THREE ESSAYS ON RESOURCES IN EDUCATION

by

Alexandra M. Resch

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Public Policy and Economics)
in The University of Michigan
2008

Doctoral Committee:

Professor Rebecca M. Blank, Chair
Professor Paul N. Courant
Professor John E. DiNardo
Professor Brian A. Jacob
Professor Jeffrey A. Smith

to my grandfather, Hugh K. Livesay,
in appreciation for his many lessons
on the limitless value of education

ACKNOWLEDGEMENTS

First and foremost I would like to thank my dissertation committee — Becky Blank, Paul Courant, John DiNardo, Brian Jacob, and Jeff Smith. Their mentorship and incredibly constructive criticism and advice have improved this research greatly.

I would also like to thank the members of the informal working group that helped clarify my research goals and provided countless hours of advice, support, and entertainment over the years — Soren Anderson, Brian Cadena, Ben Keys, Brian Kovak and Jim Sallee. I hope we have many working lunches ahead of us, even if they're long distance.

Many other people contributed to the essays in this dissertation. Beth Brooks, Gloria Hancock, Nancy Resch, Peggy Smith, Yut'se Thomas, and Don White contributed data and institutional details that made the two essays on the Abbott reforms in New Jersey possible. In addition to my committee and working group, Sheldon Danziger, Lucas Davis, Jim Hines and many seminar participants contributed helpful comments and critiques.

The third essay is coauthored with Jim Sallee and Paul Courant; this paper would not have been possible without them. We thank Soren Anderson, John Bound, Brian Cadena, Jim Hines, Ben Keys, Mike McPherson, Michael Rothschild, Jeff Smith, Sarah Turner and seminar participants at the University of Virginia and the University of Michigan for their comments and critiques. We also thank Caroline Hoxby and two anonymous referees for excellent comments.

Finally, I'd like to thank all the people who helped make grad school, and especially this last year, a lot more pleasant. The National Poverty Center has given me a happy home on campus for the last four years and I look forward to continuing the friendships I made there. For their constant support and timely distractions, I thank Emily Gustafsson-Wright, Sarah Marsh and Robin Phinney. I thank Regina Benson, Jaime Llewellyn and Karen Munday for still being my friends even though I never have time to call. I thank Zingerman's for feeding me this last year when I haven't had time to cook. Most of all I want to thank my parents, Peter and Nancy, and sister, Ingrid, for their unfailing support and encouragement, and the occasional Zingerman's gift cards.

I would also like to acknowledge generous financial support from the National Science Foundation, Rackham Graduate School, and the Spencer Foundation.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER I

Introduction

The three essays in this dissertation look at the distribution and effects of resources in education. The first two essays examine the effects of the Abbott school finance reforms that increased the resources available to low-wealth school districts in New Jersey. The third essay provides a framework for thinking about the distribution of resources in a system of public higher education.

The first essay is an empirical analysis of the effects of the Abbott school finance reform on educational expenditures in New Jersey. This reform dramatically increased the funding available to poor, urban schools with the goal of improving achievement in those districts. My analysis suggests that districts directed the added resources largely to instructional personnel. They hired additional teachers and support staff.

The second essay asks the obvious next question: Did this increase in funding and spending improve the achievement of students in the affected school districts? I focus primarily on the statewide 11th grade assessment that is the only test that spans the policy change. I find that the policy improves test scores for minority students in the affected districts by one-fifth to one-quarter of a standard deviation.

The third essay, written with James M. Sallee and Paul N. Courant, looks at the optimal provision of public higher education. We provide a simple model that shows that a tiered system of higher education, where students are sorted into schools by ability as we see in the U.S., is optimal with just a few assumptions. Our results are driven by an assumption that ability and resources are complementary.

CHAPTER II

The Effects of the Abbott School Finance Reform on Education Expenditures in New Jersey

2.1 Introduction

Despite a large literature on the effects of resources in education, there is no consensus on the effects of money in schools. Research has been inconclusive because many of the inputs to student achievement are difficult or impossible to measure and also because the observable aspects of schooling are the result of choices of individuals and several levels of government, with different incentives and goals. A recent court case in New Jersey offers the opportunity to shed new light on the topic of how increased state funding for local school districts affects their educational expenditures. In 1997, the fourth ruling in the Abbott v. Burke school finance lawsuit in New Jersey ordered an immediate, unexpected and dramatic increase in funding to 30 poor, urban school districts. This ruling provides a natural experiment which I use to explore the effects of money in schools, and to ask what happens to failing districts when they receive a sudden increase in available funds.

There are several explanations for the lack of clear conclusions about the effect of school spending on student outcomes.[1] One is difficulty in measuring both the inputs and outputs of school systems. The expenditure data is often aggregated to

[1] See Hanushek (1996) for a review of this literature.

the state level, which ignores important variation within states and relies on between-state differences for identification. The appropriate outcome measures are also not always obvious; states and researchers tend to rely on easily available standardized test scores, but not all reforms are meant to raise test scores. Another important factor is omitted variables. Wealthy students tend to go to well-funded schools and have high achievement, but this is in part due to family preferences, choices, and out-of-school investments. Disadvantaged students often receive large amounts of compensatory funding, leading to high levels of aggregate spending, but still do poorly in terms of measured achievement. The institutional structure of individual schools is also typically omitted from these studies, but there are large differences in the efficiency with which different school districts use their resources. In this study I provide evidence on the effect of the finance reform both on aggregate expenditures in the affected districts and on specific types of spending within schools. This allows me to predict what student outcome measures are likely to be affected and to conclude that the commonly used math and English test scores are unlikely to tell the whole story in this case.

There is a consensus that poor and urban schools have worse inputs and outcomes, even when their schools are well funded. For example, schools serving poor students tend to have less experienced teachers and are more likely to have high year-to-year teacher turnover.[2] In this literature it is hard to convincingly disentangle the contribution of schools to achievement from the contributions of families and other external factors. To the extent that the Abbott reform changes the circumstances of schools without affecting students' families, the analysis in this paper makes progress on this puzzle.[3] Even without the strong assumption that the Abbott reforms affected

[2]See Hanushek, Kain, and Rivkin (2004) or Reed, Rueben, and Barbour (2006).
[3]It may be too strong an assumption to say that the reforms didn't affect families, but this is a plausible assumption in the short term.

only the in-school environment, a careful examination of the particular expenditure changes allows me to identify outcome measures that offer a more nuanced view of the effects of resources on student outcomes.

Through school finance reform, there has been a general trend since the 1960s toward equalizing funding across districts within states. The vast majority of states have had legal challenges to their school finance systems and most have instituted significant reform whether or not their system was found unconstitutional.[4] Murray, Evans, and Schwab (1998) look at the effects of school finance litigation across the U.S. and show several broad effects. Court-ordered reforms reduced intrastate inequality in school spending, and did this by increasing aggregate education expenditures. Hoxby (2001) argues that, while within state inequality is reduced in all equalization reforms, the vast majority of states have lower mean spending than they would in the absence of equalization plans. In general, poorer districts receive more funding as the result of these reforms because the changes weakened the link between local wealth and school funding.

The Abbott reform is unlike the reforms in other states. Because the state legislature resisted the court's mandate to increase funding to urban schools for more than 10 years, the court eventually instituted a new regime only for the districts covered by the lawsuit. This ruling gave immediate increases in state aid to the 30 Abbott districts but left the most recent school finance law in place for all other districts. This feature allows me to use the funding mandated by the court ruling as an instrument for state aid to isolate and identify the effects of increased aid to poor districts. The policy induced a significant change in funding to Abbott districts, amounting to about 10% of average per-pupil spending statewide. Relying on

[4]See Minorini and Sugarman (1999); Evans, Murray, and Schwab (1997); Hoxby (2001); West and Peterson (2007) and Hanushek (2006) for discussions of school finance litigation.

within district variation over time and using longitudinal data from the Census of Local Governments, I find that 60-70% of the increased funding goes to increased educational expenditures. This fraction is high when compared to the predictions of public finance theory, but it is similar to other empirical estimates of the flypaper effect in school funding.[5]

The earlier studies can rarely look at where added dollars were spent inside schools, but understanding the value of having the money "stick" in schools requires an understanding of both how much money was retained and how the schools used it.[6] I have put together data from multiple sources and have an unusually detailed dataset which I use to examine how Abbott districts spent the additional aid. The increased spending is focused on current K-12 expenditures, with about equal amounts going to instruction and supplemental services. These categories seem like the most relevant for improving student outcomes, but do not inform us about how the classroom experience of students changes with the added funding. Using data on school personnel from the New Jersey Department of Education, I find that Abbott districts increase hiring in several categories relative to a comparison group of districts. Abbotts hire more teachers, tutors, and counselors. In high schools, Abbott districts increase the numbers of music and art teachers, but seem to employ no additional math and English teachers. This suggests that Abbott districts are focusing on broader education programs and not focusing only on the core subject areas covered by No Child Left Behind, as reports have suggested about national trends.[7] I find some evidence that districts adjust their spending over time. The instructional expenditures scale up

[5] Hines and Thaler (1995) review the flypaper literature and report estimates of 40 to 58% for flypaper effects in studies of effects of state aid to school districts.

[6] This is not a new idea in the education literature. See Cohen, Raudenbush, and Loewenberg Ball (2003). The economic literature on school finance reform, however, tends to look directly from the finance change to student outcomes without considering the precise uses of funds.

[7] See, for instance, Center on Education Policy (2007), which reports that instructional time for math and English has increased since 2002 and instructional time for subjects not tested under NCLB has decreased.

over several years, while Abbott districts direct money to capital spending only in the first year of the policy.

Measuring the effects of the Abbott rulings is important because they provide an example of one policy option available to states looking to improve education in failing districts: increased resources. The Abbott ruling provided both dramatic funding increases and specific mandates[8] for the uses of those funds. These policies improved the resources available to the affected districts, but the state has done no credible evaluation of the results. In fact, the state currently has no mechanism in place to allow for a credible evaluation.

The Abbott ruling, by creating an exogenous change in state aid, also allows a clean test of what happens when states increase aid to school districts. Unlike many other cases, the Abbott reform lets us identify the effects on low-wealth districts specifically. In practice, the bulk of state aid is targeted to poor districts so these are exactly the districts whose behavior we want to understand.

The remainder of the paper is organized as follows: Section 2 discusses the Abbott case, the school finance system in New Jersey and my data. Section 3 reviews the literatures on resources in schools, the flypaper effect in school finance and previous work on the Abbott case. Section 4 presents the empirical approach and results and Section 5 concludes.

2.2 Background: The Abbott Reform and School Finance in New Jersey

This section discusses the changes to New Jersey school finance over the last 30 years and the current system as it applies to Abbott districts and all other school

[8]A subsequent ruling in 1998 required additional funding to support specific mandated programs, including phased-in universal preschool for three and four year olds and the implementation of whole school reform. Funding also increased to cover the costs of these prescriptions.

districts. I then describe the data I use to look at the effects of the Abbott reforms and the selection of a comparison group of school districts used in the empirical analysis.

2.2.1 Legal Challenges to NJ School Finance

Until 1976, schools in New Jersey were financed solely through local property taxes.[9] As in many states, this led to large disparities in funding levels depending on local wealth. In 1970, the Robinson v. Cahill lawsuit challenged property tax financing on the grounds that urban schools could not afford the disproportionately high tax burden required to provide the "thorough and efficient" education mandated by the state constitution. In 1973 the New Jersey Supreme Court agreed, leading the legislature to pass the 1975 Public School Education Act (PSEA). Among other things, the PSEA instituted a new funding formula to be supported by a state income tax which, after significant contention between the legislature and the courts, was instituted in 1976. While PSEA changed the source of funds, it did not successfully reduce the funding disparity between rich and poor districts. State legislators from this period suggest that local governments chose to use the income tax inflows to reduce local property taxes without increasing school budgets (Comments at public forum, May 23, 2006).

Abbott v. Burke was filed in 1981 on behalf of a group of students in four urban New Jersey school districts. The suit argued that the new state financing law failed to remedy the existing funding disparity and that these students were still not receiving an adequate education. In 1985, the court found in the students' favor, noting that wealthy districts spent 40% more than poorer districts; the court also expanded the districts covered by the suit to a group of 28 poor urban school districts, which

[9] This ignores Federal Aid, which was 9% of K-12 funding nationally in 1976 and is only about 4% of school funding in New Jersey throughout the 1990s. (Digest of Education Statistics, 2006, Table 158.)

are now known as the Abbott Districts.[10] The first factor in designating a district an Abbott district was that they must fall into one of the two poorest categories in the state's SES grouping scheme. The second factor was "persistent educational failure." In practice, this was measured by high failure rates on the state assessment tests which had been instituted in PSEA. The Abbott districts shared several other characteristics that were noted in defining the group: high local tax rates, high percentage of non-white students and a large percentage of special needs students.

Abbott II (1990) was the first decision to impose a specific goal for the state to meet. The ruling ordered parity in foundation funding between the Abbott districts and wealthy suburban districts and required additional programs to address special needs in Abbott districts. The court ruling defined parity in terms of "regular education" expenditures and noted that the Abbott districts also have other needs which require more funding than wealthier districts. This parity definition remains the benchmark for funding in the Abbott case.

The early legislative responses attempted to legitimize the (relatively) low levels of spending in the urban districts. The state passed several laws intended to satisfy the court mandate without significantly changing the amount of funding for Abbott districts. Enacted in 1990, the Quality Education Act (QEA) set a "foundation amount" of funding that represented the level required to provide the mandated education to all students. The court found in Abbott III (1994) that the QEA was unconstitutional because the funding amounts had no relationship to the actual cost of providing an education or the differences between urban and suburban districts. The court overruled this legislative action but gave the state until the 1997-98 school year to comply with the previous rulings requiring equivalence in funding. In 1996,

[10]Two additional Abbott districts, Neptune and Pemberton, were added in 1998 and one, Salem City, in 2004, bringing the current total to 31. Since Salem City was added in 2004 and my data ends there it is not considered an Abbott district in my analysis.

the state responded by enacting the Comprehensive Education Improvement and Financing Act (CEIFA). This new act added clear standards for what constitutes a thorough education in terms of defined educational standards and instituted testing to assess compliance with the standards. However, the CEIFA definition of adequate funding was based on a model school district that did not reflect the needs of Abbott districts and arrives at funding values very close to the figures in the QEA. In both laws, the funding values were not related to the education standards but seemed to be based on the past spending levels.

The court found in Abbott IV (1997) that the funding levels set in CEIFA were chosen arbitrarily and based on no careful study of what is necessary or adequate to reach the goal of providing a thorough education. The Abbott IV ruling specifically required that the state undertake a study to determine what is necessary to provide an appropriate education in the Abbott districts and how much this will cost. Later rulings mandated additional programs, including building improvements, universal pre-school and specific whole school reform programs. These additional programs restricted districts' choice in the use of funds, but districts still had the option of reevaluating and changing their existing spending patterns as these new funds and mandates were added.

Most importantly, Abbott IV instituted "parity aid" to bring per-pupil regular education funding in the Abbott districts to the level of funding in the wealthiest school districts. The specific amount of aid a district gets is determined by the gap between that district's regular education funding and the average regular education funding of the group of wealthy districts.[11] In addition to this parity aid, poor schools in general and the Abbotts specifically get significant additional funding from

[11] Regular education funding is defined as the foundation amount of funding plus optional local levies.

the state and federal governments based on the characteristics of the local student population.

The parity aid is the portion of funding that can potentially be thought of as exogenous. The state had successfully held off the court mandate for more than 15 years, and while funding to Abbotts began to increase in 1992 as a result of QEA, the funding changes starting in 1997-98 were dramatically larger. The state passed a new funding law, CEIFA, by the 1997 deadline set by the court and could have reasonably expected another round of litigation but not an immediate funding mandate. To demonstrate the magnitude of the mandate, Figure 2.1 shows average parity aid over time. The parity aid adds an average of $1000 per pupil to the budgets of the Abbott districts in 1998 and this increases to nearly $1500 per pupil in 2004. This is a significant budget increase for affected districts, representing roughly 10% of mean total expenditures per pupil in New Jersey.

2.2.2 The Current Finance System in NJ

Although the 1996 school funding law, CEIFA, was challenged in court and overturned with respect to the Abbott districts, the basic funding formula from this law remained in effect through 2007-08 for all other districts.[12] The foundation budget, which represented 53% of the total state education budget in 1999, sets a base amount of per pupil funding for the entire state. A district's actual funding weights this amount by the grade level of students, where elementary students are assigned a weight of 1.0, and middle and high school students are weighted more heavily. This is the foundation amount meant to support the core curriculum standards laid out in CEIFA. In 1999, the base amount was $6,899 per student and this is indexed to the CPI. The source of funds (state vs local) depends on local property wealth. Poorer

[12] A new finance law way passed in January 2008 which aims to end the Abbott designation and separate funding scheme. The NJ Supreme Court has not yet ruled on the constitutionality of this law.

districts get a much larger portion of their budget from the state, while wealthier districts pay more locally. In addition to the foundation budget, districts receive "Adjustments for Special Factors," which are largely made up of various types of aid to poorer districts plus the Abbott parity aid. Other budget categories such as transportation and special, bilingual, and early childhood education are added on to the foundation budget. These ajustments are not included in the parity calculation because they are not part of the "regular education" budget.

Figure 2.2 is an illustration of the basic finance system in place in New Jersey, with district property wealth on the X-axis and per pupil "regular education" spending on the Y-axis. The flat solid line is the foundation amount. In New Jersey this is calibrated to the cost of providing a "thorough and efficient education" and is referred to as the "T & E budget."[13] Any district can choose to tax itself more heavily to increase spending above the foundation amount.[14] The full amount of regular education spending is the sum of the foundation amount and the optional local levy and this is shown by the upward sloping dashed line. The share of the foundation funding that is raised locally is determined by local property wealth — poorer districts receive more state aid and raise correspondingly less locally.[15] The state share is given by the downward sloping dotted line.

In Figure 2.2, the amount paid by the state to a given district is shown by line segment AB and the amount raised locally is segment BC. The amount above the

[13]This foundation amount is meant to be the amount that supports mastery of the Core Content Curriculum Standards, codified in CEIFA, and these standards also define the current understanding of a "thorough and efficient" education. It is important to note that this benchmark is far above the interpretation of the constitutionally required public education in many other states. For example, the ongoing case in New York, Campaign for Fiscal Equity v. State of New York, refers to a "sound basic education," presumably a lower bar, but the 2006 ruling called for a funding increase of $5 billion per year for New York City alone.

[14]New Jersey does have a spending cap which limits year-to-year growth in district expenditures. This cap can be overturned by local vote and the cap can be appealed if the increase is due to enrollment growth. In practice, actual expenditure growth is significantly higher than the cap, implying that the cap is not actually binding. (Personal communication with NJDOE officials.)

[15]State aid is determined by a formula that incorporates both property wealth and community income, but only wealth is included here to simplify the graph.

T & E budget, line CD, is the optional local levy. Abbott districts tend to be grouped along the low end of the property wealth scale, although some districts have experienced increases in wealth over the years since the Abbott designation was assigned. The other poor districts overlap with the Abbotts and extend to the right, followed by the middle income districts and, at the far right, the wealthy or I and J districts. These wealthy districts are referred to as the "I and J districts" because they fall into the highest, or I and J groups, on the state's SES scale. The calculation of parity aid is shown in the top left corner of the figure. For an Abbott district, the amount of parity aid received in 1997-98 was determined by subtracting their chosen spending level from the average spending for I and J districts.

After 1997, the Abbotts automatically received the regular education budget of the wealthiest districts, regardless of their local contribution. This suggests a clear incentive for Abbott districts to reduce their optional local levy to zero and allow the parity formula to replace this with state aid. To prevent this, New Jersey included a maintenance of effort requirement. The Abbott districts cannot reduce their per-pupil property tax levy below the 1997 aggregate (as opposed to per pupil) level.

2.2.3 Data

I combine annual school district level data from several data sources to create a unique longitudinal dataset to investigate the effects of the Abbott reforms. I construct a panel covering 1990-2004 on school district funding, expenditures and student characteristics. The main financing variables come primarily from the Census of Governments Survey of Local Governments School Systems Survey, the F-33 data. I use the longitudinal file for 1990-2002 which is available from the US Department of Education's National Center for Education Statistics, and append the 2003 and 2004 files. I supplement these data with parity aid figures from the annual New Jersey

State Aid Summaries. I add district-level tax data from the New Jersey Division of Taxation Annual Reports and income and property value information obtained from the NJDOE Finance Department. All financial variables are adjusted for inflation using the Consumer Price Index for all urban consumers (CPI-U) and measured in year 2000 dollars. When referring to a year of data, I use the spring of the school year — referring to the 1990-91 data as 1991. The first year of the Abbott parity aid is the 1997-98 school year, or 1998. The aggregate finance variables have observations back to 1990 but the local tax and student variables begin in 1995 or later. I also add data from the New Jersey School Report Cards to look at student characteristics and school-level spending indicators.[16] To look at the actual staffing decisions of districts I use New Jersey's Certificated Staff Reports. These files report the salary, age, experience, education and job code for every staff member in a position which requires certification from the state. This includes all classroom teachers as well as superintendents, principals and educational support personnel.[17] This results in a richer data set than has been used before to study the effects of school finance reform, allowing me to examine the aggregate effects on school expenditures and specific school level effects of these funds.

My analysis will rely heavily on categories of spending and funding reported in the Census of Governments F-33 Survey. Table 2.1 lists the main categories and subitems covered in this survey. The parity aid created by the court ruling in 1997 does not appear separately in this form but shows up in the data as part of "Compensatory and basic skills attainment programs" (which I will call Compensatory Aid or Comp Aid), subcategory number four under State Revenue. I subtract parity aid (obtained from

[16] More information on these data is included in the data appendix.

[17] Here I use the cross sectional information from these files, collapsed to the district level. In related work, I match the information on the same teachers over time to create a longitudinal database to look at hiring and retention patterns.

New Jersey Department of Education (NJDOE) publications) from Compensatory Aid and use "Parity Aid" and "Other Comp Aid" separately in the analysis that follows.

