. (2009). *Education Reform Starts Early: Lessons from New Jersey's PreK-3rd Reform Efforts*: New America Foundation.

New Jersey Department of Education (NJDOE). (2007). *A Formula for Success: All Students, All Communities*. Retrieved from http://www.state.nj.us/education/sff/reports/AllChildrenAllCommunities.pdf.

Picus, L. O., Goertz, M. E., & Odden, A. (2008). Intergovernmental Aid Formulas and Case Studies. In H. F. Ladd & E. B. Fisk (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 257-275). New York: Routledge.

Reardon, S. F., & Robinson, J. P. (2008). Patterns and Trends in Racial/Ethnic and Socioeconomic Academic Achievement Gaps. In H. Ladd & E. Fisk (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 497-516). New York: Routledge.

Rice, J. K., & Schwartz, A. E. (2008). Toward an Understanding of Productivity in Education. In H. F. Ladd & E. B. Fisk (Eds.), *Handbook of Reserach in Education Finance and Policy* (pp. 131-145). New York: Routledge.

Score Interpretation Manual Grades 3-8. (2009). Retrieved from http://www.state.nj.us/ education/assessment/ es/njask/SIM2009.pdf.

Sciarra, D. (2009). Time to Restore, Expand, Abbot Reforms. Retrieved from http://www.edlawcenter.org/ELCPublic/elcnews_091015_TimeToRestoreAbbottReforms .htm.

School Funding Reform Act. (2008). Retrieved from http://www.njleg.state.nj.us/2006/Bills/A0500/500_I2.PDF.

Walker, E. M., & Gutmore, D. (2000). *The Quest for Equity and Excellence in Education: A Study of Whole School Reform in New Jersey Special Needs Districts*. South Orange: Department of Educational Administration and Supervision, Center for Urban Leadership, Renewal and Research, Seton Hall University.

Walker, E. M., & Gutmore, D. (2002). The Issue of Civic Capacity in Urban Educational Reform: The Case of New Jersey's Thirty Poorest Districts. *The Journal of Negro Education, 71*(1/2), 60-76.

VIII. Appendices

Appendix 1: List of *Abbott V* remedies

1. Standards-Based Reform
Per-pupil spending parity with suburbs
Standards-based curriculum, assessment, instruction, and professional development
Elementary whole school reform
School-based budgeting
Site-based management including parents and teachers
Standards-based accountability system
Middle and high school whole school reform

2. Early Childhood Education
Full-day Kindergarten
Half-day Preschools

3. Social and Health Services
On-site health and social services for schools in communities with insufficient access
After school and summer nutrition programs
Dropout prevention specialists at middle and high schools
School-to-work and college-transition programs

4. Facilities Improvements
Correcting code violations and overcrowding
Class sizes of 15 in pre-K, 21 in K-3, and 23 in 4-5
Improve school security

Appendix 2: List of non-Abbott Districts

Non-Abbott District Name	DFG in 2004
Atlantic City	A
Bass River Township	B
Bayonne	B
Bellmawr	B
Berkeley	B
Beverly	B
Bradley Beach	B
Brooklawn	B
Buena Regional	A
Carteret	B
Chesilhurst Borough	A
Clayton	B
Clementon	B
Commercial Township	A
Deerfield	B
Dover Town	B
Downe Township	A
Eagleswood	B
East Newark	A
Egg Harbor	B
Fairfield Township	A
Fairview	B
Glassboro	B
Guttenberg	B
Haledon	B
Hammonton Township	B
Keansburg	A
Lakehurst	B
Lakewood	B
Lawnside	B
Lawrence Township	A
Linden	B
Little Egg Harbor	B
Lodi	B
Lower Alloways Creek Town	B
Lower Township	B
Magnolia	B
Manchester Township	B
Mannington Township	B
Maurice River Township	B
Middle Township	B
Mount Holly Township	B
Mullica Township	B
National Park Borough	B
North Bergen	B
North Wildwood City	A
Ocean Gate	B
Ocean Township	B
Paulsboro	A
Penns Grv Carnets Pt Reg	B
Pine Hill	B
Prospect Park	B
Quinton Township	B
Riverside Township	B
Runnemede Borough	B
South Hackensack	B
South River	B
Stafford Township	B
Swedesboro-Woolwich	B
Tuckerton Borough	B
Union Beach	B
Upper Deerfield Township	B
Wallington	B
Washington Township	B
Weehawken	B
Westville	B
Wildwood City	A
Wildwood Crest	B
Winefield Township	B
Woodbine	A
Woodbury	B
Woodland Township	B
Woodlynne Borough	B