The expenditures side of the survey includes total salaries (and benefits) for each subitem under Instructional Programs. Changes in these items could result from increased hiring or from higher salaries to the same staff. I use data on the salaries of all certified staff in school districts since 1996 to further examine the patterns in hiring.

2.2.4 Selection of a Comparison Group

The court ruling, by choosing the most disadvantaged urban districts, created a situation where it would be difficult to credibly evaluate the effects of this policy against a control group of unaffected districts. However, the case does provide significant variation in parity aid among Abbotts. In 1998, the first year of the policy, three Abbott districts received no parity aid and the largest amount was $1614 per pupil in the City of Orange. There is also variation among the Abbott districts on many other dimensions. For example, the average community income per pupil in 1996 for Abbott districts is $68,663, but the highest is $280,000 in Hoboken and the lowest is only $23,400 in Camden. There is at least one district, Atlantic City, that is similar to the Abbott districts on all dimensions except the local tax base. Atlantic City receives adequate local funding from casino revenues, but the characteristics of the student population line up well with the Abbott districts.[18] Ideally, my analysis would have a control group of districts that look just like the Abbotts but were not subject to the Abbott reforms. To approximate this using the available data and

[18]For example, the average income per pupil, fraction of students receiving free lunch, fraction of students who are black, and fraction of students who are hispanic in Atlantic City are all within one standard deviation of the Abbott average.

natural experiment, I create a comparison group by selecting the districts which are most like the Abbott districts.

Since the Abbott designation was assigned in 1985 and based on the state SES classification using 1980 census data, change over time in New Jersey may have resulted in more similarities between groups in 1997 than existed in 1980. To take advantage of these similarities and reduce the likelihood that outliers among the non-Abbott districts are driving the results, I use data from 1996 to identify an appropriate comparison group. I flag districts that have values below the 10th percentile of the Abbott districts for percent of students eligible for free lunch, district enrollment, or district unemployment rate, and above the 90th percentile for percent white, district income per pupil, or equalized property value per pupil, and exclude districts that have more than two flags. This excludes one Abbott district, 28 of the other 51 districts in the three poorest state SES categories, and all but one of the districts in the five wealthier categories.

Because my empirical strategy uses district fixed effects, the key coefficients are identified using within-district variation in parity aid over time. The purpose of the comparison group is to identify the state average levels of funding and expenditures in each year. Table 2.2 presents means of key variables for Abbotts, the selected comparison group and all other K-12 districts.[19] The Abbott districts do look significantly different from other K-12 districts. They are on average larger, poorer and serve larger fractions of minority, special education and limited English proficient students. Although the differences are smaller, the Abbotts are significantly worse off than the selected comparison group on most measures. Because the comparison group is used to identify state-level effects, my empirical strategy requires that

[19] New Jersey has roughly 600 school districts, but this number includes many specialized districts such as K-6, 7-12 and special vocational districts. All Abbott districts are K-12 districts. To facilitate comparisons across districts, I include only K-12 school districts, a total of 214 districts.

statewide changes affect the Abbotts and the comparison group similarly. I present results in Section 4 using alternate selection methods to verify that the sample selection is not driving my results.[20]

Figure 2.3 shows all K-12 school districts in New Jersey and shows the location of the Abbotts, the comparison group and the districts that are being excluded based on this selection criteria. The figure demonstrates that each group of districts is spread throughout the state. It is not the case that the Abbotts are all located in one area and the comparison districts are located in a different area of the state. The larger districts in the lower portion of the map tend to be more rural and the many geographically small districts that are clustered together tend to be more urban.[21] The bulk of the population of New Jersey is located across the middle of the state, between Philadelphia and New York City.

The Abbott average minus the comparison group average of Parity Aid and Comp Aid are graphed in Figure 2.4. The between-group difference in Compensatory Aid is between \$200 and \$500 before 1997 and jumps to \$1000 when Parity Aid is introduced. After 1997, Other Comp Aid is the area between the two lines. The increase in Other Comp Aid beginning in 2000 is driven by additional programs mandated by later Abbott judgments. The Abbott rulings in 1997 and 1998 laid out clear mandates regarding funding levels and necessary programming. The court required the state to fund the Abbott districts at a high level, with the intention of improving student outcomes. Two necessary, but not sufficient, conditions for the money to have an effect on students are that the funding was used for school expenditures and

[20]The alternate strategies include specifications that used all districts, all "Other Poor" districts, samples that dropped all central cities (mostly Abbotts), that used all districts weighted by inverse probability weights, and samples trimmed of observations with low probabilities of being an Abbott district. Both probability measures are predicted using a logit model of the likelihood of being an Abbott district based on 1996 district level characteristics. The results are qualitatively the same over all specifications so I present the results from the most transparent selection scheme in the text.

[21]The analysis sample covers 38% of total K-12 enrollment in 2000.

that those expenditures were related to student achievement. The next section looks to the existing literature for predictions about the effects of increased state aid in the Abbott case.

2.3 Previous Literature and Predictions

This paper, while focusing on the expenditure effects of the Abbott reforms, contributes to several strands of the existing literature. The public finance literature on intergovernmental transfers and the flypaper effect offers predictions for the amount of state aid that passes through to expenditures. Several papers about the effects of different expenditure types and on New Jersey speak to the possible uses of funds and the ultimate effects of this spending. I discuss each of these in this section.

The public finance literature about the flypaper effect dates back to Bradford and Oates (1971) who model intergovernmental grants as the result of majority voting decisions. They conclude that local governments, acting on behalf of local citizens, should treat lump-sum grants from the state as increases in local income. This implies that local governments would spend the same fraction of grant income as they do a dollar of additional local income. A common estimate of this is 5%. The flypaper effect refers to the phenomenon that “money sticks where it hits,” in the sense that intergovernmental grants targeted to a specific purpose are spent at a higher rate than the propensity to consume out of local wealth or other income.[22] The puzzle of the flypaper effect is that the empirical evidence shows consistently that localities spend much more than 5% of lump-sum state grants. Hines and Thaler (1995) report estimates of 40% to 58% for flypaper effects in studies of the effects of state aid to local school districts.

Moving to education expenditures and funding, Fisher and Papke (2000) review

[22]See Hines and Thaler (1995) or Bailey and Connolly (1998) for a review of the flypaper literature.

the theory and empirical evidence on how local governments react to different kinds of education grants. They find that there is increased spending associated with all types of grants, but also leakage across the board; local tax relief is a common use. Feldstein (1978) and Gordon (2004) look at the effect of Title 1, the federal aid targeted to poor students. Title 1 funding may be the funding stream that is most like parity aid, in that it is directed specifically to poor students. However, school districts may react differently to federal aid (Title 1) than to state aid (parity aid). Feldstein finds that about 70% of Title 1 funding passes through to increased educational expenditures. Gordon reports that expenditures increase dollar-for-dollar in the first year following a change in Title 1 funds, but after three years there is full crowd out of local revenue. Focusing on different types of local government structure, she finds that the crowd out is occurring mostly in the county or city government contribution to the school district. For districts with no city or county layer (as in New Jersey where school district boundaries closely match municipal boundaries), she cannot rule out zero crowd out.

Bedard and Brown Jr. (2000) and Betts (1995) provide evidence that different types of expenditures have different effects on student outcomes. Bedard and Brown Jr. (2000) use data from California to test the effects of different expenditure patterns across districts on student test scores, arguing that the school finance system and expenditure limitations in place there make district level total expenditures orthogonal to district preferences. They conclude that a reallocation of $100 per pupil from administration to instructional spending or a $100 per pupil increase in total expenditures targeted to instruction moves a district up 5 percentiles in the state test score distribution. Betts (1995) uses data from the Longitudinal Study of American Youth and finds that traditional input indicators such as class size and teacher experience have little effect on student test scores, but that spending on computer technology does predict student achievement.

Several papers look at local responses to past education finance changes in New Jersey. Goodspeed (1998) examines the reaction of local property taxes to increases and decreases in the New Jersey state income tax, using changes in the early 1990s. In New Jersey the state income tax does not go into the general fund; it is required to be used for "property tax relief" which means that it is distributed to localities (as state aid to school districts and municipalities) and can be used to increase spending by these local governments or to offset property taxes. He finds evidence of a flypaper effect in both directions: the local property tax cut does not fully offset an increase in the state income tax, and property taxes are not increased to completely make up for a cut in the income tax. The magnitude of the flypaper effect he estimates is similar to the studies noted above: roughly 70% of funding sticks.

Firestone, Goertz, Nagle, and Smelkinson (1994) and Firestone, Goertz, and Natriello (1997) use eleven case study districts to describe how New Jersey school districts used their funding from 1992-97, when funding changed as a result of the Quality Education Act (QEA) passed in 1990. This legislation was ultimately overturned as inadequate to remedy the Abbott disparity, but it did increase funding to poorer districts and lessened the funding gap by roughly 10%. In response to these funding increases, the authors found the largest increases in support for at-risk students and for teachers. They also found sizable increases in health and social services and supplemental after school, weekend and summer programs. The Abbotts also improved their academic programs, adding advanced classes and hiring teachers for specialized electives. The authors also discuss differences in decision making processes between Abbott and non-Abbott districts. In Abbotts, the main communication is between the central district office and the state but in non-Abbotts it was between the district and the community. This is partly because the Abbotts were not meeting assessment targets and therefore were under state monitoring, but may also reflect differences in the organizational structure of the school districts. They also find that, in all districts, allocation decisions seem to be at the district level, not at the school or classroom level. This suggests that the district is the appropriate level of analysis for looking at funding and spending decisions. Goertz and Natriello (1999) find that districts tend not to alter their spending patterns much, even following large reforms. If this holds true in this case, then we can expect that the added spending after Abbott IV went to similar programs. The support programs for at-risk students could be expected to improve achievement or, if not improving measured achievement, perhaps reduce dropout rates. The teacher support programs might improve work conditions for teachers and improve teacher retention.

In sum, the previous literature makes several predictions for the fraction of the state aid increase that passes through to school expenditures in the Abbott case, the uses of these funds, and the eventual effects on student achievement. The flypaper literature predicts that 40 to 70% of the state aid will go to increased school spending. Many of the explanations for the flypaper effect rely on citizen misperception of or inattention to the actions of local goverments, in this case the school districts. The public nature of this case and continuing court oversight make it unlikely that citizens were unaware of the added funds. The effects of this attention are unclear; the court oversight would likely lead to a higher flypaper effect as it creates pressure to spend the money on schools, but citizen attention could tend to create a smaller flypaper effect as local voters may prefer tax relief.

The results from Bedard and Brown Jr. (2000) and Betts (1995) imply that the exact uses of the added state funding in the Abbott case will be useful in predicting student outcome effects. If the bulk of the money was spent on instruction, as the results from earlier funding increases studied in Firestone et al. (1994) and Firestone et al. (1997) suggest, we will predict that Abbott students will perform better than if the bulk of the added money was spent on non-instructional uses. The Firestone et al. papers also imply that it is important to look at the specific focus of even instructional expenditures — tutors for remedial education will affect different parts of the student distribution and different outcome measures than increased Advanced Placement offerings.

2.4 Empirical Analysis

This section presents my empirical analysis, beginning with a discussion of the difficulties in estimating causal effects in education, my estimation strategy and the instrumental variables used to estimate the effect of state aid on school district expenditures. After describing the instruments I use, I present the results for aggregate expenditures, measuring the flypaper effect in the Abbott case. I then examine how other sources of funding changed with the added state aid and how specifically districts spent the funds. Finally, I present suggestive evidence from personnel data to describe hiring patterns for Abbott districts relative to the comparison group of districts.

2.4.1 Difficulties in Estimation

In education spending, it is widely acknowledged that school districts with low local wealth (which therefore tend to receive larger equalizing grants from the state government) have relatively high education costs stemming from large percentages of students with special needs (Minorini and Sugarman, 1999). These additional costs complicate comparisons of aggregate expenditures across school districts because, conditional on a given spending level, low-wealth districts spend a larger fraction of their budget on compensatory programs and fixed costs, leaving less money for "regular education." Put differently, if the cost of achieving a given level of education varies across districts, then district-level spending figures conflate preferences for high levels of education with high costs for providing any given level of education.

A school district's total expenditures are a function of two things: local preferences for a level or quality of education and the costs of educating the students in the district. School districts vary on both of these dimensions. Empirically, both educational costs and preferences for education are related to local wealth and/or income.[23] Desired education is rising with local wealth and the costs of a given level of education tend to be inversely related to local wealth. In principle, we can measure and control for wealth and costs directly, but preferences are unobservable. In practice, even measuring educational costs is not straightforward (Yinger, 2004).

The funds used to support the chosen level of education expenditures come from both external and local sources. External sources are primarily Federal and State Aid but can also include grants from non-governmental organizations. Federal Aid is almost exclusively compensatory and meant to compensate districts that have high costs (such as Title 1 and Bilingual Education Aid) or low revenues (such as Federal Impact Aid, which offsets lost property tax revenue from government-owned property). These funds are determined by the characteristics of the student population or district and are not a choice variable for districts. Local Revenue consists mostly of property taxes but also includes payments from other school systems, revenues from lunch programs, textbook sales and interest on assets.

[23] I will use wealth from here forward to refer to both income and wealth since the two are closely related and local property wealth is used in most school funding formulae.

Most of these funding sources are also correlated with wealth. The parts of state and local revenue that support basic or "regular education" funding are directly determined by local wealth through the statutory funding formula. The factors which trigger compensatory programs are correlated with wealth, with high compensatory aids typically associated with low wealth. The optional part of local funding, the amount above the foundation level, tends to be higher in high wealth communities and is based on the local preferences for education services. In New Jersey, the regular education budget (which is the reference category for Abbott parity) is the sum of the foundation budget and the optional local levy. This is the funding that is intended to produce the locally chosen level of education. If compensatory programs were accurately accounting for the different educational costs across districts, we could subtract those programs out and the remaining funding would tell us about the education preferences of localities. Since, as Duncombe and Yinger (2005) point out, the compensatory programs are not accurately scaled to the true costs of educating disadvantaged students, we cannot separate out preferences and costs.

2.4.2 Empirical Strategy

Regardless of local preferences or needs, the question at hand is how a district responds to a change in external aid. If the district has already chosen the optimal level of educational spending, a new grant should be treated as income to the locality. The district may choose to increase spending on education slightly, but would also be expected to increase spending on other public goods, including local tax relief. If all preferences and costs could be measured, I would want to estimate the following equation to measure the average propensity of municipalities to spend state dollars on local education:

$$Expenditures_{jt} = \alpha_0 + \beta StateAid_{jt} + \gamma LocalIncome_{jt} + \delta_1 X_{jt} + \epsilon_{jt}, \qquad (2.1)$$

where j indexes districts and t indexes time. X_{jt} encompasses all characteristics of districts that affect expenditures, including local preferences for education and other public goods, and any factors that lead to differential costs of education. A significant difference between β and γ, the coefficients on state aid and local income, would be evidence of a flypaper effect.

The feasible range of β is zero to one, with zero representing complete crowd-out of other funds and one representing complete pass-through to educational spending.[24] State Aid has several components. The largest are formula aid, determined by the statutory foundation level of spending and the local wealth level, and compensatory programs meant to help districts pay for needier students — both economically needy and in terms of specific educational needs. Since these funds are meant to offset high costs, this aid will go to higher-spending districts (controlling for other characteristics), thus biasing the estimate of β upwards if the determinants of high costs are not fully measured and controlled for. Optional expenditures are rising in local wealth, so districts which choose to spend more for unobserved reasons will tend to be ones with low state aid. This will tend to bias β toward zero. The net effect of these biases is ambiguous.

Two main strategies are typically used to deal with these biases. One is to use detailed cross-sectional data to attempt to control for all differences across districts. Often studies are done at the state level, which allows researchers to abstract away from residential sorting among districts within a state. This strategy relies on dif-

[24] β could be larger than one if increased funds crowd-in additional external funding. One could imagine a locality using a grant as seed money to spur fund raising for a one-time project like a new building or sports facility. See Andreoni (1998) for evidence that seed money increases charitable contributions.

ferences in funding across districts or states for identification, and an unbiased estimate requires that there are no additional omitted factors. The second strategy is to use longitudinal data and an exogenous shock to funding to identify within-district changes over time. Longitudinal data allows the researcher to control for stable differences between districts over time. Feldstein (1978) is an example of the first strategy and Gordon (2004) of the second.

I will use the second strategy. The Abbott IV ruling in 1997 induced a sudden change in state aid. Although the case had been ongoing for many years, the parity aid mandate was a surprise. The parity aid induced by the Abbott IV court decision provides exogenous variation in state aid that I will use to examine the relationship between state aid and school district spending choices. This instrumental variables strategy isolates the variation in state aid that comes from the court ruling and does not use the variation that is associated with district preferences and needs. The policy offers several possible choices for an instrument.

Figure 2.4 shows the variation that will be used under each of these instruments. Instruments 1 and 2 use different combinations of the two variables graphed here. Instrument 1 uses parity aid and other compensatory aid per pupil separately and relies on the increase in parity aid between 1997-98 and 1998-99 and the increase in other compensatory aid between 1999-2000 and 2001-02. The increased compensatory aid starting in 2000 was arguably less fungible than the parity aid because the later Abbott ruling mandated specific programs to be added. Instrument 2 uses only the variation in parity aid, giving a better estimate of the choices of districts if their ability to direct the use of other compensatory aid is indeed constrained. Since the increased compensatory aid is going to the same districts that receive parity aid, Instrument 2 might be picking up some of the variation in compensatory aid and therefore gives an overestimate of the effect of parity aid. For this reason, I prefer Instrument 1.

Instruments 3 and 4 are indicator variables for receiving the treatment of parity aid. IV 3 uses the interaction of Abbott and a Post-1997 indicator. IV 4 adds an indicator for Abbott*Post-2000 to capture the second increase in aid. These instruments do not use the between-district variation in parity and compensatory aid, but they have the advantage of isolating the variation in compensatory aid that comes from the 1997 ruling. The estimates using these two instruments are equivalent to a difference-in-differences estimate comparing Abbott and non-Abbott districts before and after 1997.

Table 2.3 presents the results of the first stage regressions for each of these instruments. The first stage regression is the following:

$$StateAid_{jt} = \alpha_0 + \rho Instrument(s)_{jt} + \alpha_j + \alpha_t + \epsilon_{jt}, \tag{2.2}$$

where j indexes districts and t indexes years. α_j and α_t are district and year fixed effects. The values presented in Table 2.3 are the estimated ρ coefficients on the instruments. All four instruments strongly predict State Aid to school districts. Column 1 implies that a dollar of parity aid leads to an increase of \$1.17 in state aid and that one dollar in other compensatory aid leads to an increase of \$1.16 in state aid. The F-statistic for the joint significance of these two instruments is just over 300. Column 2 reports the coefficient on parity aid when it is used alone, the instrument in specification 2. When parity aid is used alone, one dollar of parity aid leads to \$1.53 in state aid. This coefficient is larger than the one in column 1 because it picks up the other compensatory aid that is increasing starting in 2000. Columns 3 and 4 present the first stage for the two instruments based on exposure to the Abbott reforms. These two instruments give equal weight to all Abbott districts including those that receive little or no parity aid.

Instruments 1 and 2 are valid if the amount of the parity aid is not related to unmeasured characteristics of the districts that are related to expenditures and change over time. If you are concerned that the comparison group is not adequately controlling for year-to-year state changes[25], you will prefer the estimates from Instruments 1 and 2 because these use the variation in state aid and parity among Abbotts over time to identify both the year fixed effects and the coefficient on parity aid. These first two instruments implicitly weight the districts that get the most aid more heavily. Since the variation in parity aid among Abbott districts comes in part from pre-ruling local revenue differences, IVs 3 and 4 may be less susceptible to contamination by other unobserved variables.

2.4.3 Effects on Total Expenditures, or the Flypaper Effect

Table 2.4 presents the results for the second stage regression of total school district expenditures on state aid using the four instruments described above. The OLS estimate is included in row 1 for comparison. The regression equation is the following:

$$Expenditures_{jt} = \alpha_0 + \beta \widehat{StateAid}_{jt} + \alpha_j + \alpha_t + \epsilon_{jt}, \tag{2.3}$$

where the Abbott IV decision and the related funding are used as instruments for State Aid. The district fixed effects, α_j, control for differences between districts that are fixed over time. These fixed effects attempt to control for all permanent differences between districts, including potentially observable differences in scale that may lead to different expenditure needs and patterns, differences in the educational

[25] Because the district fixed effects control for time invariant differences between districts, the concern must be that there are changes in state aid after 1997, correlated with but not caused by the court rulings, that affect the Abbott districts differently than they affect the comparison districts.

needs of students, and also unobservable differences in preferences for educational spending and the efficiency of districts in using their resources. If these characteristics are changing over the sample period, the fixed effects will not capture these aspects and the estimate of β could be biased. The estimate will be biased only if the changes over time are correlated with the instrument.[26] The year indicators control for state level changes from year to year that affect all districts similarly.

The coefficient of interest, β, tells how much of the added money from the policy is spent on education in a given district. β is identified using within-district variation over time for Abbott districts. The non-Abbott districts, through the year fixed effects, help identify changes over time that affect all school districts equally. The standard errors are clustered at the school district level to allow for within district correlation over time.

Column 1 presents the estimates of the effect of state aid on total expenditures. The coefficient of 0.66 in row 2 implies that 66 cents of every dollar of state aid are spent on schools. The coefficient on row 3, which uses only the parity aid as an instrument, is smaller than the estimate that uses both parity and compensatory aid. One explanation for this is that the parity aid is indeed more fungible than the other compensatory aid added later. The standard error in row 3 is also larger, so I cannot rule out with any of these instruments that 80% of the funding goes to education expenditures. Rows 4 and 5, using the policy dummy, give similar results. The IV estimates in this table are not significantly different from the OLS results. This may be because the differences between districts in preferences and costs that may bias the OLS estimate are averaged out at the level of total expenditures. Some subcategories, shown in subsequent tables, do show sizable differences. The estimates

[26] Inclusion of measurable district characteristics including enrollment, fraction of students classified limited English proficient, and fraction of students classified as special education does not affect the estimate of β.

do not vary much over the different instruments either. This pattern holds true in the remaining tables, so to focus on meaningful differences, I report only the results from instruments 1 and 3 in all remaining tables. This allows me to see any difference between the instrument that uses within-district variation over time (IV 1) and the one that uses only policy-level variation (IV 3).

To verify that these estimates of the flypaper effect are not being driven by the sample selection, Table 2.5 presents the corresponding results using several different comparison groups. The first row repeats the coefficient from Table 2.4. Rows two and three report the coefficient from models that include only districts in the three poorest SES categories (using New Jersey's classification system) and that include all K-12 districts. Rows 4-6 present estimates from specifications that exclude districts that have a low probability of being an Abbott based on logit regressions of the propensity of being an Abbott based on 1996 data. The results in rows 7 and 8 use similar logit models to create propensity score weights that are used to weight the data from all K-12 districts. The results using these different sample selection strategies range from 0.55 to 0.76 but all include 0.66 easily within the confidence interval.[27] These results suggest that the sample selection is not driving the estimates of β.

Taken together, the estimates in Table 2.4 suggest that around two-thirds of state aid increases go to increase educational spending. Column 2 provides the 95% confidence interval for each estimate. The estimates are all also significantly smaller than 1, meaning that there is some crowd out of other funding. Conservatively, using the upper bounds of the confidence intervals, at least 10-20% of the funding is not used for school expenditures.

[27]The data appendix provides more detail on these alternate selection strategies.

2.4.4 Crowd-Out of Other Funding

To determine what funding is crowded out, Table 2.6 presents estimates of the effects on other funding streams. Column 1 shows results for federal revenues. The small and insignificant coefficient of 1.4 cents per dollar of state aid in row 2 supports the conclusion that there is no effect on federal revenues. This is a reassuring result because federal revenues are formula based and triggered by student characteristics. No change here implies that the characteristics of the student population are fairly stable across the policy change. The larger OLS coefficient in row 1 implies that there is a relationship between high state aid and high federal aid that is not present using the exogenous variation captured by the IV. Any leakage from the policy seems to be coming through reduced local effort. Column 2 shows the estimates for total local revenues per pupil, the funding stream over which the municipality has direct control. This category is dominated by the local property tax, but also includes tuition payments, fees for services, and revenue from school lunches and textbook sales. Figure 2.5 shows the pattern over time in local property taxes per pupil. The Abbotts were constrained by the court ruling from lowering their aggregate levy from the 1997 level because of the maintenance of effort requirement. Unable to reduce their levy because of the mandate, the Abbotts were able to hold it steady in nominal terms; this translates to a gradual decrease in real terms, while the non-Abbott districts raised their levies slightly over the period.

Column 3 of Table 2.6 presents the regression results for local property tax levies for schools. The overall pattern shows a small negative effect, but this is not statistically distinguishable from zero in most cases. Column 4 shows the results for the total local effective property tax rate. For this column, state aid is measured in thousands of dollars per pupil. This combines all municipal property taxes, adding the local funding for all other public goods to the school taxes. Across the instruments, the effect here is consistently about negative 0.10 percentage points per $1000 of state aid per pupil. The Abbotts were restricted from lowering the property tax rate but are lowering their overall tax rates for all municipal operations. Figure 2.6 shows this quite dramatically.[28] This suggests some fiscal substitution across municipal expenditures. One explanation is that schools are providing services that were previously provided by the city, which allows the locality to reduce spending on these services. For instance, one of the requirements of the 1998 Abbott ruling was that each school employ a social worker and this may have replaced similar community services.[29]

2.4.5 Expenditure Results

Broad Expenditure Subcategories

Table 2.7 returns to the expenditure side and presents estimates for the effects of increased state aid on the major expenditure subcategories. For reference, columns 2 and 3 are subcategories of column 1, and columns 1, 4, 5, and 6 add up to total expenditures. There are large and statistically significant increases in current K-12 expenditures and its subcategories, instructional and support services expenditures. The significant increases in expenditures as a result of an increase in aid are strongly concentrated in the expenditures most directly related to student education — those that affect the classroom and supplemental services. Row 2 of column 2 shows that an increase of $1 in state aid per pupil leads to 42 cents of increased spending on instructional expenditures. Instructional expenditures include salaries and benefits

[28] Another explanation for this is differential increases in property values in Abbott districts. Statewide property values fell through the 1990s and recovered starting in 2000. The recovery for the Abbott districts began a few years earlier than for the comparison group, but values in the comparison group grew more sharply starting in 2002 leading to a rewidening of the gap between the two groups.

[29] Courant and Loeb (1997) discuss fiscal substitution as it relates to the Michigan school finance reform in the mid-1990s.

for instructional staff, so an increase in this category means that districts are spending more money on teachers. This could mean that districts are hiring more teachers or that districts are paying teachers more, so this finding alone does not inform us about the number or quality of classroom teachers. The other significant increase in expenditures is for support services, with an increase of about 35 cents per dollar of state aid. This category covers a variety of functions which are broken out in the next table. The results in columns 4 and 5 support a conclusion that none of the added state aid goes to other current expenditures or non- K-12 expenditures. There is a moderate, but statistically insignificant, decrease in capital outlays.[30]

Expenditures on Salaries

Support services, the category in column 3 of Table 2.7, encompasses several different functions. Table 2.8 confirms that, even within support services, the increased expenditures are concentrated on categories that can be considered educational support. The pattern is the same as in the broader expenditure categories. The spending increases that result from increased state aid are concentrated in the categories most related to students' learning environment. Instructional salaries show the largest magnitude increase, 17 to 20 cents per dollar of increased aid, followed by pupil support and instructional support. There are small increases in administrative support and operations and maintenance support and no change in transportation support. For reference, column 1 in this table is a subset of column 2 in Table 2.7 (Instructional Expenditures) and columns 2-6 are subcategories of column 3 in Table 2.7 (Support Services Expenditures).

The fact that the OLS and IV estimates for total expenditures are very similar

[30]This may be due to the fact that the state took over all facilities construction for Abbott districts starting in 2000. This was also a result of the Abbott rulings, because the court found that adequate facilities were a necessary part of a thorough and efficient education.

while there are differences in the estimates for some subcategories implies that the biases may wash out in the aggregate, but the specific spending patterns might depend more on unobservables. For instance, the OLS estimates for the current K-12 expenditure subcategories in columns 1 to 3 of Table 2.7 are smaller than the IV estimates. This is likely because urban districts, which tend to receive more state aid, have higher fixed, non-educational expenditures for things like building and land costs, security, and transportation. This means they have less money to put toward current educational expenditures, all else equal. Since these other costs are fixed, the exogenous increase in aid allows them to focus on the educational programs. This same pattern is seen in Table 2.8. The OLS estimates of the effect of state aid in the first three columns, those most related to student learning, are all smaller than the IV estimates. Columns 4 through 6 show no significant difference between the OLS and IV estimates, implying that districts with different unobservable characteristics spend similarly in these categories.

Within-School Spending Indicators

The results for expenditure and salary items from the Census of Governments School Systems data support an interpretation that the money went primarily to uses that could be expected to improve the educational classroom environment for students. These expenditure categories are not entirely transparent and leave the following question: How does an increase in instructional salaries translate to the classroom? I use school-level data from the New Jersey School Report Cards to look at several indicators within schools.

Table 2.9 presents the results for the effect of state aid on computer access and the relative numbers of students and faculty. State aid is measured in thousands of dollars, so the coefficient in row 2 of column 1 implies that $1000 of state aid per pupil reduces the average class size by 0.07 students. The results in column 1 suggest that class size is not affected in a statistically or economically meaningful way in response to state aid. Column 2 shows a statistically significant decrease in student-faculty ratio of a magnitude of about 10 times larger than the effect on class size. The combination of these two results implies that Abbott districts have added extra staff but not classroom teachers. If these extra staff are tutors or remedial instructors they may significantly improve student learning for the students at the lower end of the achievement distribution.[31] Columns 4 and 5 attempt to measure whether districts increased student access to computers. The data on the student computer ratio and the fraction of classrooms with internet access are not reported before 1997, so these estimates are quite noisy and do not provide any convincing evidence that access to computers has changed for Abbott districts.[32]

Staffing Levels

To look more closely at the types of staff hired in Abbott districts, I present graphs of the number of staff per 1000 students in several staff and teacher categories over time. To create these graphs, I use the New Jersey Department of Education's Certificated Staff data collapsed to broad teacher categories at the district level.[33] Figure 2.7 shows the average number of principals per 1000 students for Abbott and comparison districts. Figures 2.8 to 2.13 show several other staff categories.

The numbers of principals (including vice principals), elementary school teachers, high school arts and high school reading specialists are all increasing in Abbott

[31] Data on the number of AP subjects offered and number of students in AP classes would indicate whether educational opportunities were also changing at the top of the student achievement distribution. These data are only reported in the New Jersey School Report Cards for recent years, but in principle they exist for earlier years. I am working on obtaining these data.

[32] I am pursuing additional data on technology resources from the state.

[33] These data provide significant detail on the characteristics and job assignment of every staff member in a position that requires a certificate from the state. This includes all teachers as well as school administrators and many student support positions.

schools. In several of these categories any differential increase for Abbotts over the comparison group occurs after 2000, perhaps as the additional funding is added. The immediate increase in the number of principals makes sense as districts might want more administrators to evaluate and direct how the new resources were used. The number of elementary school teachers increased dramatically for Abbott districts, from about 29 per 1000 students to about 39 per 1000 students, while the comparison group rose from about 24 to 29.

Elementary school teachers is the only staffing category where the scale of the change over this period is economically significant. The addition of one principal for every 2500 students translates to a small increase in per pupil spending, about $35 per pupil. In contrast, the number of elementary school teachers rose from about 30 per 1000 students to about 40, or an increase of 1 teacher per 100 students. This implies a much larger increase in spending on teacher salaries, matching the results in Table 2.8 that show that the largest salary category increase is for instructional salaries. Using the mean characteristics of Abbott districts in 2000, this would imply hiring 94 more elementary teachers at a cost of $4,692,640, or $498 per pupil, per year.[34] Surprisingly, the number of teachers in the core subjects of high school English and math decreased slightly for Abbott districts over this time. The changes do not correspond exactly with the policy changes, and I discuss some explanations for these trends in the next section.

Table 2.10 presents the regression results that quantify the changes shown in these graphs. The table includes results for 22 categories of teachers and other school district staff. The first column of results uses parity aid and compensatory aid, IV 1. The coefficient of 1.4 in the third row says that for every $1000 in state

[34]For Abbott districts in 2000, the average number of pupils was 9419, the average salary of principals was $86,305, and the average salary of elementary teachers was $49,385.

aid per pupil and every 1000 students, a district hires 1.4 elementary school teachers. The second column of results uses IV 3, the policy dummy, which can be interpreted as a difference-in-differences estimate of the effect of the policy. Abbott districts hired about 1.5 more elementary school teachers per 1000 students per year, relative to the comparison group. There are only two years of pre-reform data, so I am hesitant to interpret these coefficients as an accurate accounting of the effect of added state aid on staffing levels, but these numbers provide a way to judge whether the graphs are displaying a substantively important difference between the Abbott and comparison groups. The only staffing categories that show significant increases using both instruments are elementary school teachers, high school special education teachers, and high school nurses. With IV 1, there is also a significant increase in the number of teachers for high school art, music, physical education, and social studies and in the number of high school counselors.[35]

2.4.6 Timing of the Expenditure Effects

To look at the dynamics of the effects of the Abbott policy on expenditures, I estimate the following equation using three different time windows:

$$\Delta Expenditures_j = \alpha_0 + \beta \Delta \widehat{StateAid}_j + \epsilon_j, \tag{2.4}$$

where j indexes districts and Δ represents a 1-, 3-, or 5-year change with 1997 as the starting year. The results of these regressions are summarized in table 2.11. Each entry in the table is the coefficient β in the above regression for the time span noted at the top of the column and the outcome variable listed in that row. For each outcome, the first row is the coefficient for the regression with no controls and the

[35] The estimates using IV 1 have smaller standard errors, and therefore higher significance levels, because they use the variation in parity aid within districts over time.

second row includes 1995 values of total per pupil expenditures and total enrollment as controls. These results show that there is significant variation in the timing of the aggregate effects presented in the earlier tables. For instance, the estimate for the one-year change in total expenditures for a $1 change in state revenue is 1.36, roughly double the corresponding estimate from using the fixed effects specification and all years of data. Reading across the row, the three-year change is 0.99 and the five-year change is 0.76, implying that as time goes on the districts are diverting more money away from education expenditures. The estimates for current K-12 expenditures and its subcategories show that districts did not move to the eventual high levels of spending immediately, but took time to ramp up. This is likely because the money became available after staffing decisions were made for the first year of the policy. In contrast, capital outlays, which show a negligible impact in the main analysis show an almost dollar-for-dollar increase in the one-year change, but drops off to negative but statistically insignificant changes in the 3- and 5-year analyses. This suggests that districts may have used the influx of money in 1997-98 to complete outstanding capital projects or to move up projects that were planned for the following several years.[36]

The second grouping of variables shows the effects of state aid on other finance streams. The first variable is total revenue, which is the sum of federal, state and local revenue. The effects on total revenue are similar over the different time frames, with a dollar in state aid translating to 75-85 cents of revenue available to the district based on the point estimates. As is shown in the main analysis, the leakage is coming through crowd-out of local revenues but not exclusively, or even primarily, through property tax reductions.

[36] It is also relevant that the state took over capital expenditures in later years.

The third grouping of variables, salaries, suggests that districts were scaling up their spending in pupil and instructional support, but for instructional salaries and operations and maintenance support the timing of the effects is nonlinear. Instructional salaries increase only about 11 cents per dollar of state aid in the one-year time frame, but increase by 30 cents per dollar in the three-year window. The five-year change is 0.16, in line with the estimate of 0.17 from the main analysis. Given the large standard errors on the three-year changes, it cannot be ruled out that the effect is around 0.16 at all time periods.

2.5 Discussion and Conclusions

Throughout this analysis, I compare disadvantaged districts that received a large increase in state aid to disadvantaged districts that did not receive this funding increase. While this does not inform us about how schools in general will react to aid increases, it does provide evidence on the segment of schools that are the subject of most concern. These are the districts that are struggling to meet state and national achievement benchmarks and consistently show poor life outcomes for their students.

The results described in Section 4 are promising for urban education policy, in the sense that the bulk of an aid increase to disadvantaged schools goes into spending that can be reasonably expected to improve the learning environment for students. I find evidence that about two-thirds of Abbott parity aid passes through to schools as increased expenditures on education. This is in line with, but at the high end of, previous estimates of the flypaper effect in K-12 schooling. I find that the increased spending is targeted to instructional expenditures and support services and that the one-third of aid that doesn't reach schools is explained by decreased local contributions. There are some puzzles to be explored and important caveats to these results.

The staffing results are surprising in that there are no increases in staffing for the core high school subjects that are the focus of current accountability policies. The New Jersey Core Curriculum Standards, formalized in 1996, include standards for art, sciences, and world languages at all grade levels, so all districts must cover these subjects to meet their constitutional obligation to students. The national No Child Left Behind (NCLB) law added pressure to focus explicitly on the core subjects of reading and math. This could explain the leveling off in the number of art teachers in 2002, but does not explain the decline in the number of high school English and math teachers from 1999-2003. These categories begin to rise at the end of the time period, perhaps in response to the NCLB pressures. Returning to the art and music results, the comparison districts reduce staffing in both of these subjects after 2002 while the Abbotts continue to hire music teachers and level off in arts staffing. This suggests that whatever pressures were introduced by NCLB to focus on core subjects were muted for the Abbotts.

In further work on this topic, I explore the implications of these results for students and schools. The next chapter in this dissertation looks at whether student achievement in the Abbott districts improves following the 1997 funding change. Because I find no change in staffing for high school math and English, the state assessments and national assessments commonly used to measure student achievement are unlikely to show improvement in this case.[37] My results point to other measures that may be more appropriate. One hypothesis is that the increased numbers of counselors and social workers may affect student retention or course choices. The dropout rate

[37] This is not consistent with what I find in the next chapter. I argue there that districts may have focused on high schools in the short term but then pulled back in favor of elementary schools. The HS staffing estimates here are for the full time period - I need to look at shorter windows.

or a cumulative retention rate could be used to investigate whether Abbott districts after the reform are more successful in retaining students. Data on the number of Advanced Placement subjects offered and the number of AP tests taken would shed light on increased opportunities for higher performing Abbott students. A further examination of heterogeneity in spending among Abbotts may inform us about precisely which expenditures do improve student achievement. Given that the Abbott reforms significantly increased spending on instructional and support salaries, a natural next question is whether these increases led to changes in the characteristics of instructional staff. The effects of the added spending within schools on achievement and teacher labor markets will be important to a comprehensive evaluation of the Abbott reform.

Significantly, New Jersey has conducted no comprehensive evaluation of the effects of the Abbott reforms. State officials point to promising improvements in early literacy and to narrowing achievement gaps in elementary grades (MacInnes, 2005). These results come from trends on tests that begin in 1998, after the Abbott reforms were instituted. The court mandated an evaluation but the state has failed to provide it; this is an important policy failure. If states are interested in learning which expenditures have high returns they need to plan evaluations and define appropriate outcome measures before reforms are implemented. The court mandated that money flow to the Abbotts immediately, but districts had several years to implement specific programs. At a minimum, the state could have undertaken a comprehensive baseline accounting of the programs and expenditures in place in 1997, and developed a strategy to evaluate the subsequent changes.

The Abbott case was unique in its selection of wealthy districts as a benchmark for spending adequacy, but a recent case in New York City promises even larger funding increases.[38] Substantial funding increases will improve schools only if the money is spent in ways that improve the education of students. The good news in this paper is that the money provided to disadvantaged districts in the Abbott case did largely go to schools, and it was spent on things that can be reasonably expected to improve student achievement: instruction and support services. The bad news is that the state has not evaluated these changes in a comprehensive or convincing way. Lessons from the Abbott case could inform the pending and future reforms in other states. States undertaking major reform efforts may learn much from the Abbott case, and approach reform with a better understanding of how difficult it is to measure the effects of these reforms and with a plan for meaningful evaluation.

[38]In the Campaign for Fiscal Equity v. State of New York case, the New York Supreme Court has ruled that an adequate education for New York City schoolchildren requires an additional $5 billion of state aid per year. This is roughly $5000 per student per year.