Appendix 3. Charter Schools in New Jersey

Name[41]	Grade(s)	District	DFG A/B[42]	Abbott[43]	Year Opened[44]	Enrollment
Academy CHS	9-12	Multiple	Y	Y	1998	220
Academy for Urban Leadership CHS	9	Perth Amboy	Y	Y	TBD	100
Adelaide L. Sanford CS	K-5	Newark	Y	Y	2007	288
Bergen Arts and Science CS	K-8	Bergen	N	N	2007	540
Burch CS of Excellence	K-1	Irvington	Y	Y	2008	108
Camden Academy CHS	9-12	Camden City	Y	Y	2001	500
Camden's Pride CS	K-1	Camden City	Y	Y	2008	84
Camden's Promise CS	6-8	Camden City	Y	Y	1998	300
Capital Preparatory CHS	9-10	Trenton/Ewing	Y	Y	2008	200
CREATE CS	9-12	Jersey City	Y	Y	2001	500
Central Jersey Arts CS	K-8	Plainfield	Y	Y	2006	282
Central Jersey College Prep CS	7-12	Multiple	Y	Y	2006	336
CHS for Environmental and Civic Studies	9-12	Teaneck	N	N	TBD	120
Charter Tech HS for the Performing Arts	9-12	Multiple	N	N	1999	250
Classical Academy CS of Clifton	6-12	Clifton	N	N	1998	260
Community CS of Paterson	K-3	Paterson	Y	Y	2008	500
D.U.E. Season CS	K-7	Camden City	Y	Y	2005	500
Discover CS	4-8	Newark	Y	Y	1999	75
East Orange Community CS	K-4	East Orange	Y	Y	1998	600
Elysian CS of Hoboken	K-8	Hoboken	Y	Y	1997	270
Emily Fisher CS of Advanced Studies	6-12	Trenton	Y	Y	1998	245
Englewood on the Palisades CS	K-6	Englewood City	N	N	1998	240
ECO CS	K-4	Camden City	Y	Y	2005	245
Excellence CS	K-2	Camden City	Y	Y	TBD	324
Foundation Academy CS	5-6	Trenton	Y	Y	2007	160
Freedom Academy CS	5-8	Camden City	Y	Y	2004	320
Galloway Community CS	K-8	Galloway	N	N	2000	405
Gray CS	K-8	Newark	Y	Y	2000	261
Greater Brunswick CS	K-8	Multiple	Y	Y	1998	180
Greater Newark CS	5-8	Newark	Y	Y	2000	112
Hatikvah International Academy CS	K-2	East Brunwick	N	N	TBD	108
Hoboken CS	K-12	Hoboken	Y	Y	1998	300
Hoboken Dual Language CS	K-2	Hoboken	Y	Y	TBD	132
Hope Academy CS	K-8	Asbury Park	Y	Y	2001	135
Institute for Excellence CS	K-2	Winslow	N	N	2009	240
International CS of Trenton	K-4	Trenton	Y	Y	1998	85
Jersey City Community CS	K-8	Jersey City	Y	Y	1997	594
Lady Liberty Academy CS	K-8	Newark	Y	Y	2001	444
LEAP Academy Univeristy CS	K-12	Camden City	Y	Y	1997	702
Learning Community CS	K-8	Jersey City	Y	Y	1997	314

41 CS stands for Charter School; CHS stands for Charter High School
42 Whether the charter school was DFG A or B in 2001
43 Whether the charter school was in a Abbott district in 2001
44 Some schools are approved but have not opened. I labeled them with "TBD."

Liberty Academy CS	K-8	Jersey City	Y	Y	1999	450
Maria L. Varisco-Rogers CS	4-8	Newark	Y	Y	1999	144
Marion P. Thomas CS	K-8	Newark	Y	Y	1999	540
New Horizons Community CS	K-5	Newark	Y	Y	1999	500
Newark Educators Community CS	K-3	Newark	Y	Y	2009	200
Newark Legacy CS	K	Newark	Y	Y	TBD	90
North Star Academy CS of Newark	K-12	Newark	Y	Y	1997	488
Oceanside CS	PreK-8	Atlantic City	Y	N	1999	365
PACE CS of Hamilton	K-3	Hamilton	N	N	1999	120
Paterson CS for Science and Technology	6-11	Paterson	Y	Y	2004	200(?)
Paul Robeson CS for the Humanities	4-6	Trenton	Y	Y	2008	180
PleasanTech Academy CS	K-8	Pleasantville	Y	Y	1998	681
Pride Academy CS	5-8	Multiple	Y	Y	2008	240
Princeton CS	K-8	Princeton	N	N	1997	280
Princeton International Academy CS	K-2	Multiple	N	N	TBD	170
Queen City Academy CS	K-8	Plainfield	Y	Y	2000	177
Red Bank CS	K-8	Red Bank	N	N	1998	162
Renaissance Regional Leadership CS	K-8	Multiple	N	N	TBD	138
Ridge and Valley CS	K-8	Multiple	N	N	2004	135
Riverbank CS of Excellence	K-2	Florence	N	N	2009	108
Robert Treat Academy CS	K-8	Newark	Y	Y	1997	450
Schomburg CS	K-5	Jersey City	Y	Y	2000	500
Soaring Heights CS	K-8	Jersey City	Y	Y	1997	184
Sussex County CS of Technology	7-8	Sparta	N	N	1997	100
TEAM Academy CS	5-8	Newark	Y	Y	2002	320
Teaneck Community CS	K-8	Teaneck	N	N	1998	240
The Barack Obama Green CHS	9-10	Plainfield	Y	Y	TBD	120
The Ethical Community CS	K-1	Jersey City	Y	Y	2009	120
Trenton Community CS	K-8	Trenton	Y	Y	1997	380
Trillium CS	K-4	Multiple	N	N	TBD	152
Union County TEAMS CS	K-8	Plainfield	Y	Y	2005	180
Unity CS	K-8	Morristown	N	N	1998	105
University Academy CS	9-12	Jersey City	Y	Y	2002	500
University Heights CS	K-5	Newark	Y	Y	2006	240
Village CS	K-8	Trenton	Y	Y	1999	323
Vineland Public CS	K-2	Vineland	Y	Y	2009	108
Visions Academy CHS	9	Newark	Y	Y	TBD	150