2.6 Figures

Figure 2.1: Parity Aid by Abbott Status

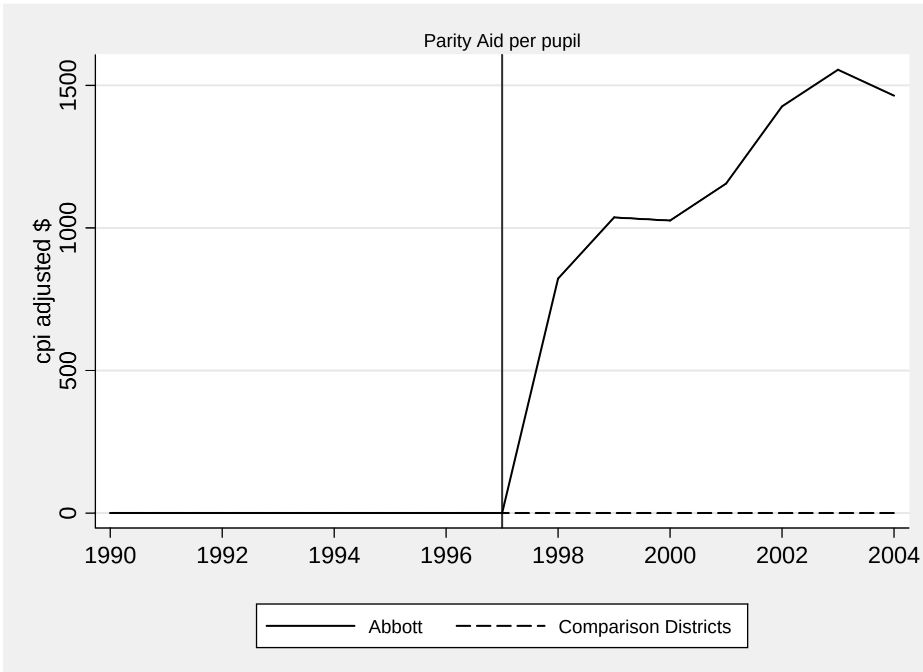


Figure 2.2: Existing School Finance System in New Jersey

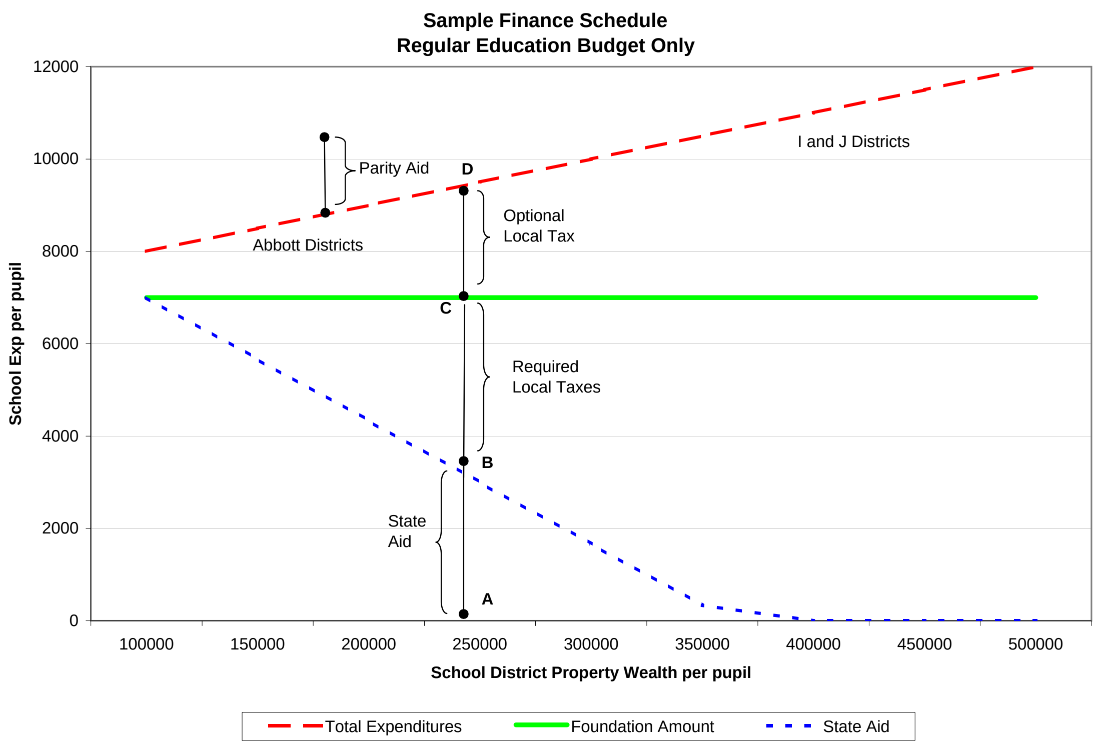


Figure 2.3: Map of Abbott and Comparison Districts

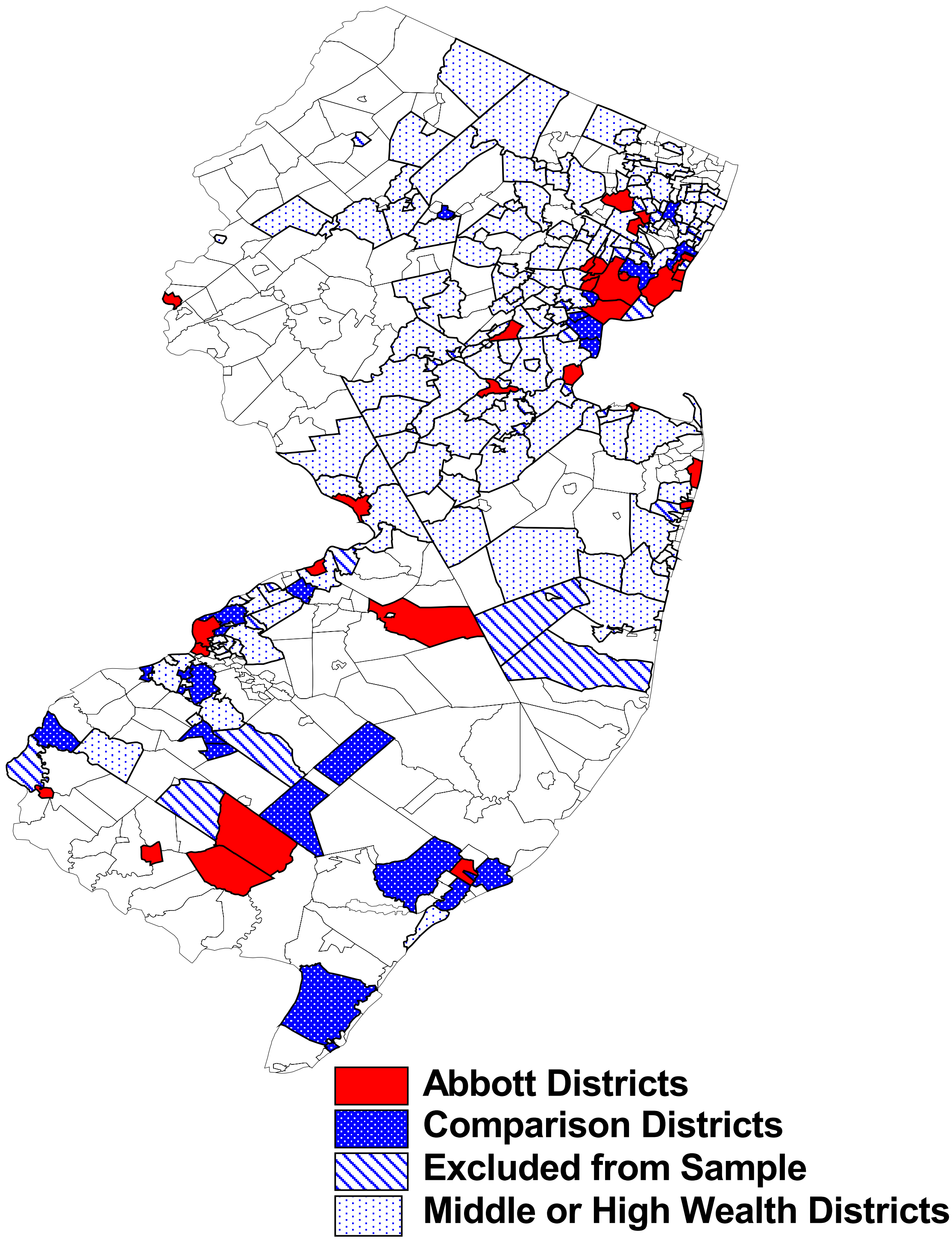


Figure 2.4: All State Compensatory Aid by Abbott Status

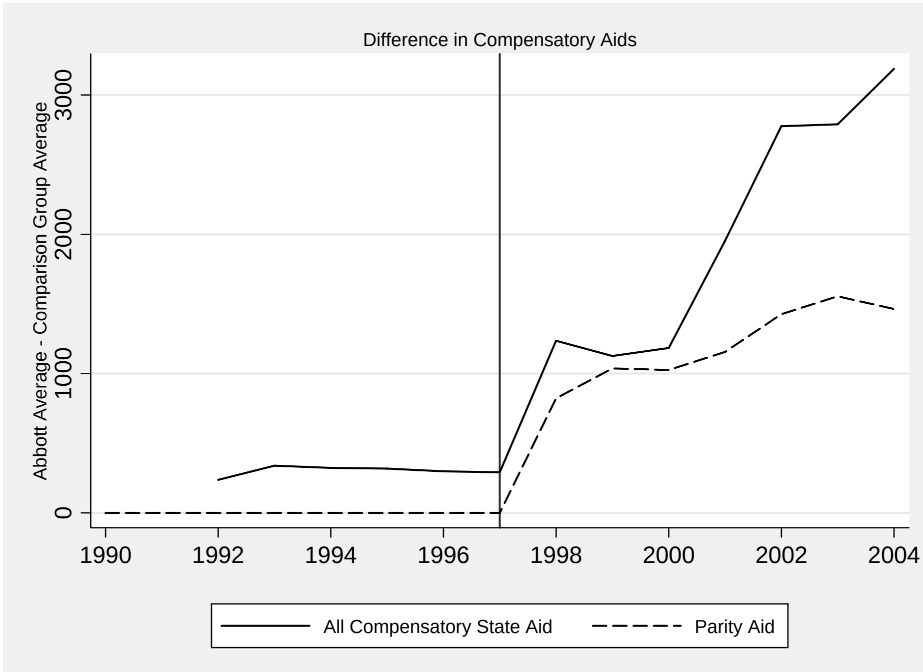


Figure 2.5: Property Tax Levy Per Pupil

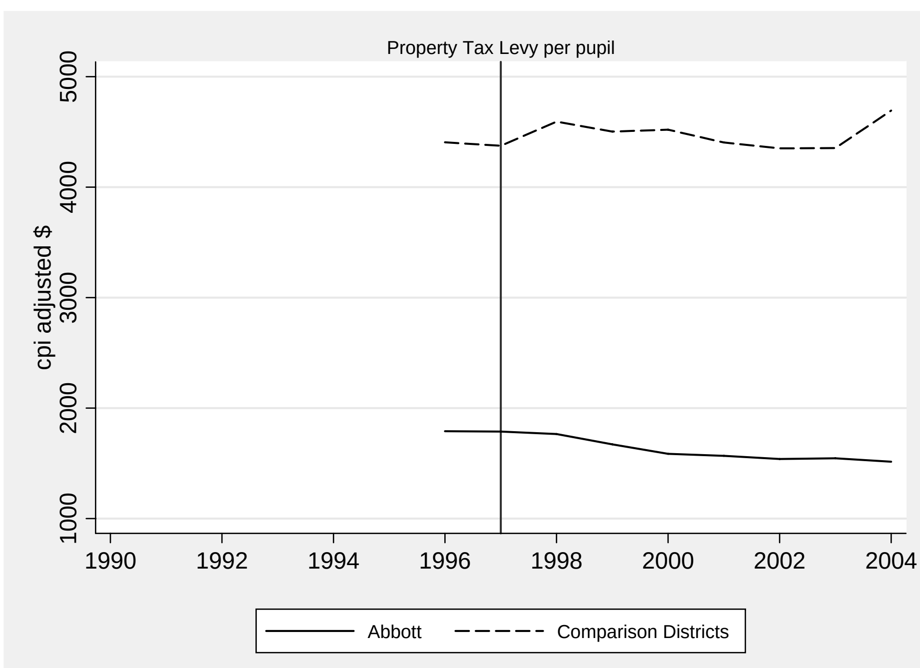


Figure 2.6: Total Effective Property Tax Rate

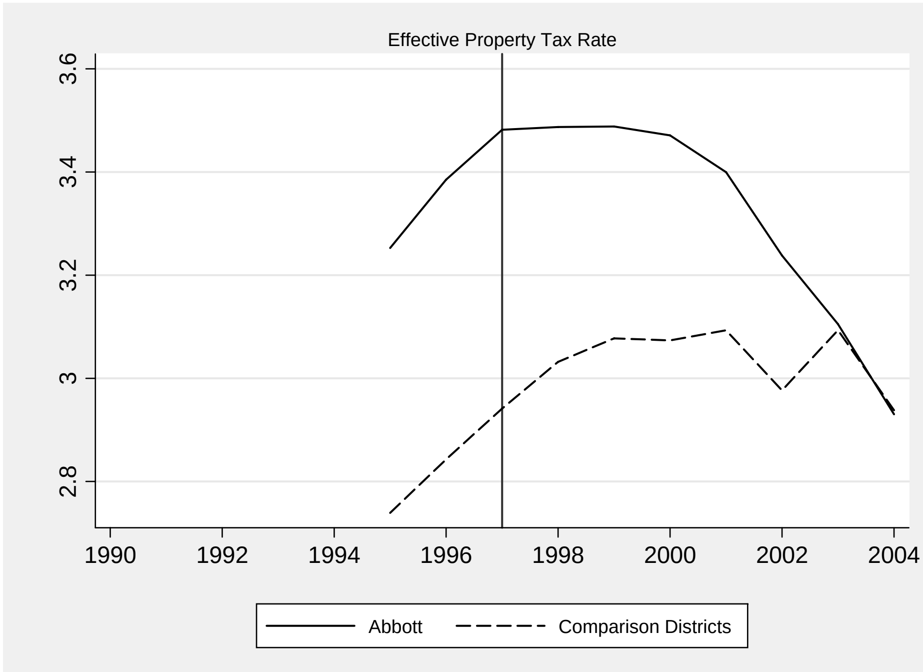


Figure 2.7: Number of Principals per 1000 Students

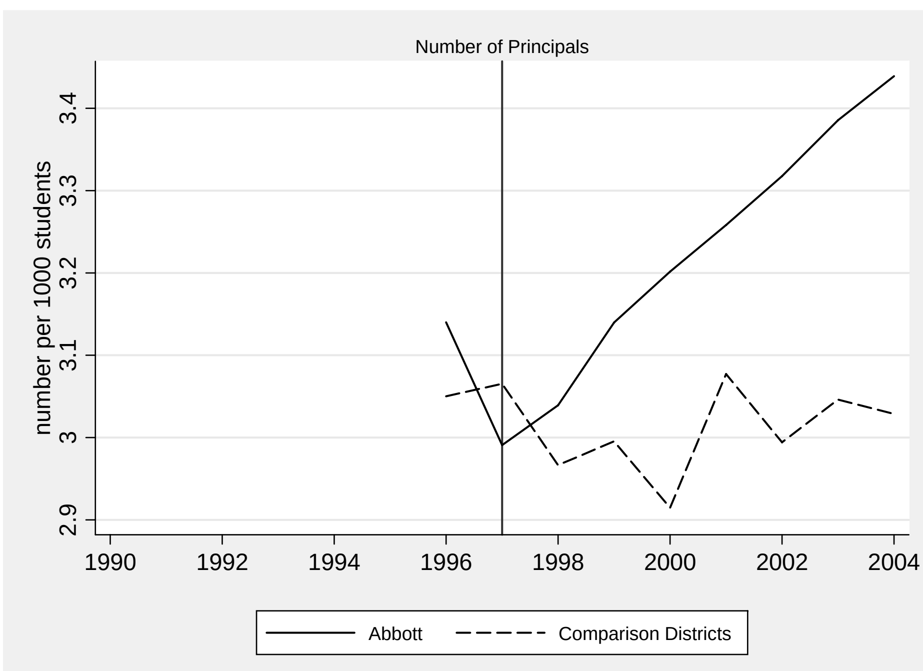


Figure 2.8: Number of Elementary School Teachers per 1000 Students

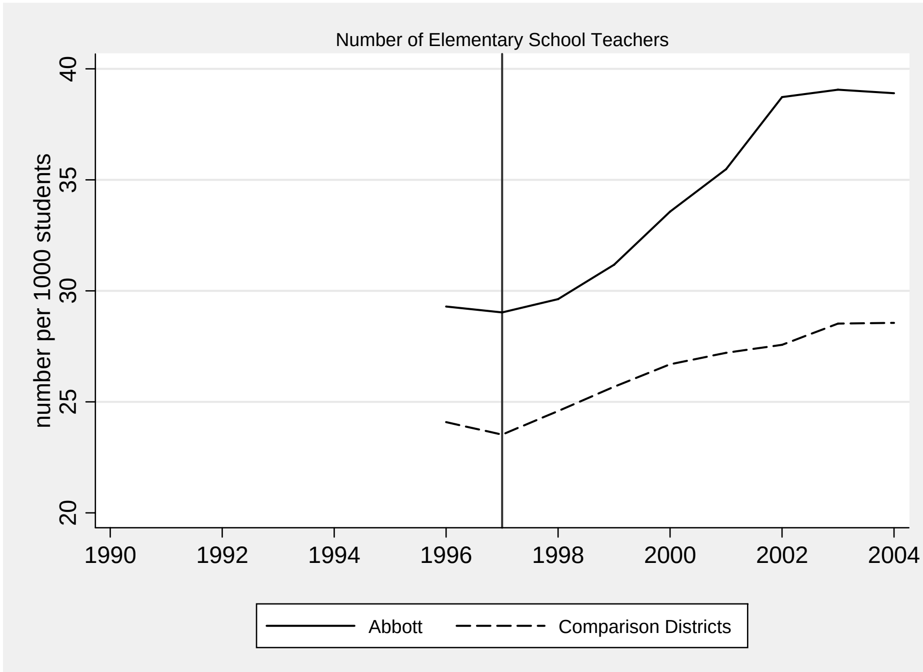


Figure 2.9: Number of High School English Teachers per 1000 Students

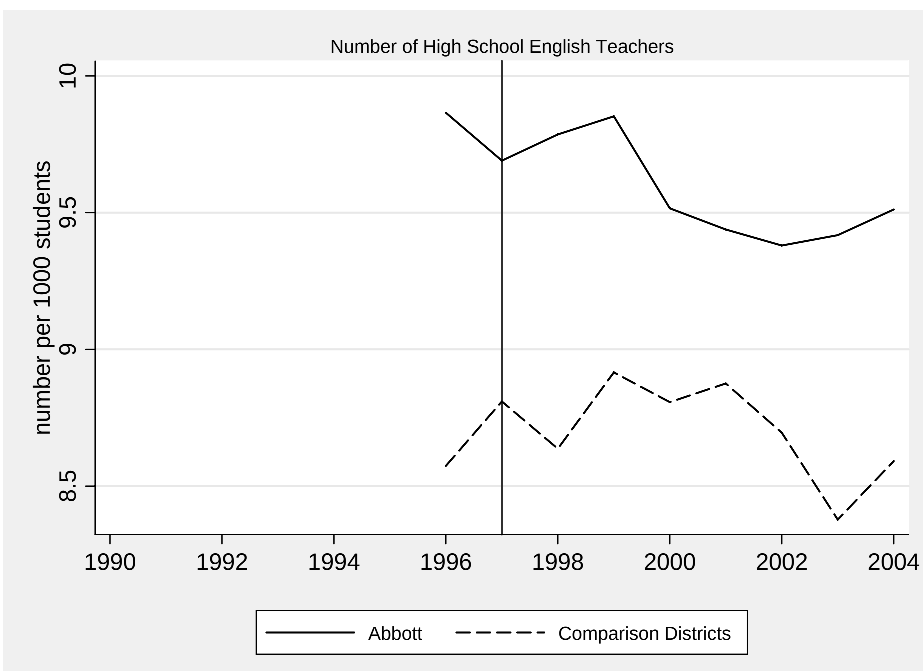


Figure 2.10: Number of High School Math Teachers per 1000 Students

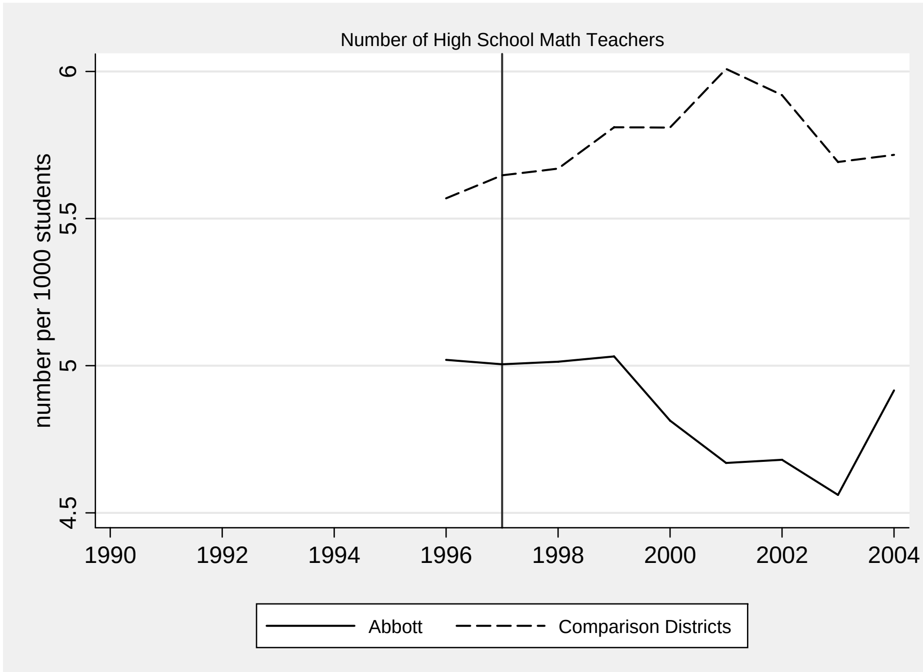


Figure 2.11: Number of High School Arts Teachers per 1000 Students

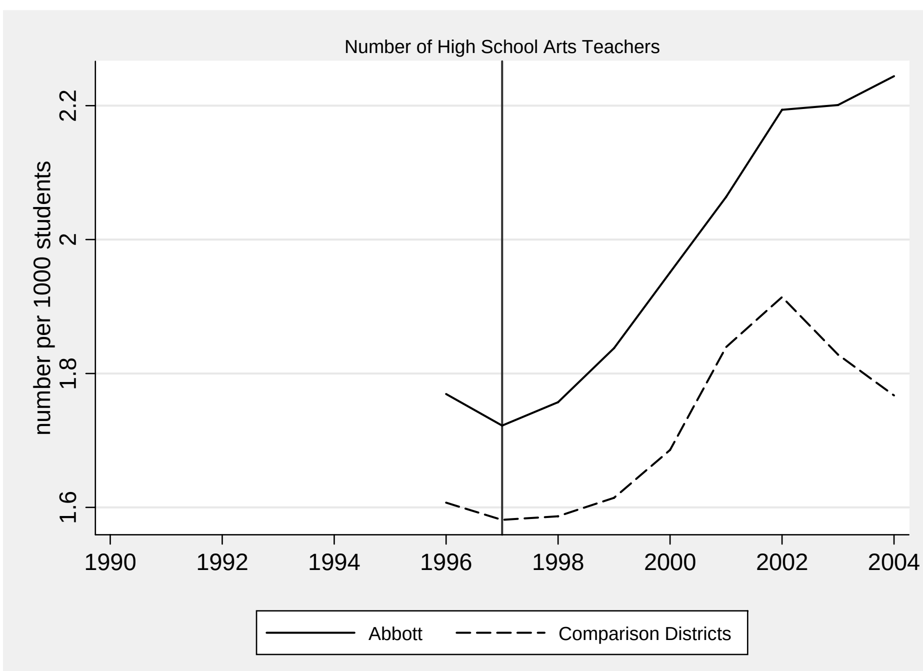


Figure 2.12: Number of High School Music Teachers per 1000 Students

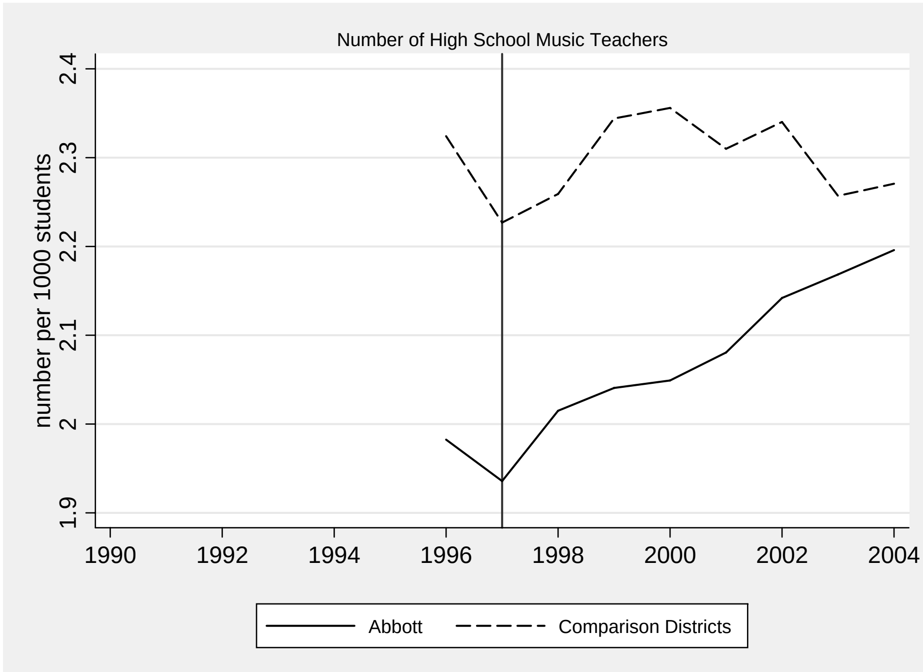


Figure 2.13: Number of High School Reading and Speech Specialists per 1000 Students

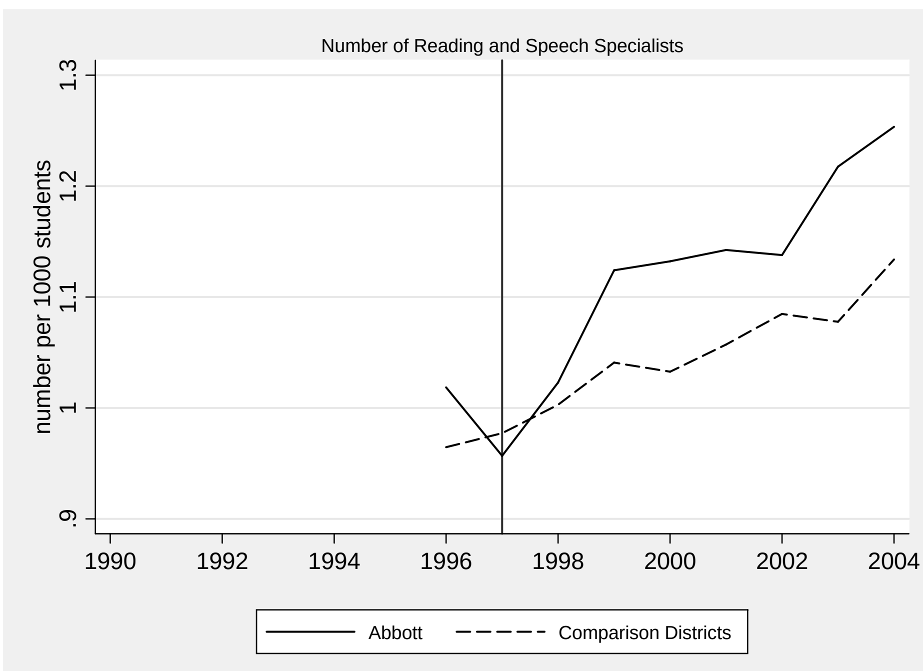


Table 2.1: Revenue and Expenditure Items from F-33 Survey

Revenue and Expenditure Items from the F-33 Form

Revenue

Main Categories		Subitems
Total Local Revenue		
		1. Property taxes
		2.- 5. Other Taxes
		6.- 8. Revenue from cities, counties, other districts
		9. Tuition fees from pupils and parents
		10. Transportation fees from pupils and parents
		11. Textbook sales and rentals
		12. School lunch revenues
		13. Student activity receipts
		14. Other sales and service revenues
		15. Interest earnings
		16. Miscellaneous other local revenue
Total State Revenue		
		1. General formula assistance
		2. Staff improvement programs
		3. Special education programs
		4. Compensatory and basic skills attainment programs
		5. Bilingual education programs
		6. Gifted and talented programs
		7. Vocational education programs
		8. School lunch programs
		9. Capital outlay and debt service programs
		10. Transportation programs
		11. All other revenues from state sources
Total Federal Revenue		
	Through State	
		1. Title I
		2. Children with disabilities – IDEA
		3. Eisenhower math and science
		4. Drug free schools
		5. Title VI
		6. Vocational and technical education
		7. Child nutrition act – exclude commodities
		8. All other federal aid through the state
	Direct to District	
		1. Impact aid (P.L. 815 and 874)
		2. Bilingual education
		3. Indian education
		4. All other direct federal aid

Expenditures

Main Categories		Subitems
Current Operations Expenditures		
	Instructional Programs	
		1. Instruction
		2. Support services, pupils
		3. Support services, instructional staff
		4. Support services, general administration
		5. Support services, school administration
		6. Support services, operation and maintenance of plant
		7. Support services, student transportation
		8. Business/central/other support services
	Non-Instructional Programs	
		9. Food services
		10. Enterprise operations
		11. Other
	Non-Elementary Secondary Programs	
		12. Community services
		13. Adult education
		14. Other
Capital Outlay Expenditures		
		1. Construction
		2. Land and existing structures
		3. Instructional equipment
		4. All other equipment

Table 2.2: Summary Statistics for Estimation Samples

Means of Key Variables For Abbott Districts and Comparison Group, 1997			
	Abbott Districts	Comparison Group	All Other K-12 Districts
N	29	23	162
Enrollment	9351	3389	3317
Federal Revenue per pupil	$699	$424	$189
Local Revenue per pupil	$3,401	$5,470	$8,307
State Revenue per pupil	$7,520	$4,165	$2,267
Compensatory Aid per pupil	$608	$328	$97
Property Tax Revenue per pupil	$1,791	$4,483	$6,767
Total Expenditures per pupil	$11,599	$10,331	$11,136
District Personal Income per pupil	$77,755	$91,671	$180,391
District Property Value per pupil	$200,769	$341,759	$556,957
Effective Municipal Tax Rate*	3.45	2.96	2.57
Equalized County Tax Rate*	0.678	0.579	0.480
Equalized School Tax Rate*	1.336	1.453	1.500
Equalized Municipal Tax Rate*	1.523	0.996	0.621
Total Equalized Property Tax Rate	3.537	3.029	2.602
Class Size	22.16	22.90	21.60
Student-Faculty Ratio	13.66	14.62	13.93
Student-Administrator Ratio	308.28	305.60	322.42
Fraction of Classrooms with Internet*	28.96	24.22	26.59
Student-Computer Ratio*	8.63	14.84	11.69
Student Mobility Rate	4.2%	3.7%	2.2%
Percent Special Ed	4.8%	3.6%	2.6%
Percent Limited English Proficient	9.9%	5.1%	3.0%
Percent Asian	2.0%	2.9%	7.2%
Percent Black	39.4%	32.8%	7.0%
Percent Hispanic	32.2%	17.0%	6.7%
Percent White	26.8%	48.2%	77.5%
Percent Free Lunch	57.8%	33.8%	9.1%

*These variables are measured for 1998, the first year for which I have data.

Table 2.3: First Stage Results for Instruments

OLS estimates of First Stage for Instruments
Y = State Aid per pupil

Instruments
IV 1: Parity aid per pupil and other compensatory aid per pupil
IV 2: Parity aid per pupil
IV 3: Abbott*Post97 indicator
IV 4: Abbott*Post97 and Abbott*Post00

	First Stage Estimates			
	(1) IV 1	(2) IV 2	(3) IV 3	(4) IV 4
Parity Aid	1.172 (0.087)**	1.531 (0.102)**		
Other Comp Aid	1.159 (0.070)**			
Abbott *Post 1997			2,268.454 (155.206)**	1,309.909 (187.810)**
Abbott* Post 2000				1,674.744 (203.431)**
District Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Constant	9,445.348 (319.670)**	12,342.234 (321.133)**	11,951.250 (333.687)**	11,539.487 (320.747)**
Observations	676	676	676	676
R-squared	0.93	0.90	0.90	0.91
F statistic	300.55	227.48	213.62	152.37

Standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 2.4: IV Results for Effect on Total Expenditures

Instrumental Variable Specifications - Y=Expenditures per pupil

Model: Exp per pupil = α + βState Aid per pupil (IV) + γDistrict FE + δYear FE + ε

Entry in table is coefficient on State Funding for the corresponding IV.

	(1) Total Expenditures	(2) 95% Confidence Interval	
OLS	0.642 (0.106)**	0.43	0.85
Instrument			
IV1- parity per pupil and other comp per pupil	0.660 (0.092)**	0.48	0.84
IV2 - parity aid per pupil	0.585 (0.118)**	0.35	0.82
IV3- abbott*post97 indicator	0.697 (0.110)**	0.48	0.91
IV4- abbott*post97 and abbott*post00	0.650 (0.114)**	0.43	0.87
Observations	676		

Robust standard errors in parentheses (clustered at the district level).
* significant at 5%; ** significant at 1%

Table 2.5: IV Results for Effect on Total Expenditures Using Alternate Control Groups

Sensitivity of Effect on Total Expenditures to Alternate Control Groups

			coefficient	std error	N	number of districts	number of years
(1)	from Table 3		0.660	(0.092)**	676	52	13
(2)	Poor K-12 Districts		0.733	(0.065)**	1027	79	13
(3)	All K-12 Districts		0.715	(0.041)**	2301	177	13
Alternative Selection 1: Excluding districts with less than .1 probability of being an Abbott district+							
(4)	sample 1		0.662	(0.094)**	676	52	13
(5)	sample 2		0.605	(0.112)**	689	53	13
(6)	sample 3	Excludes central cities	0.632	(0.166)**	468	36	13
Alternative Selection 2: Weighting by propensity to be an Abbott district++							
(7)	Inverse P-score weights	All K-12 Districts	0.548	(0.165)**	2301	177	13
(8)	Inverse P-score weights	Excludes Atlantic City	0.763	(0.088)**	2288	176	13

Robust standard errors in parentheses (clustered at the district level); * significant at 5%; ** significant at 1%

+Alternate samples were constructed by excluding districts that have less than .1 predicted probabilty of being an Abbott district, using several different logit models. The regressors are listed below.

(4) sample 1 - median hh value in 1989 and 1999, median income in 1989 and 1999, 1996 effective municipal tax rate, 1996 average sale price of homes, 1996 labor_force, 1996 unemployment rate

(5) sample 2 - all in sample 1 plus number of schools, number of teachers, enrollment, fraction white, fraction free lunch, fraction with individualized education plan

(6) sample 3 - same as sample 1 but excluding all districts classified as "central cities". This excludes about half of the Abbotts.

++ Inverse propensity score weights constructed using the predicted probability of being designated an Abbott district, using a logit model on district level characteristics. For non-Abbotts, w = p/1-p. For Abbotts w=1. The weights are rescaled to sum to 1 within group. The variables included in the regression are:

(7) - median hh value in 1989 and 1999, median income in 1989 and 1999, 1996 effective municipal tax rate, 1996 average sale price of homes, 1996 labor_force, 1996 unemployment rate

(8) - same as (7), but Atlantic City is excluded.

Table 2.6: IV Results for Effect on Revenues

Instrumental Variable Specifications - Y=Revenues per pupil

Model: Revenues per pupil = α + βState Aid per pupil (IV) + γDistrict FE + δYear FE + ε

Entry in table is coefficient on State Funding for the corresponding IV.

	(1)	(2)	(3)	(4)
	Federal Revenue	Local Revenue	of which:	
			Local Property Taxes	Effective Municipal Prop Tax Rate
OLS	0.023	-0.161	-0.023	-0.0815
	(0.010)*	(0.047)**	(0.050)	(0.0248)**
Instrument				
IV1- parity per pupil and other comp per pupil	0.014	-0.223	-0.116	-0.0825
	(0.015)	(0.072)**	(0.070)	(0.0409)*
IV3 - abbott*post97 indicator	-0.006	-0.252	-0.144	-0.1053
	(0.020)	(0.117)*	(0.084)	(0.0425)*
Observations	676	676	413	520

Robust standard errors in parentheses (clustered at the district level).
* significant at 5%; ** significant at 1%

Table 2.7: IV Results for Effect on Expenditure Subcategories

Instrumental Variable Specifications - Y=Expenditure (subcategories) per pupil

Model: Expenditures per pupil = α + βState Aid per pupil (IV) + γDistrict FE + δYear FE + ε

Entry in table is coefficient on State Funding for the corresponding IV.

	(1)	(2)	(3)	(4)	(5)	(6)
	Current K-12 Expenditures	of which:		Other Current Expenditures	Non-K-12 Expenditures	Capital Outlays
		Instructional Expenditures	Support Services Exp			
OLS	0.569	0.318	0.242	0.013	0.002	0.060
	(0.087)**	(0.054)**	(0.033)**	(0.008)	(0.005)	(0.069)
Instrument						
IV1- parity per pupil and other comp per pupil	0.774	0.421	0.342	0.014	-0.004	-0.116
	(0.063)**	(0.043)**	(0.028)**	(0.011)	(0.006)	(0.072)
IV3 - abbott*post97 indicator	0.857	0.483	0.365	0.011	-0.004	-0.128
	(0.097)**	(0.064)**	(0.041)**	(0.012)	(0.010)	(0.076)
Observations	676	676	676	572	572	676

Robust standard errors in parentheses (clustered at the district level).
* significant at 5%; ** significant at 1%
Note: Column 1 is the sum of columns 2 and 3. Columns 1, 4, 5, and 6 sum to total expenditures, the variable presented in Table 4.

Table 2.8: IV Results for Effect on Salary Expenditures

Instrumental Variable Specifications - Y=Expenditures per pupil (subcategories)

Model: Expenditures per pupil = α + βState Aid per pupil (IV) + γDistrict FE + δYear FE + ε

Entry in table is coefficient on State Funding for the corresponding IV.

	(1)	(2)	(3)	(4)	(5)	(6)
	Instructional Salaries	Instructional Support	Pupil Support	Admin Support	Operations and Maintenance Support	Transportation Support
OLS	0.110 (0.026)**	0.040 (0.005)**	0.034 (0.006)**	0.003 (0.002)	0.020 (0.005)**	-0.002 (0.003)
Instrument						
IV1- parity per pupil and other comp per pupil	0.174 (0.033)**	0.057 (0.008)**	0.061 (0.009)**	0.007 (0.003)**	0.026 (0.009)**	-0.006 (0.004)
IV3 - abbott*post97 indicator	0.207 (0.042)**	0.053 (0.009)**	0.065 (0.010)**	0.006 (0.004)	0.025 (0.011)*	-0.003 (0.006)
Observations	676	676	676	676	676	676

Robust standard errors in parentheses (clustered at the district level).
* significant at 5%; ** significant at 1%

Table 2.9: IV Results for Effect on Expenditure Indicators

Instrumental Variable Specifications - Y=District Spending Indicators

Model:Spending Indicator = α + βState Aid per pupil (IV) + γDistrict FE + δYear FE + ε

Entry in table is coefficient on State Funding for the corresponding IV.

	(1)	(2)	(3)	(4)	(5)
	Class Size	Student-faculty Ratio	Student-Admin Ratio	Student-computer ratio	Fraction classrooms with internet access
OLS	-0.0906	-0.1827	-4.9794	-0.313	1.332
	(0.0994)	(0.0716)*	(2.9022)	(0.222)	(1.243)
Instrument					
IV1- parity per pupil and other comp per pupil	-0.0682	-0.2830	-4.5571	-0.014	0.483
	(0.1392)	(0.1054)**	(4.5599)	(0.595)	(2.296)
IV3 - abbott*post97 indicator	-0.0860	-0.2960	-6.6365	-3.354	-3.695
	(0.2538)	(0.1485)	(6.1439)	(2.131)	(6.063)
Observations	520	520	520	364	364

Robust standard errors in parentheses (clustered at the district level).
* significant at 5%; ** significant at 1%

Table 2.10: IV Results for Effect on District Staffing

Estimates of Change in Staffing Levels Associated with Abbott Rulings

Model: Staff per 1000 pupils = α + βState Aid per pupil (IV) + γDistrict FE + δYear FE + ε
Coefficient is Effect of $1000 State Aid per pupil (instrumented with IVs 1 and 3)

	Staff per 1000 students	IV 1 - parity aid and other comp		IV 3 - abbott* post97	
(1)	superintendents	-0.0047	(0.0168)	-0.0009	(0.0232)
(2)	principals	0.0195	(0.0470)	0.1082	(0.0651)
(3)	elementary	1.4171	(0.3220)**	1.5456	(0.5313)**
(4)	hs art	0.1077	(0.0473)*	0.0679	(0.0535)
(5)	hs business	0.0645	(0.0325)	0.0416	(0.0446)
(6)	hs english	0.2676	(0.1758)	-0.0823	(0.2646)
(7)	hs foreign lang	0.0159	(0.0572)	-0.0388	(0.0770)
(8)	hs phys ed	0.1234	(0.0543)*	0.0153	(0.0800)
(9)	hs home ec	0.0031	(0.0245)	-0.0008	(0.0354)
(10)	hs industrial arts	-0.0542	(0.0376)	-0.1118	(0.0822)
(11)	hs math	-0.0386	(0.1158)	-0.2324	(0.1717)
(12)	hs music	0.0704	(0.0326)*	0.0315	(0.0542)
(13)	hs science	0.0765	(0.0516)	0.1364	(0.0915)
(14)	hs social studies	0.0858	(0.0420)*	0.1205	(0.0710)
(15)	hs special ed	0.5914	(0.2352)*	0.5908	(0.2742)*
(16)	hs counselors	0.1200	(0.0500)*	0.1262	(0.0775)
(17)	hs librarians	0.0286	(0.0188)	0.0423	(0.0421)
(18)	hs nurse	0.0822	(0.0211)**	0.0747	(0.0339)*
(19)	hs psychologist	0.0369	(0.0231)	0.0361	(0.0316)
(20)	hs social worker	0.2874	(0.0580)**	0.2608	(0.0744)**
(21)	hs reading/disab specialist	0.0391	(0.0438)	0.0422	(0.0730)
(22)	hs speech specialist	0.0669	(0.0448)	0.0337	(0.0607)

Robust standard errors in parentheses (clustered at the district level).
* significant at 5%; ** significant at 1%

Table 2.11: IV Results for Timing of Expenditure Effects

Timing of the Expenditure Effects: 1-, 3-, and 5-year Changes
Model: Change in Exp per pupil = = α + βChange in State Aid per pupil (IV) + ε

Entry in table is coefficient on State Funding for IV =parity aid and other comp aid .

Expenditure Outcomes	1-year change 1997 to 1998		3-year change 1997 to 2000		5-year change 1997 to 2002	
Main Expenditures						
Total Expenditures	1.36	(0.25)**	0.99	(0.46)*	0.76	(0.19)**
with controls	1.20	(0.23)**	0.89	(0.44)+	0.60	(0.19)**
Current K-12 Exp	0.32	(0.09)**	1.14	(0.36)**	0.90	(0.12)**
with controls	0.31	(0.08)**	1.07	(0.34)**	0.78	(0.11)**
Instructional Exp	0.23	(0.07)**	0.70	(0.25)**	0.41	(0.07)**
with controls	0.27	(0.06)**	0.71	(0.24)**	0.39	(0.07)**
Support Services Exp	0.09	(0.05)+	0.43	(0.13)**	0.48	(0.07)**
with controls	0.04	(0.05)	0.34	(0.11)**	0.38	(0.06)**
Capital Outlays	1.00	(0.24)**	-0.11	(0.36)	-0.13	(0.15)
with controls	0.88	(0.22)**	-0.13	(0.35)	-0.16	(0.16)
Revenues						
Total Revenue	0.76	(0.17)**	0.74	(0.28)*	0.86	(0.10)**
with controls	0.78	(0.16)**	0.69	(0.26)*	0.79	(0.11)**
Local Revenue	-0.24	(0.16)	-0.29	(0.27)	-0.16	(0.10)
with controls	-0.23	(0.15)	-0.34	(0.26)	-0.23	(0.10)*
Property Tax Revenue	-0.07	(0.04)+	-0.15	(0.13)	-0.05	(0.06)
with controls	-0.05	(0.03)	-0.14	(0.12)	-0.09	(0.07)
Salaries						
Instructional Salaries	0.11	(0.03)**	0.30	(0.17)+	0.16	(0.04)**
with controls	0.12	(0.03)**	0.33	(0.16)+	0.18	(0.04)**
Pupil Support	0.03	(0.02)+	0.07	(0.03)*	0.08	(0.01)**
with controls	0.02	(0.01)	0.06	(0.03)*	0.06	(0.01)**
Instructional Support	0.01	(0.01)	0.05	(0.03)*	0.08	(0.01)**
with controls	0.01	(0.01)	0.05	(0.02)+	0.07	(0.01)**
Operations and Maintance Support	0.02	(0.01)**	0.08	(0.02)**	0.04	(0.01)**
with controls	0.01	(0.01)*	0.07	(0.02)**	0.03	(0.01)**

Standard errors in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

Note: For each expenditure category, the first row has no controls and the second row adds controls for 1995 total expenditures per pupil and 1995 total district enrollment.

Table 2.12: Summary Statistics

Sources and Summary Statistics for Estimation Sample
All financial variables are per pupil and adjusted for inflation (to year 2000 dollars)

Variable	Source	Dates Covered	Mean	Std. Dev.	N
Federal Revenue	F33	1990-2004	705	339	780
Local Revenue	F33	1990-2004	4332	2335	780
State Revenue	F33	1990-2004	6809	3066	780
Parity Aid	NJ State Aid	1990-2004	317	603	780
Compensatory Aid	F33	1992-2004	1252	1263	676
Comp Aid net of Parity Aid	created	1992-2004	886	793	676
Property Tax Revenue	NCES	1996-2004	2895	2457	468
Total Expenditures	F33	1990-2004	11720	2663	780
K-12 Expenditures	F33	1990-2004	10324	2259	780
Instruction Expenditures	F33	1990-2004	6303	1412	780
Support Service Expenditures	F33	1990-2004	3692	898	780
Non-K-12 Programs	F33	1992-2004	161	117	676
Non-Instructional Current Expenditures	F33	1992-2004	320	121	676
Capital Outlays	F33	1990-2004	716	1107	780
Enrollment	F33	1990-2004	6685	7887	780
Percent Limited English Proficient	NJ Report Card	1995-2004	0.079	0.077	520
Class Size	NJ Report Card	1995-2004	21.28	2.57	520
Student-Faculty Ratio	NJ Report Card	1995-2004	12.92	1.92	520
Student-Administrator Ratio	NJ Report Card	1995-2004	295	66	520
Student Mobility Rate	NJ Report Card	1995-2004	0.039	0.017	520
Percent Special Ed	NJ Report Card	1995-2004	0.049	0.026	520
Effective Municipal Tax Rate	NJ Tax Report	1995-2004	3.17	0.71	520
District Personal Income	NJDOE	1995-2004	89588	75845	520
District Property Value	NJDOE	1990-2004	305913	236463	780
Instructional Salaries	F33	1990-2004	4397	731	780
Instructional Support Salaries	F33	1992-2004	277	149	676
Pupil Support Salaries	F33	1992-2004	554	223	676
Administration Support Salaries	F33	1992-2004	110	63	676
Oper. and Maintenance of Plant Supp Salaries	F33	1992-2004	468	177	676
Transportation Support Services	F33	1992-2004	90	99	676
Equalized County Tax Rate	NJ Tax Report	1998-2004	0.66	0.33	364
Equalized School Tax Rate	NJ Tax Report	1998-2004	1.26	0.32	364
Equalized Municipal Tax Rate	NJ Tax Report	1998-2004	1.15	0.67	364
Total Equalized Property Tax Rate	NJ Tax Report	1998-2004	3.08	0.73	364
Fraction of Classrooms with Internet	NJ Report Card	1998-2004	0.738	0.367	364
Student-Computer Ratio	NJ Report Card	1998-2004	6.71	5.92	364

Source	Description
F33	Annual Census of Governments Survey of School District Finances.
NCES	Downloaded from NCES website. Mostly items from F33 that were not included on the longitudinal file.
NJ State Aid	Annual report of State Aid to districts. Published on NJDOE website.
NJ Report Card	Annual report of student and school characteristics and achievement results.
NJ Tax Report	Annual Report of the New Jersey Division of Taxation. Published on NJ Department of Treasury website.
NJDOE	Requested files from NJDOE Finance Department.

2.7 Data Appendix

2.7.1 Data Sources

The data sources are summarized in Table 2.12 and are described in more detail in this section.

National Center for Education Statistics Data

Census of Governments F-33 Data The school funding variables are taken from the Census of Governments Survey of Local Government Finances School Systems Survey, the F-33 data. I use the longitudinal version of these data produced by NCES (William, McLaughlin, Glander, and Fowler Jr., 2006a), which includes data from 1990-2002, and add the more recent single-year files for 2002-03 and 2003-04. This dataset includes detailed revenue and expenditure breakdowns and enrollment counts and is collected annually at the school district level. The longitudinal file does not include the subcategories of local revenues. I downloaded property tax receipts from the NCES website to fill this in for 1996-2002. The F-33 data are the same as the finance data released by the state in the school report cards. These files can be found at the following internet addresses:

yearly files: http://nces.ed.gov/ccd/f33agency.asp

longitudinal file: http://nces.ed.gov/edfin/links/findata.asp

New Jersey Department of Education Data

State Aid Summary The State Aid Summary is a breakdown of state funding for schools, reported by statutory funding categories. Beginning with the 1997-98 school year, these data are posted on the NDJOE website. I use parity aid figures from these files. The 1997-98 State Aid Summary does not include parity aid because the State Aid data was released before the court mandated parity aid. I construct parity aid figures for 1997-98 by subtracting total state aid in the 1997-98 file from the reported total for 1997-98 in the 1998-99 file. These constructed values line up with media reports of parity aid by district. These data are found at:

http://www.state.nj.us/education/stateaid/

School Report Card The School Report Cards are published annually by the NJDOE and include, among other variables, school-level data on enrollment, staffing, expenditures, results on student assessments. A longitudinal historical file covers the school years 1994-95 to 2001-02. These files do not, however, include data on all items back to 1995. Many items were only collected in later years. Yearly files are available from 2002 to 2006.

I merge school-level data from 1995-2005 into one file. I collapse these data to the district level, summing counts and taking enrollment-weighted means of rates and percentages. The school-level data include vocational, special services and charter schools, but I exclude these schools when collapsing to district level observations.

These data are found at:

http://education.state.nj.us/rc/

Certificated Staff Data The Certificated Staff Data are annual files of all teachers in positions that require certification. The files cover all classroom teachers, because emergency certifications are only used for educational services positions such as school social workers or school nurses. I requested these files from the NJDOE Public Information Office and received them as electronic files on CD.

These data include name, year of birth, salary, experience (overall, in NJ, and in the district), education (highest degree received), and information describing where and what the teacher teaches.

Other New Jersey State data

Division of Taxation Annual Reports provide effective total property tax rates by municipality. These reports are available on the NJ Department of Treasury website starting in 1996 (which provides statistics for 1995).

These data are found at:

http://www.state.nj.us/treasury/taxation/pubs.htm

2.7.2 Alternate Sample Selection in Table 2.5

Trimming Sample of Districts With Low Abbott Probability

Alternate samples were constructed by excluding districts that have less than .1 predicted probability of being an Abbott district, using several different logit models. The regressors that lead to each sample are listed below, but the broad method was to run a logit where Abbott status is the 0/1 dependent variable and the regressors include district characteristics or district and student characteristics. Then the sample is created by excluding districts that have $\hat{p} < .10$. Sample 3 excludes districts classified as central cities in both the logit and the sample selection.

For Sample 1, the logit model includes median home value in 1989 and 1999, median income in 1989 and 1999, 1996 effective municipal tax rate, 1996 average sale price of homes, 1996 labor force, and 1996 unemployment rate as regressors. Sample 2 adds the following school and student level variables: number of schools, number of teachers, enrollment, fraction white, fraction free lunch, and fraction with individualized education plan. Sample 3 uses the same regressors as sample 1 but excludes all districts classified as “central cities”. This excludes about half of the Abbotts.

Propensity Score Reweighting

I construct inverse probability weights using the predicted probability of being designated an Abbott district, using a logit model on 1996 (pre-treatment) district level characteristics. For non-Abbotts, the weight is $w = p/1 - p$. For Abbotts the weight equals 1. The weights are rescaled to sum to 1 within group. The regressors for the logit models that lead to the weights are median home value in 1989 and 1999, median income in 1989 and 1999, 1996 effective municipal tax rate, 1996 average sale price of homes, 1996 labor force, and 1996 unemployment rate.

The goal of this reweighting is to make the non-Abbott districts look more like what the Abbotts would look like had they not received the treatment of additional funding. The reweighting places more importance on non-Abbott districts with a high probability of being an Abbott and less weight on districts with a low probability. This method allows me to use all the information from all districts.

There is one non-Abbott, Atlantic City, with a very high ($> .9$) likelihood of being an Abbott. The student and city characteristics of Atlantic City match the Abbotts quite well, but the city has a much larger tax base than the Abbott districts so did not meet the municipal overburden criteria for selecting Abbotts. However, Atlantic City has a higher weight than any other city and seems to drive some of the results when using this method. I run a second version of the logit model that excludes Atlantic City to generate a second set of weights.

CHAPTER III

New Evidence on School Funding and Student Achievement

3.1 Introduction

The Abbott v. Burke court ruling in 1997 caused a dramatic increase in school funding for 30 poor, urban school districts in New Jersey. This ruling provides a unique natural experiment with which to examine the effect of increased resources on student achievement. The policy has several nice features from a researcher's perspective. The ruling was unexpected and the change sudden; schools were wrapping up the 1996-97 school year when the court released the decision and the additional money was in place for the start of the 1997-98 school year. The change in funding was large, averaging over $1000 per pupil and increasing the regular education budget by about 10% for affected districts. Finally, the change affected only 30 school districts, leaving the statutory school finance system in place for all other districts and creating a natural comparison group.

In 2002, the New York Times Editorial page stated that the Abbott case "may be the most significant education case since the Supreme Court's desegregation ruling nearly 50 years ago" (The New York Times, February 9, 2002). While this may be true, there has been little research into the effects of the Abbott reforms.[1] The

[1]There is a body of research about the effects of the "high-quality" universal preschool that was mandated for Abbott districts in a subsequent ruling. See Frede, Jung, Barnett, Lamy, and Figueras (2007) and Lamy et al. (2005). These papers show increased achievement for the Abbott students but I have not seen a comparison between

Abbott rulings speak of the need to “reduce the disparity in outcomes by reducing the disparity in inputs.” The 1997 decision dramatically increased funding but there is no comprehensive evaluation of the effect on student outcomes. In fact, because of series breaks in the elementary and middle school assessments in New Jersey, it’s not clear that the state has the data required to test whether the Abbott reforms have had an effect. I use the publicly available, but imperfect, district-level data and restricted student-level data to examine the effects of the policy on the one exam that spans the policy change. I supplement this analysis with data on other outcomes at the district level and trends on the state and National Assessment of Educational Progress (NAEP) tests in earlier grades.

There is consensus in the economics of education literature that academic achievement matters for lifetime earnings and adult success (Hanushek, 2003). Urban schools continue to struggle in matching the achievement levels seen in suburban schools and increased funding continues to be a remedy directed to failing schools. One unresolved question is how to use that funding to improve the achievement of students in poor and urban schools. Hanushek (1996, 2003) surveys the literature through 1994 on expenditures and student performance, finding that a large majority of the serious studies find no statistically significant effect of school expenditures on student achievement, and, of those finding significant effects, a sizable portion are of the “wrong” sign. Hanushek also notes that other studies find significant differences among teachers and that teachers impact the performance of students, but there is no reliable link between school spending and the quality or output of teachers. He concludes that resources have not been shown to be a useful policy instrument in improving student achievement. Krueger (2003) criticized Hanushek’s meta-analysis

the Abbott districts and other districts. These papers do support the consensus in the literature that high-quality early education is of great value in improving the education of needy students.

and argues that the evidence on class size reductions in particular has shown consistent positive effects. He used the results from Project STAR, a class size reduction experiment, to calculate the lifetime earnings effect of small class sizes in early grades and performs a cost-benefit analysis. He concludes that even at a cost of around $7000 per student, class size reductions pay off in lifetime earnings and notes that there are other private and societal benefits that are not captured by this estimate.

Most of the literature, including many of the studies Hanushek reviews, are studies of aggregate resources in a state or locality, not direct evidence on the effect of resources in the most needy schools. Studies of specific programs do suggest that particular expenditures can improve performance. Ferguson (1998) summarizes the research on programs or choices within schools that potentially affect the test score gap between black and white students. He reports that there is a consensus that early-education interventions, such as Head Start, the Perry Preschool experiment and the Child-Parent Center, increase achievement and IQ scores for participant students. These measured gains are large, ranging from one-third to one standard deviation, but the literature is mixed on whether these gains are sustained over time. Research on instructional interventions and class size reductions report gains of similar magnitude, with larger effects for minority children. Brooks-Gunn (2003) reviews the literature on early-childhood interventions and also concludes that high-quality interventions increase achievement particularly for poor children and that the effects, while diminishing, do persist through childhood and later schooling. A remaining question is whether the effects of these programs justify their high costs.

Although there is a long and active literature on resources and schools, questions remain unanswered. There are individual and societal benefits to higher educational achievement. Money can matter in increasing achievement, but not all expenditure changes lead to educational improvements. The Abbott case gives an opportunity to see what school districts do when given new resources and free reign in spending choices. The Abbott case was not a test of a carefully controlled program but instead an opportunity for the various Abbott districts to spend the money where they thought it was needed. News reports from the first year of the program described districts and schools that bought books for classrooms that were lacking them and hired teachers aides to help struggling students (Newman, March 29, 1998; Rhor, December 25, 1998). These types of expenditures seem likely to add to the school experience of the treated students, and measuring the impact of these choices would be valuable to policy makers and school officials. As I will discuss later, subsequent Abbott rulings dramatically reduced districts' discretion over spending.

While the natural experiment created by this ruling should be ideal for investigating the effects of money on achievement, New Jersey's lack of attention to the quality and consistency of their school data makes a comprehensive analysis challenging. New Jersey instituted annual statewide assessments in the early 1980s in elementary, middle and high school. Through the early 1990s districts used different nationally available basic skills tests. The statewide elementary assessment did not begin until 1999, making a before-after comparison impossible.[2] Eighth grade students have been assessed since the early 1990s, but two separate and very different exams were in place before and after the reform and there was no overlap period that might allow for norming across the two test regimes. The high school assessment, given in 11th grade, did span the policy change and covers several years in each period; I use this assessment to evaluate the Abbott reform.[3]

The state has recently highlighted improved achievement in fourth grade assess-

[2]District-level results for the earlier 4th grade assessments are available on paper in the state archives in Trenton. I plan to obtain these data in the coming months.

[3]See http://www.nj.gov/education/assessment/history.shtml for an overview of assessment testing in New Jersey.

ments in Abbott schools, but evidence of improvement in higher grades is notably lacking (MacInnes, 2005)[4]. Indeed, 11th grade may be the least likely place to find effects on achievement. If there is a cumulative effect of schooling, for example if the skills tested on the 11th grade test build upon skills learned in each previous year, the Abbott students who are exposed to the funding increase in high school will have only a few years under the high-spending regime and will have obtained most of their schooling with lower funding. Many of the students who could benefit from extra resources may have already dropped out by high school. For the students who have left the system, this policy is unlikely to draw them back, but it might prevent students in later cohorts from dropping out. If the students who are deterred from dropping out are at the lower end of the achievement distribution, this would tend to lower average scores after the policy, working against measuring any achievement improvement for the students who would have been in school under either policy.

I use publicly available district-level and restricted student-level data for the 11th grade assessments. Using the district-level averages on the 1994-2001 High School Proficiency Test, I find a statistically significant positive effect of about one-seventh of a standard deviation on math scores and small and insignificant effects on reading achievement. Using the student-level data to examine these effects more closely, I find large and statistically significant improvements in achievement exclusively for black and hispanic students in the Abbott districts, on the order of one-fifth to one-quarter of a standard deviation. These groups of students consistently score one standard deviation below their white classmates, so these are particularly important gains, reducing the gap that remains after controlling for other characteristics by nearly one-half. In all likelihood this measured effect averages successful programs

[4]NJDOE officials cite a narrowing gap in pass rates between Abbott and non-Abbott districts for the fourth grade assessment, but the data start in 1999 and it is unclear whether the convergence is coming from better performance by Abbotts or weaker performance by non-Abbotts.

and unsuccessful ones. Better data and planning from the NJDOE would allow for a more comprehensive analysis and the identification of which specific programs or expenditures have substantial positive effects.

The remainder of the chapter is organized as follows: Section 3.2 discusses the Abbott case and the expected effects on student achievement. Section 3.3 describes my data and the empirical approach. Section 3.4 presents my empirical results and Section 3.5 concludes.

3.2 The Abbott Case and School Finance in New Jersey

Until 1976, schools in New Jersey were financed solely through local property taxes. As in many states, this led to large disparities in funding levels depending on local wealth, and spurred a series of litigation aimed at increasing funding to low-wealth districts. The 1970 Robinson v. Cahill lawsuit was the precursor of the Abbott case and it challenged the disproportionately high tax burden required for urban districts to provide the "thorough and efficient education" mandated by the state constitution. The court agreed with the urban districts, leading to a new funding formula to be supported by a state income tax which was instituted in 1976.

The Abbott v. Burke case, first filed in 1981, argued that the new state financing plan failed to remedy the funding disparity, and that students in the four urban districts originally named in the case were still not receiving an adequate education. In 1985, the court found in the students' favor, noting that wealthy districts spent 40% more than poorer districts; the court also expanded the districts covered by the suit to a group of 28 poor urban school districts, which are now known as the Abbott Districts.[5]

[5] Two additional Abbott districts, Neptune and Pemberton, were added in 1998 and one, Salem City, in 2004, bringing the current total to 31.

3.2.1 The Abbott IV Ruling Increased Funding to Abbott Districts

The court found in Abbott IV (1997) that the funding levels set in CEIFA were chosen arbitrarily and did not reflect the cost of providing a thorough education. The state had failed to significantly reduce the funding disparity through the mid-1990s, so the court ordered in this ruling that the state fund the Abbott districts at the level of the wealthiest districts in the state. The rationale behind this decision was that the wealthy districts were successfully educating their students, so in the absence of concrete figures for the cost of a "thorough and efficient" education, the expenditures of the wealthy districts were the only available benchmark.[6]

The money that the state must provide to bridge the gap between an Abbott district's regular education budget and the average spending of the wealthiest districts is called "parity aid." In addition to this parity aid, poor schools in general and the Abbotts specifically get significant additional funding from the state based on the characteristics of the local student population. The parity aid is the portion of funding that can potentially be thought of as exogenous. The state had successfully held off the court mandate for more than 15 years, and while funding to Abbotts began to increase in 1992, the funding changes starting in 1997-98 were dramatically larger.

Figure 3.1 shows the money driven by the policy. The initial funding increase is nearly \$1000 per pupil on average and the amount increases to about \$1500 per pupil over the next few years. This is a significant budget increase for affected districts, representing roughly 10% of mean total expenditures per pupil in New Jersey and in the Abbott districts. There is significant variation among Abbott districts in the amount received, with a few districts receiving no parity funding and one district

[6] For more background on the Abbott case and the New Jersey education finance system see the first chapter of this dissertation and the Abbott history on the Education Law Center website at http://www.edlawcenter.org/index.htm.

receiving $1700 per pupil in the first year of the policy. Abbott school districts receive the added funds and decide how to distribute the money throughout the district. In principle, the funds could be targeted very narrowly within districts so only specific students are affected.

3.2.2 Did the Policy Affect Actual Expenditures?

In the first chapter of this dissertation, examining the expenditure effects of this reform, I show that much of the funding increase is spent in schools and in particular is spent on items that could be expected to increase student performance. Table 3.1 summarizes the effects of the reform on expenditures in the Abbott districts. The comparison group used in this analysis and throughout this paper, described in section 3.3.2, contains other poor districts that are most like the Abbott districts. The numbers in this table estimate the effect of $1 of state aid on the listed expenditure category where the Abbott policy is used as an instrument for state aid. The estimate in row 1 says that for a dollar increase in state aid, school expenditures increase by 70 cents. Current K-12 expenditures increase by an even larger amount, about 85 cents per dollar of state aid. As shown in rows 3 and 4, the increased funds go largely to Instructional and Support Services Expenditures.

These expenditures are focused on hiring more teachers and student support staff. The largest numbers of teachers are hired into the elementary grades, reflecting a stated focus on the early grades by the state and the Abbott districts. The Abbott districts hired one extra elementary school teacher per 100 students and showed significant increases in the numbers of high school nurses and social workers. Abbott districts also showed significant increases in hiring of teachers of nonacademic subjects in high school such as art, music and physical education.

3.2.3 Expected Effects of Additional Spending

There are several potential avenues through which the added funds might affect student outcomes. Several types of expenditures affect the classroom environment directly, such as books and other educational materials, hiring or retaining teachers and aides, and expansion or improvement of course offerings. Increasing classroom resources through any of these channels could improve student achievement directly by affecting the amount of academic material students learn. These seem to be the main expenditures that increased with the policy. The Abbott districts did hire more teachers. The early news articles reported that districts hired in various categories and bought textbooks and other instructional supplies (Newman, March 29, 1998; Rhor, December 25, 1998).

Other uses affect the school but not necessarily the classroom; these include spending on facilities, administration, and internal data collection and use. I find little evidence of increased spending in these categories. Some expenditures, like expanded student supports including health care and meals and expanded extra-curricular activities, affect the students directly but outside the classroom. This last set of expenditures is perhaps most likely to affect the composition of the 11th grade class by affecting who drops out and it is also a category on which the Abbott districts were spending money.

There are at least two reasons why effects might also have a time dimension. One is that the effect on the student is cumulative and the other is that the district learns how to use the money over time or has some lag before it can implement new spending or policies. The evolving effect could be nonlinear; money at different ages or grades might have different effects or part of a district's learning process might be learning how to divert this money for other purposes. If districts have clear unmet needs at the time of the policy change, such as unfilled teaching slots or necessary supplies, the funding may have a short run effect as these needs are filled. In the longer term, districts may or may not alter spending patterns to the optimal uses. Goertz and Natriello (1999) report, looking at previous funding changes in New Jersey, remarkably little adjustment to spending patterns. They show that districts tend to increase spending proportionally across all uses, but do not seem to reoptimize.

With enough data, we could get at the time pattern of these effects and the interactions of the effects with grade level. With only one test score, the time effect means that later cohorts of students get more years of the Abbott "treatment." If there is indeed a cumulative effect, the effects should get stronger with more years of the policy. In other words, students who received more years of the parity aid treatment should show larger effects than students with only one year under the policy.

Newspaper articles during the 1997-98 school year reported on how some Abbott districts used the parity aid in the first year. An article about one Abbott district, Perth Amboy, describes that teacher aides were added to large classes, books and other instructional materials were purchased for all classrooms, teachers were paid to conduct intensive after-school remediation classes, and the district hired attendance officers to reduce high absenteeism in particular schools (Newman, March 29, 1998). These reports suggest that, at least in the first year, districts were filling existing immediate needs.

The news reports and my results imply that districts were spending the added money in classrooms, but not exclusively on academic subjects. My expenditure results cannot differentiate among spending at different grade levels within a district, so I cannot predict the expected results for the high school tests for which I have data. The following section describes the data and strategy I use to examine the effects empirically.

3.3 Data and Empirical Strategy

3.3.1 Data

State assessment testing began in New Jersey in the early 1980s with one assessment in the elementary years, one in middle school and one in high school, but the tests changed frequently. The assessments at all levels were completely overhauled beginning in 1998 to reflect the new Core Curriculum Standards that were introduced in 1996. In this paper I focus on student performance on the 11th grade state assessment test (the HSPT) because this is the only test that spans the policy change[7]; testing using the HSPT began in 1991 and continued through 2000. The state reports the average scale score and fraction of students in each district who pass each state assessment in the annual school report card data. An electronic version of these data is available starting with the 1997-98 school year.[8] I extend the 11th grade assessment series back in time using paper copies of the results published by the state in earlier years. I have these district-level scores for the tests given in 1993 to 2000, which gives me 5 pre-reform years and 3 post-reform years. I also obtained restricted student-level results for this exam from the New Jersey Department of Education. These data include individual-level demographic information that allows me to determine whether the policy disproportionately affects different groups of students, but several years of the pre-reform results are missing so I am left with only 2 years of pre-reform data and 5 years total.

[7]The 11th grade assessment that I use is the High School Proficiency Test (HSPT). This was replaced in 2002 by the High School Proficiency Assessment (HSPA), a measurably more difficult test.

[8]The full Report Card dataset includes state assessment results, enrollment, graduation and dropout rates and other indicators of student achievement, as well as some summary data about teachers.

The HSPT was given in October of each school year. Student performance on an October exam is a test of their learning through the previous year, so I relate each October test to the previous school year. This means that the October 1998 exam is the first one for which Abbott districts had received the parity aid "treatment." For all other variables, data labeled 1998 refers to the 1997-98 school year.

3.3.2 Comparison Group

In the achievement analysis I use the same comparison group that I use in chapter 2. To reiterate the selection strategy, I flag districts that have values below the 10th percentile of Abbott districts for percent of students eligible for free lunch, district enrollment, or district unemployment rate, and above the 90th percentile for percent white, district income per pupil, or equalized property value per pupil using the 1996 values for these variables. I exclude districts that have more than two flags. This excludes one Abbott district, 28 of the other 51 districts in the three poorest state SES categories, and all but one of the districts in the five wealthier categories leaving an analysis sample of 29 Abbott districts and 23 comparison districts. Because my empirical strategy uses district fixed effects, the coefficients of interest are identified using within-district variation over time. The comparison group is used to identify the year effects and the impact of the control variables. Alternate selection methods both more and less formal or sophisticated, discussed in chapter 2, result in a very similar sample so I present results based on one particularly straightforward method here. The choice of the comparison group will only affect my estimates if the true relationship between achievement and the control variables differs between the Abbott and comparison groups. The results presented in the following sections are not sensitive to the exclusion of various outlier districts.[9]

[9]Specifically I separately exclude Hoboken, the wealthiest and most successful Abbott district, and exclude all "large" districts.

3.3.3 Characteristics of Sample and Patterns in Outcome Variables

Table 3.2 presents means of key variables for Abbott districts and the selected comparison group in 1997, the school year before the policy change. The first two sets of variables show characteristics of the communities and schools, measured at the district level. The Abbott districts are slightly poorer and have much lower property wealth than the comparison districts. Enrollments are much higher on average in Abbott districts, although this is driven by a handful of large cities. The characteristics of students are fairly similar - Abbotts have more limited English proficient students than the comparison districts but similar numbers of special education students and similar attendance and student mobility rates. The Abbotts do slightly worse than the comparison group on the cumulative promotion index (CPI), a measure of student promotion (and implicitly retention) used here as an alternative to the dropout rate which is highly inaccurate in these data.[10]

The next group of variables includes the main outcome variables at the district level. The math scale score is the district average score on the math assessment. The scale score is an overall score for each subject that combines scores on several subsections and is scaled (or normed) to be comparable across different years of the test. This ranges from 100-500, with 300 being the cutoff for proficiency. The math proficiency rate is the fraction of tested students in the district who score above this cutoff. More than a third of students in Abbott districts are not proficient in math or reading. The comparison districts do better with only about 20 percent of students failing to reach proficiency. The number of regular-education students tested for each subject is also reported. This excludes special education and limited English proficient students who are also excluded from the reported scores. Statistics for

[10]The reported dropout rate is 1.9 on average for both groups, a number that is not credible. More information in the CPI is included in the data appendix.

performance and participation on the SATs are also included. The Abbott districts perform roughly one standard deviation below the comparison districts on all of the achievement measures listed here.

The final panel shows the assessment scores and student demographics calculated from the student-level assessment data. The micro data show much greater variation in assessment scores than the district level aggregates, with student scores spanning the full range of possible scores. The Abbott districts also have larger fractions of students eligible for Title 1 funding, which is directed to students who are both low-achieving and low-income. The patterns in these data and the interactions between demographics and test scores are explored in figures 3.2 and 3.3.

Figure 3.2 shows the patterns in math and reading scores on the 11th grade assessment for Abbott and comparison districts. These numbers are the raw averages calculated from the student-level data. Note that there is no data point for 1996; each other year from 1995-2000 is covered. On average, math scores rise over the period and reading scores are flatter and decreasing in the last few years. The differential patterns after 1997 are coming through increasing math scores for the Abbott districts, while the reading scores are falling for the comparison group.

Figure 3.3 shows the same outcomes by student ethnicity and then by ethnicity and Abbott status. Panel A of this figure highlights the large achievement gap between white and minority students. Black students score more than 80 points lower than white students and hispanic students do only slightly better than blacks. These gaps are roughly one standard deviation of the distribution in the full sample. This gap is in line with estimates of the achievement gap nationally. This panel also shows that the post-policy increases in math scores are coming from black and hispanic students. Panel B suggests that these achievement improvements are concentrated in the Abbott districts. Panel B also highlights that achievement gaps exist between Abbott and non-Abbott students of the same race. This gap, at about 20 points, is much smaller than the gap between white and minority students. The reading scores are less dramatic, but also show significant achievement gaps and an uptick for black and hispanic students in the years after the policy change.

These time patterns suggest that the Abbott policy had an effect on achievement on the 11th grade assessment, at least for some subgroups. The following section discusses the empirical strategy that will test whether these changes are statistically significant and whether they can be attributed to the policy.

3.3.4 Estimation Strategy

The ideal assessment of the effect of additional spending on students would require longitudinal data by student with multiple assessments and information on the school attended by each student for all years (including nontested years). With longitudinal data by student and tests at various grade levels, I could use student fixed effects to control for unmeasurable factors like innate student ability and motivation that are constant over time but impossible to measure and control for directly. Comprehensive data would also allow me to identify and control for teacher and school effects. Even without multiple test scores for each student, longitudinal enrollment information would allow me to more precisely define the amount of policy that each student was exposed to. The theoretical model that underlies this strategy is that individual student achievement at any grade is determined by the accumulated effects of school and family inputs over all previous years. This is the production function model described in Todd and Wolpin (2003). The Abbott IV ruling creates an exogenous increase in the school-based inputs available to the affected students. This exogenous change allows me to identify the effect of increased school resources on student achievement. Todd and Wolpin (2003) lay out the important distinction between identifying parameters of a production function for education and identifying a policy effect that does not make clear the causal mechanisms driving the effect. In my main analysis I am estimating a policy effect that combines the effects of many different choices and programs in different school districts. The additional analysis attempt to further examine what factors within districts might be driving these changes.

The main shortcoming of the available data is that the state does not follow students over time.[11] Without longitudinal data by student it is impossible to control for the unobserved student characteristics. This will not bias the estimate of the effect of the policy unless the policy also alters the composition of the student body. Ideally, even without longitudinal data, we would like to know if each student in the cross section has switched schools or districts in the last year. This would allow us to estimate whether there is a different effect for movers and stayers. In the actual data, district level mobility is defined only as the number of students who leave or enter during the year.

A second shortcoming of the actual data is that the full series of results is available only for the district-level aggregates. The student-level data are missing three of the pre-policy years. I first use the district-level data to take advantage of the longer time series and then use the more detailed microdata to investigate whether the effects differ across student characteristics.

To evaluate the effect of the policy on student test scores using the district-level data, I estimate the following equation:

$$AverageScaleScore_{jt} = \alpha_0 + \beta Abbott * Post1997_j + \alpha_j + \epsilon_{jt}, \tag{3.1}$$

[11] Longitudinal student-level data exist statewide in New Jersey starting in 2004. The districts with the best record keeping might have them for some period of time before that date.

where j indexes the district and t indexes years. The district fixed effects, α_j, control for stable differences between districts over time. Abbott*Post1997 is an indicator variable for being in an Abbott district while the policy is in effect.

To tease out the time path of the effects I use two alternate specifications. The first adds the term $\gamma YrOfAbbottFunding_j$, which is a linear trend for Abbott districts that turns on with the policy change. Here β measures the immediate effect of the Abbott funding policy on assessment scores and γ is a linear learning effect. The time path could be more flexible. The second version adds separate indicator variables for each year of the policy and for being an Abbott district affected by each year of the policy, allowing for different effects for each year of the policy. For each specification of the policy effect, I also estimate models that include the few district-level control variables that are plausibly unaffected by the policy. The student-level version adds demographic information (race, gender, and Title 1 status) about the student and interactions of these variables with the policy indicator.

With the district-level data, several assumptions are necessary for this estimated β to capture the true effect of the policy on achievement. First, student mobility and the effects of mobility on test scores do not change as a result of the policy. Second, the likelihood of dropping out conditional on ability or achievement is not changed by the policy. Both of these assumptions deal with the composition of the tested group. If the policy changes the ability distribution of students who are tested, any measured effect could be the result of this composition change. To get at the causal relationship between money and student achievement we want to measure the effect on students who would have been tested under either regime. The student-level data includes demographic information on each student that allow me to relax these assumptions. In the specifications that use these more detailed data, the necessary assumption is that these variables fully control for any changes in composition.

3.4 Results

This section presents the empirical results. First I present the effects on the 11th grade assessment at the district and student level and then look at other high school outcomes and potential mediating variables. To get at the effects on earlier grades I also present trends on the New Jersey assessments in fourth and eighth grade and the state NAEP.

3.4.1 The Abbott Reform Increased Math Achievement

Table 3.3 presents results for the effect of the Abbott reform on 11th grade math scale scores.[12] Panel A shows the basic relationship between money and achievement and highlights the mistake that observers make when looking at this policy from afar. The OLS results in column 1 show a negative relationship between the policy and achievement. This is the raw correlation that newspapers and critics of the reform frequently highlight: the Abbott districts perform worse than the other districts, even after receiving a large infusion of money. Column 2 adds district fixed effects and is the weakest version of being careful about the counterfactual. Controlling for the average performance of each district over the whole period, β is positive and statistically significant. The policy improves math scores in the Abbott districts by nearly six points. This is about fourteen percent of a standard deviation in the district-level data. The post 1997 indicator shows that all districts score about seven points higher on average after 1997.

Columns 3 and 4 present the two alternate specifications that tease out the timing of the effects. Column 3 adds a linear trend interacted with the policy indicator.

[12] All standard errors are clustered at the district level to allow for correlation within district over time.

This variable equals zero for all districts for the pre-policy years and equals one for Abbott districts in 1998-99, the second year of the policy, and increases by one each additional year. The Abbott*Post1997 indicator captures the immediate effect and this time trend allows for an evolving effect over time. Abbott students in later years receive a combination of these two effects, β and γ. This allows for the fact that 11th grade Abbott-district students in 2000 received three years of schooling under the new regime while students in 1997-98 only received one. The addition of this time trend does reduce the size of β but the effect remains positive. γ, the coefficient on the trend is positive, with a point estimate of 1.53, but statistically insignificant. Over time, however, this effect accumulates so for students in 2000, the total effect is β plus 2 times γ or 5.85, the same as the estimate of β in the previous column. Because this trend is linear it only identifies whether the effect of the policy is growing or shrinking over time. Column 4 breaks this linear trend into effects for each year. The point estimate is biggest for the second year of the policy, although the three coefficients are not statistically different from one another. The coefficient on each Abbott*year indicator is interpreted as the difference between the average score for Abbotts and the statewide average, controlling for each district's overall average through the district fixed effects. In 1998, Abbott districts score 2.9 points higher than the comparison group, 8.7 points higher in 1999 and just under 6 points higher in 2000. In this specification only the second year of the policy is significant, but the year effects show statewide increases in math scores in the post-policy years.[13]

Panel B adds the district-level control variables that are unlikely to be affected by the policy.[14] These variables are district-level enrollment and enrollment squared

[13]An alternate version of this specification adds a statewide annual trend to control for the possibility that the measured changes are coming from an overall trend in scores and are unrelated to the policy. The estimates on the Abbott*year variables do not change significantly in this alternate specification.

[14]Table 3.8 examines the effects of the policy on potential mediating variables.

to control for changes in district scale over the time period. District size does have a small positive but diminishing effect; the net effect is positive at the average enrollment for each group. The estimate of the effect of the policy is slightly smaller, but is still nearly 5 points in column 5. Columns 6 and 7 add the timing variables and the results are very similar to those in panel A. The policy has a small effect in the first year, 2.3 points or less than one-tenth of a standard deviation, a much larger and statistically significant effect of 7.8 points in the second year, and the effect falls in the third year to 4.4 points.

Panel C presents the same regressions for the restricted set of years that matches the data available at the student level. As described above, several years of the pre-reform student-level data are unavailable. There are 5 years of data in the student-level versions compared to 8 years in the unrestricted district-level regressions. For the policy variables, the move to the restricted sample does not dramatically affect the results. Comparing the estimates in Panel C to those in Panel B, the point estimates change slightly but the qualitative story is similar. The effect of the policy is slightly larger using the restricted sample.

Table 3.4 presents estimates for the remaining HSPT outcomes at the district level. In this table I present results for the basic policy effect, using just the indicator for being affected by the policy, and the results that break out each year separately. In each case I present the results for the full sample and for the restricted sample; the full sample is first in the odd numbered columns and the restricted sample follows in the even columns. Panel A shows results for the fraction reaching proficiency on the math exam. The pattern is the same as for the scale score, but the magnitudes are slightly smaller with the largest estimate around five points as the dependent variable is measured on a smaller range from 0 to 100. Panels B and C present the corresponding results for the reading assessment. For reading, the effect of the policy, shown in column 5 for the scale score and column 9 for proficiency, is positive but statistically insignificant. On average, the policy seems to have little effect on reading performance. The results in columns 7, 8, 11 and 12, which break out the year effects, show that the effect is evolving over time. These results mirror the trends seen in figure 3.2 where the performance of the comparison group deteriorates in the later years and performance for the Abbotts increased somewhat. In contrast to the math results, the reading results do show a growing, although statistically insignificant, impact of the policy for Abbott districts. In all cases, the estimates of policy effects are not significantly impacted by use of the restricted sample.

3.4.2 The Abbott Reform Increased Math and Reading Performance for Black and Hispanic Students

Table 3.5 uses the student-level test scores to estimate the effect of the Abbott reform and allows identification of differential effects for different groups of students. I restrict the sample using the same criteria that the state uses for their published reports. Only first-time 11th grade test takers are included, and special education and limited English proficient students are excluded.[15] The specifications included here mimic those presented in table 3.4 for the aggregated data.[16] Panel A presents the math results. Looking at column 1, the effect of the Abbott policy on math scores is about 5 points and statistically insignificant in this specification, but the coefficients on the covariates highlight the disparities between different groups of students. Black and hispanic students perform far worse than their white counterparts on the math exam. Female students do worse than males on the math assessment. Black girls perform significantly better than black boys but hispanic girls do worse

[15] Inclusion of these students and indicator variables for their status does not affect the main results. The performance of the special education and LEP students is not differentially affected by the policy.

[16] In specifications with no control variables, the β estimates are nearly identical to those in tables 3.3 and 3.4.

than hispanic boys.

Although the size of the policy effect is similar to that seen in the district-level data, the time pattern is negative when using the micro data. The results in column 3 shows a 7 point gain in the first year of the policy and then a drop off in both magnitude and significance in the 2nd and 3rd years of the policy. Again, these effects are not significantly different from one another. In contrast, the overall trend given by the year indicators shows monotonically increasing performance statewide.

The even numbered columns add interactions of the policy indicator with the student-level demographics. In both cases β falls in size and significance and the interactions of the policy with black and hispanic are quite large and statistically significant at the 0.01 level. This pattern is consistent across the different specifications. The policy improves the scores of black and hispanic students by around 17 points on the math assessment. This is one-fifth of a standard deviation in the math score and about one-quarter of the achievement gap between white and minority students highlighted in figure 3.3. It represents an even larger fraction of the achievement gap after controlling for other characteristics. This impact is 40% of the coefficient on race for black students and 60% of the coefficient for hispanic students.

Panel B repeats the same specifications for the reading scores. The demographic variables show that the relationship between student characteristics and reading achievement is similar to the relationship with math achievement with one exception. Girls perform significantly better than boys on average on the reading exam while boys do better on the math exam. The main effects of the policy, shown in the odd columns for the various specifications, conform with the results from the district-level data. There is a small and weak effect on reading scores, but the time path is positive and contrasts with a negative trend statewide. When the policy is interacted with the demographic variables the coefficient on the main policy effect becomes more negative but remains statistically insignificant and retains the positive time trend. The individual level data does reveal that there is a positive effect of the policy for the black and hispanic subgroups. The magnitude is about 12 points, which is smaller than the effect on math scores, but represents about one-quarter of the pre-policy gap between Abbott and non-Abbott achievement.[17] For both races, it is about half of the gap measured by the uninteracted coefficient on race.

3.4.3 Other District-Level Outcomes - Enrollment, Attendance, SATs

Although the HSPT is the only state exam that is consistent over the policy change, there are available data for other indicators of student achievement. The following two tables examine the effects on other variables at the district level that may be outcomes of interest. These include SAT participation and achievement which are hard to interpret as direct achievement measures because students elect to take the exam, but provide some information about plans for college-going and the performance of college-bound students. The second set of outcomes includes several enrollment and retention measures.

Table 3.6 presents effects of the policy on participation and performance on the SAT exams. Looking at the odd numbered columns, which use the policy dummy, there is little effect of the policy on participation or average performance, but the point estimates for all variables except the 75th percentile math score are negative. The time pattern shows that the policy dummy is averaging positive and negative effects in different years. The time pattern in participation and scores supports the interpretation that increased participation is causing the lower scores. If new

[17] When the specification with year*treatment indicators is run on each race separately there is no evidence of an increasing effect over time for black or hispanic students in reading or math. For white students, there does seem to be some positive trend for reading.

participants are coming from the lower part of the achievement distribution, they would be expected to reduce average scores. Participation for Abbott districts is highest, controlling for average participation through the district fixed effects, in 1998 and falls in 1999 and 2000. The average math SAT scores show the exact opposite pattern. The 25th percentile math score shows the largest effect, -12.6 points or about one-third of a standard deviation, in 1998 when participation is the highest, suggesting that greater numbers of low-scoring students are taking the test.

Table 3.7 presents the effect of the Abbott policy on other district level outcomes. The first two outcomes examine whether the tested class is changing in size, both measured in absolute terms as 11th grade enrollment and relative to each cohort's 9th grade enrollment two years earlier. For 11th grade enrollment, the estimate in column 1 suggests that 11th grade enrollment falls by about 46 students or 10% of the Abbott average. This appears to be coming from a change in cohort size as there is little effect on 11th grade enrollment as a fraction of lagged 9th grade enrollment. Columns 3 and 4 show the effect on 11th grade enrollment as a fraction of each cohort's ninth grade enrollment. The estimate using the policy dummy is 0.8 percent, suggesting that the enrollment changes are occuring within each cohort; the more flexible time effects show that there are overall declines for all districts. The Abbotts show a postive, although statistically insignificant, trend indicating that the within cohort effect of the Abbott policy on retention may be positive. Columns 5 and 6 give the effect on the cumulative promotion index, an alternative to the dropout rate that uses enrollment data from all four high school grades to estimate the likelihood that a ninth-grade student will graduate on time. The estimated effect of 8.8 in column 5 is quite large, about 13% of the pre-policy average, and column 6 shows that this average effect is masking a more dramatic trend in the year effects.

The effect on the CPI is growing for Abbott districts from about 5 points in the first year of the policy to 14 in 2000, relative to statistically significant declines in the promotion index for all districts in the sample. The effect on the attendance rate is also positive and statistically significant, but small at less than one-third of one percent and the year effects show that the effect for the Abbotts is only statistically significant in the first year. Taken together, these results suggest that more enrolled students are in school, both on a daily basis measured through attendance and in the longer term measured through retention, but that enrollment overall has fallen in Abbott districts.

3.4.4 Effects on Mediating Variables at the District Level

To examine the paths through which the Abbott policy affected student achievement, table 3.8 presents the effects of the policy on potential mediating variables at the district level. These data are all reported by the school districts and published by the NJDOE in the school district report cards. As in the previous tables, two regressions are presented for each outcome variable. Columns 1 and 2 show the results for average class size. The estimate using the policy indicator is -0.12 and is statistically insignificant, implying that class size fell by a negligible amount for Abbott districts. Column 2 shows that class sizes were falling for all districts, and by an additional amount for the Abbotts. The implied effect of the policy is small in this specification, but does increase over time. The results for the student-faculty ratio in columns 3 and 4 are similar, but are larger than the effects on class size and even larger as a fraction of the baseline level. The class size and student-faculty ratio results both show that these ratios were improving for the comparison group as well as the Abbotts over this period. The point estimates imply that by 2000, class size decreased by 1 student for all districts and 1.3 for the Abbotts while the student-faculty ratio fell by 0.9 for the comparison group and 1.73 for the Abbotts. The effect on the student-administrator ratio are larger in point estimates, and are about 5% of the baseline of about 300 students per adminstrator. In contrast to the pattern for class size and student-faculty ratios where the effects for Abbotts reinforce the overall trend given by the year indicators, the effects on the student-administrator ratio in Abbott districts offset an increasing ratio for all districts. The comparison districts show a sizeable increase in the student-administrator ratio and the effect grows over the time period.

The remaining outcomes in columns 7 to 12 are district-level measures of student characteristics. Student mobility is reported as the number of students transitioning in or out of a district during the school year and the rate is calculated by dividing mobility by total district enrollment. Mobility is increasing for Abbott districts through the policy years, but the effect is small and not statistically significant. The special education rate shows a positive but statistically insignificant effect when the policy is estimated using the policy dummy, but the effect is smaller and decreasing when measured by year. The percent limited English proficient is relatively flat, with no significant change using either specification.

Looking at the results in table 3.8, the main effects seem to be decreases in the student-faculty and student-administrator rates and a modest decrease in class size. The class size change is relatively small when compared to the class size interventions reviewed by Krueger (2003). It is important to note that these results are at the district level, so may not accurately reflect changes at the high school level. It is possible (although unlikely) that the class size reduction in tenth grade was large enough to create the large achievement gains described above.

3.4.5 Trends in Earlier Grades

While these high school results are promising, the news coverage and my previous work show that the bulk of the expenditures went to the elementary grades. Because there is no consistent elementary school test until 1999, I cannot assess the impact of the policy on achievement in early grades. Figures 3.4 and 3.5 show the trends in proficiency rates for the 4th grade exam that began in 1998-99. The general upward trend for both groups in both subjects is explained by the fact that this was a new exam introduced to test proficiency relative to the new curriculum standards. The content and scoring of the test was likely adjusted over the period to better align with the state's definition of proficiency and over time schools and teachers learned how to prepare for this assessment.[18] The Abbott districts do seem to be closing the gap in the math assessment as the gap falls from 13 percentage points in 1999 to 5.7 points in 2004, but the change is not as dramatic as the overall trend. The differential trend in the reading exam is less clear, although the gap is smaller in 2004 than in 1999, dropping from 13.5 to 5 percentage points.

Figures 3.6 and 3.7 show the trends for the eighth grade assessment that began in 1999. As with the 4th grade test, the overall trends overwhelm any change in the gap between Abbott and non-Abbott districts. The general trend is negative for both subjects. The gap in math scores falls from 11.6 to 5.6 percentage points from 1999 to 2004 and the gap in reading scores falls from 9 to 7.

Table 3.9 shows the trends in aggregate New Jersey NAEP scores since 1992. The data presented here are average scale scores for the 4th grade math and reading assessments and the 8th grade math assessment.[19] Where results are missing for a

[18]This is often criticized as "teaching to the test" but this is the goal of assessments aligned with curriculum standards. The assessment is meant to accurately test the material the state has stated that students should learn.

[19]8th grade reading is excluded because there are no data for the pre-policy years.

test in a given year it is because that test was not given in that year or New Jersey did not meet the reporting requirements for that assessment. The scores cannot be split into Abbott and non-Abbott districts, but the school location designation of "Central City" provides a proxy for Abbott districts while "Urban Fringe" covers all other districts in New Jersey. The 4th grade scores show the gap between performance in cities and that in the other areas decreasing markedly over the time period. The gaps between black and white students and hispanic and white students show similar decreases over the time period. Although these data on achievement in the earlier grades are not ideal, they suggest that the Abbott districts are catching up to the achievement level of other districts in New Jersey.

3.5 Discussion and Conclusions

The results presented in the previous section show that there is a significant positive impact of the Abbott policy on 11th grade achievement in the Abbott districts. In the aggregate data, the results are fairly modest, with an impact of at most 5 points. This average effect for all Abbott students obscures the fact the improvement is much more dramatic for the groups of students who most need improvement. Using student-level data I show that there are significant improvements for minority students in both math and reading. The impact is roughly one-fifth of a standard deviation in test scores and one-quarter of the achievement gap between white and minority students without considering Abbott status. These effects are quite large. My estimates of 0.2-0.25 standard deviations of math and reading scores are similar in size to the 0.2 that Krueger (2003) uses as an estimate of the effect of four years of reduced class size based on the results of Project STAR. The measured effect here is almost surely an average of large positive effects in some districts and zero (or negative) effects in other districts. Better data about the specific expenditures for each district would allow for identification of the successful policies.

The effects on the HSPT are about half the size of the 0.5 standard deviation effects of the Success For All program on elementary school reading (Ferguson, 1998). Because the elementary and middle school assessments began in the 1998-99 school year and we have no measure of the pre-policy gap, I cannot measure the effect of the policy on those grade levels. The trends in the assessment results are not strong or conclusive, but do offer supporting evidence that the Abbott districts are improving relative to the comparison group.

3.5.1 Are the Effects Too Large?

The effects described here are quite large, given that they are measured within the first few years of the policy and in high school where we might expect little effect. One might question whether these effects are too large. My related work on how the Abbott money was spent showed that districts hired more teachers and instructional support staff, and teachers in Abbott districts confirm that this was the most immediate change. Newspaper articles also reported spending on books and for alleviating classroom crowding. The retention and attendance results together suggest that students are spending more time in high school. If this time in school is productive, this could explain the achievement impact even without any specific instructional intervention.

Although the class size change seen for Abbotts is too small to have created these large achievement effects, this modest reduction in class size is combined with increased individual instruction and remedial help by teacher aides targeted to the students. It is possible that these changes together could have large effects, particularly when targeted to students in need of assistance. A cynical interpretation is that the districts were focusing only on students that could make a large gain, perhaps those closest to the pass/fail threshold. The evidence here does not differentiate between actual knowledge gains by the affected students and short-term gains from specific test preparation, so I cannot judge whether these effects are likely to persist for the individual students. While the HSPT is a high-stakes test from the student perspective because it is a graduation requirement, there was no district-level accountability policy in place during my sample period so there is no district-level incentive to inflate scores artificially.

It is notable that the policy seems to affect only minority students. This is consistent with previous research that has shown that effects are particularly strong for poor and minority students. This is true of the early childhood interventions surveyed by Brooks-Gunn (2003) and the class size interventions described by Krueger (2003).

3.5.2 Why Doesn't the Effect Increase Over Time?

When the timing effect is split using year indicators instead of the linear trend, the effects are in most cases flat or even diminishing. If districts were learning how to use the money more effectively or students were benefiting from a cumulative positive effect over several years of increased funding, we would expect the impact to be increasing over time. One explanation is that the districts learned how to divert the resources away from educational purposes over time. Another is that education expenditures at the high school level do not have a cumulative effect, either because there is a limit to what students can learn or because there is a limit on what school districts know how to change. Perhaps alleviating the clear needs in the Abbott schools by fully staffing classrooms and making sure that all students have books caused a 20 point increase for affected students, but districts did not make any changes beyond that.

Another explanation, cited by officials in Abbott districts, is that the state added mandates in the years after 1998 for how the money was to be spent. The two main mandates added in 1998 were universal preschool and comprehensive school reform for Abbott districts. These were to be in effect by the 2000-01 school year but some districts began immediately; these programs were reportedly fully funded over and above the parity aid. Even with full funding, the districts may have shifted attention and possibly resources to the lower grades and away from high schools. The court's attention to early grades is attributed to the fact that there is a body of evidence about successful interventions at early grades but no comparable evidence about higher grades. The court did not intend for districts to ignore older students, but the lack of guidance about these grades may have had that result. The implication in this story is that the district's unconstrained choices in the first years were useful but that funds were diverted away from the high school grades in later years. In effect, all students covered in the HSPT sample would have received the additional funding for one year but each cohort would have received it in a different grade. The oldest affected students, who took the HSPT in October 1998 would have had the added funding in 10th grade. The next cohort would have had the added funding in the same calendar year but that would be their 9th grade year.

An implementation study of the comprehensive school reform supports the idea that Abbott districts diverted money and attention to elementary schools to the detriment of other schools in the districts. Erlichson (2005) describes the implementation and eventual abandonment of comprehensive school reform (CSR) for Abbott elementary schools. Different schools chose different CSR packages, including well-known programs such as Success for All and Community for Learning. These packages each prescribe their own curricula and associated staffing levels. For explaining the effect of this reform on high schools, one important detail is that the CSR implementation included the use of school-based budgets for the elementary schools. While the NJDOE provided significant additional funding to support this reform, the affected schools chose CSR models and created their budgets “without district interference” and without considering district-wide needs and constraints. These budgets were submitted to and approved by the NJDOE based on the requirements of each CSR package. Erlichson (2005) reports that larger Abbott districts had as many as seven different CSR models under implementation at once. It is easy to imagine that this would cause a drain on district resources, both financial and in terms of leadership. The first cohort of schools implemented CSR for the 1998-99 school year, after one year of unrestricted parity aid money for Abbotts, and additional schools started CSR in the following two years. The Abbott preschool program, leading toward universal preschool for all 3- and 4-year olds in Abbott districts began in the 1999-00 school year, further diverting attention away from secondary schools (Lamy, Frede, Seplocha, Strasser, Jambunathan, Junker, and Wolock, 2005).

3.5.3 Are the Effects Worth the Money Spent?

Compared to other estimates in the literature, these effects come quite cheaply at less than $1500 per student per year. Krueger (2003) estimates that the cost of the class size reduction in the Project STAR experiment was $3500 per pupil per year, which is substantially larger than the $1000-1500 per pupil per year of parity aid. At that cost, he estimates that 0.1 standard deviation is the critical effect size for the Project STAR program to break even, assuming costs of just over $7000 for two years of schooling in a small class and a lifetime earnings boost of 8% from a one standard deviation increase in math or reading achievement in the elementary grades. He argues the earnings effect would be larger if the achievement gain is measured later in school. These effects are only the private benefit and the aggregate social benefits are likely to be positive as well.

3.5.4 Does Money Matter and What Do We Learn from Abbott?

The evidence I present in this paper, although not ideal, clearly suggests that some students have made dramatic improvements in achievement in a short time. Combined with the consensus that interventions can have a big effect on achievement in early grades (Krueger, 2003; Brooks-Gunn, 2003; Ferguson, 1998) it is clear that money does matter for achievement. My results suggest that targeted funds can have a big impact even in high school. The remaining question is which policies and programs are worth supporting given limited budgets. To answer this, researchers need to look more closely at specific policies and interventions. Within each case there are successes and failures, but the large meta-analyses tend to average these effects and find zero or small effects on net. Ideally researchers will continue to run randomized, controlled experiments in education so we can be confident of the parameters being manipulated and the mechanisms behind the effects.

Researchers can still learn from non-experimental cases, but we need to use all the data available to look within policies to learn what pieces are actually working. This paper attempts to do this in the Abbott case, but the lack of appropriate data makes it impossible to make conclusions about specific expenditures. The first lesson from the Abbott reform is that if a state wants to measure the achievement results of a policy convincingly and cleanly, it must keep a consistent testing regime in the years around the policy change.

The second lesson is that it is important to think about the broader effects of each policy. In the Abbott case, the attention to elementary grades in the later court rulings had the unintended effect of diverting attention and resources from secondary grades. In particular, the implementation of school-level reforms without regard to the district- and state-level budget constraint seems to have been a main factor in derailing the gains in high schools.

This particular policy is not sustainable, both because of the large financial cost and because it sets all other districts in the state against the Abbott districts in the allocation of state resources. Indeed, New Jersey has passed a new school finance law that attempts to remove the Abbott designation; the NJ Supreme Court has yet to rule on the constitutionality of the new law. Although the state failed to set up a credible evaluation when the policy began, it could still learn from the Abbott policy by using internal data to identify which districts and schools were successful and why. New Jersey and other states could learn from the Abbott experience and focus resources on the more successful programs and expenditures.

3.6 Figures

Figure 3.1: Parity Aid by Abbott Status

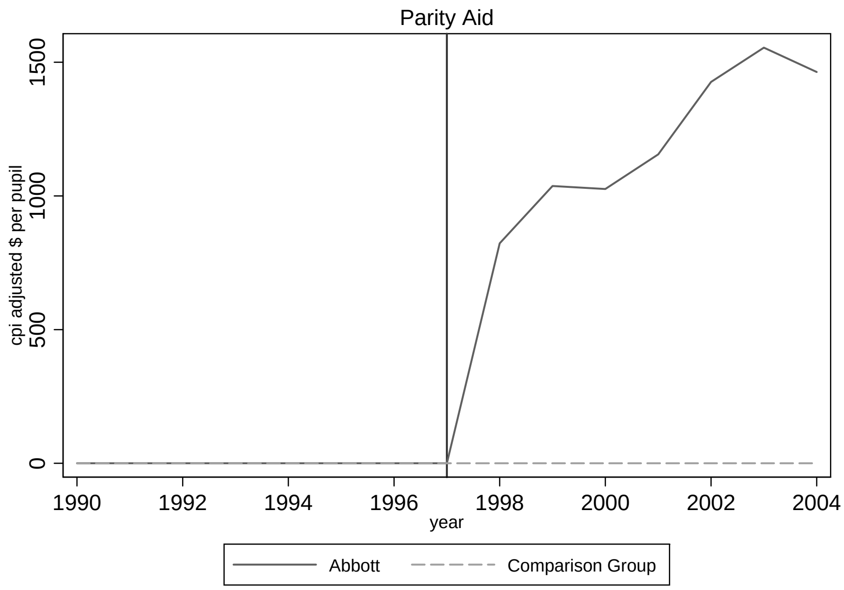


Figure 3.2: Average Math and Reading Scale Scores by Abbott Status

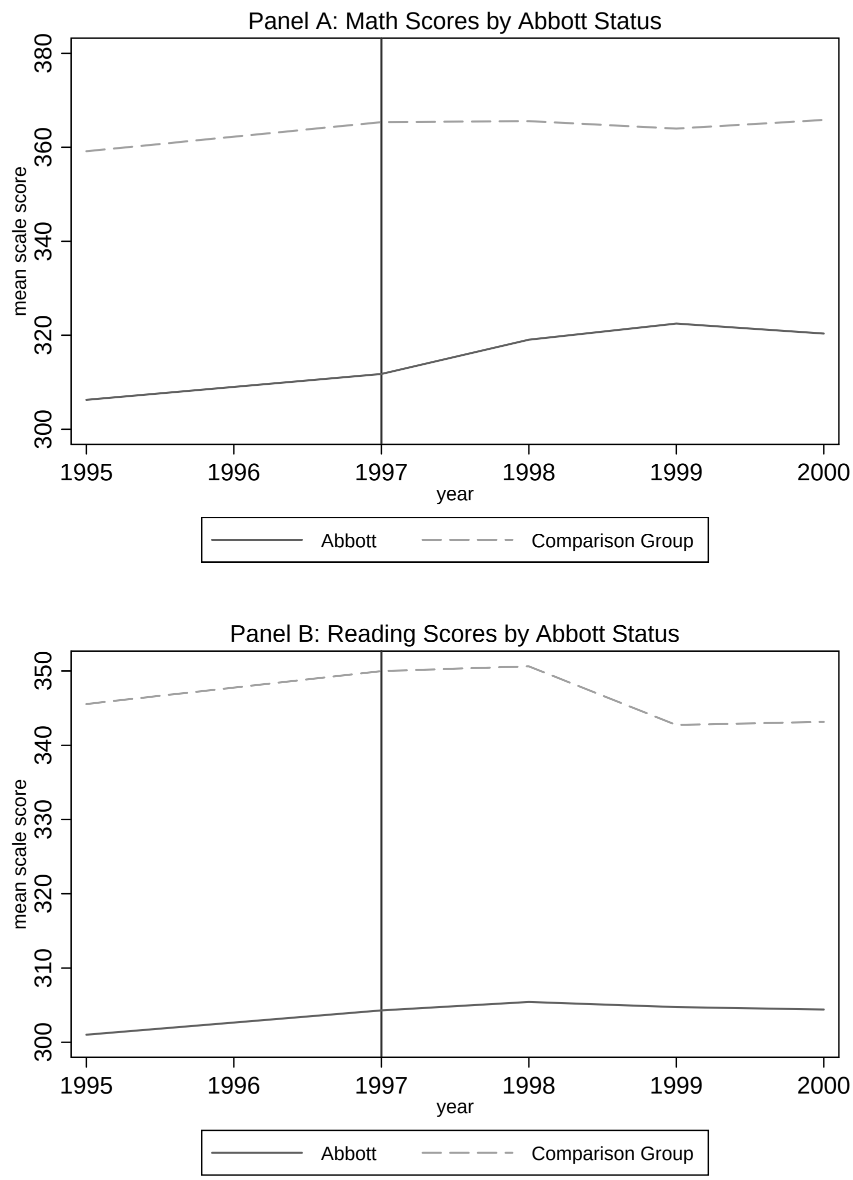


Figure 3.3: Average Math and Reading Scale Scores by Race and Abbott Status

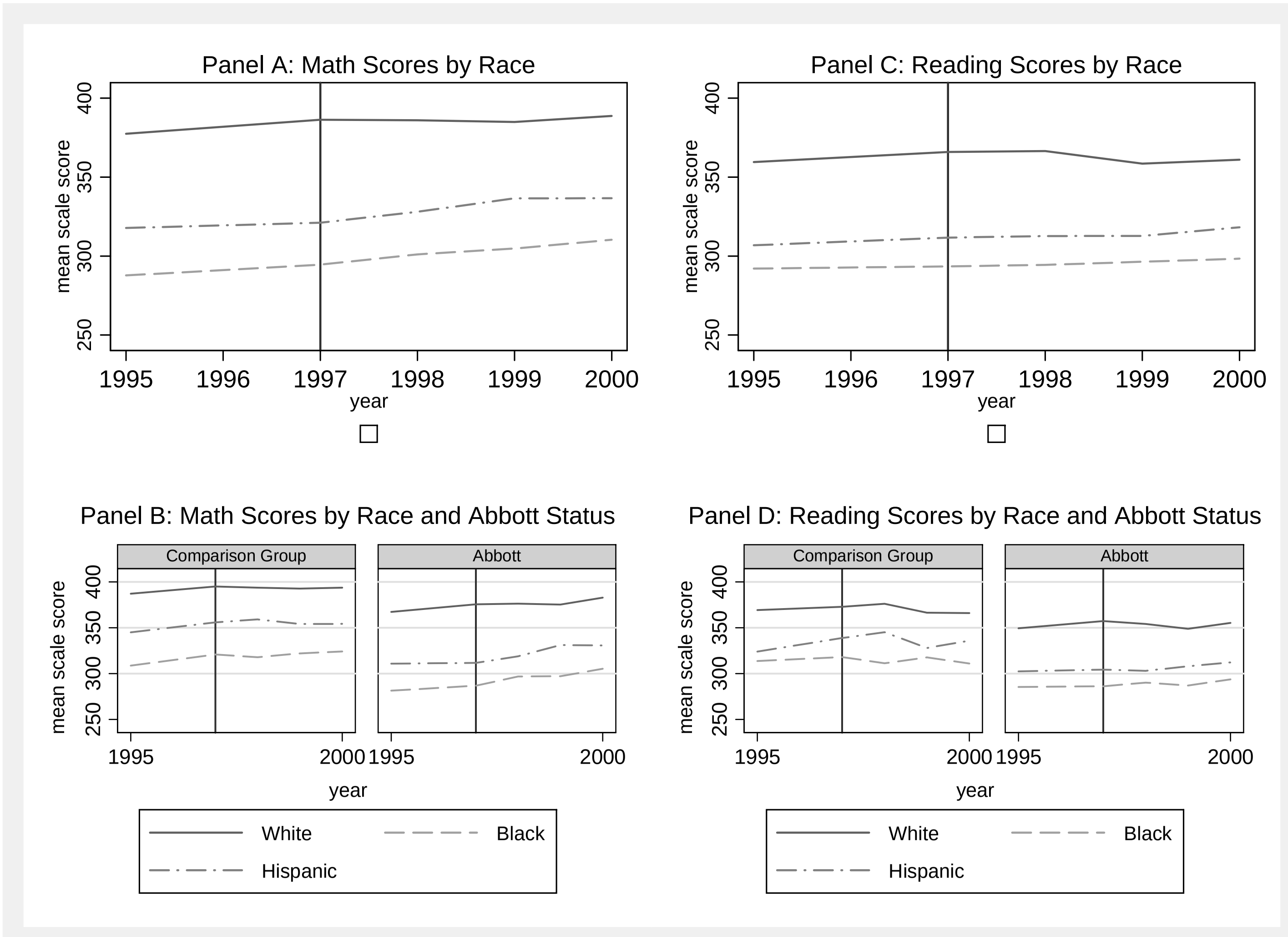


Figure 3.4: Average Proficiency Rates on Fourth Grade Math Assessment by Abbott Status

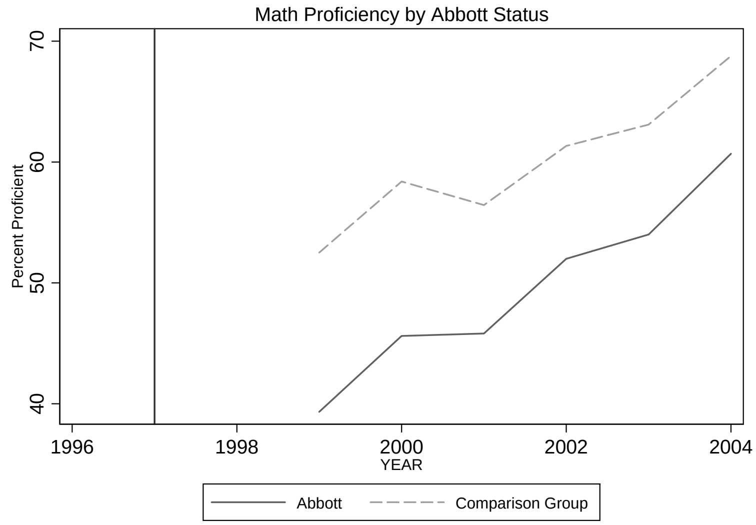


Figure 3.5: Average Proficiency Rates on Fourth Grade Reading Assessment by Abbott Status

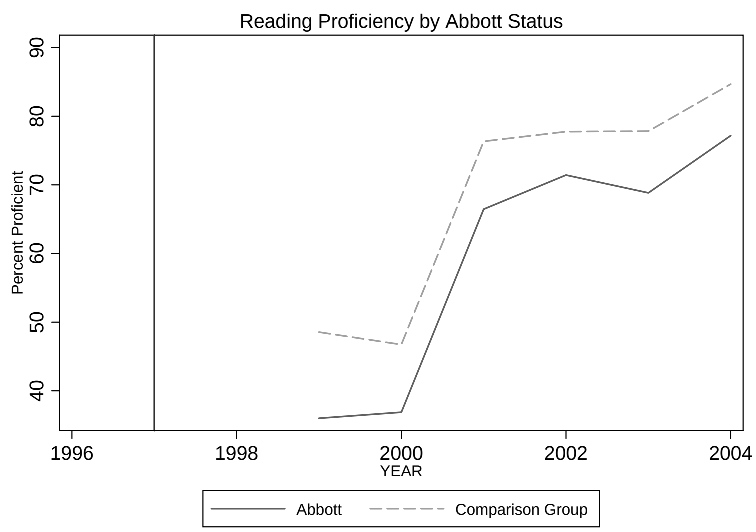


Figure 3.6: Average Proficiency Rates on Eighth Grade Math Assessment by Abbott Status

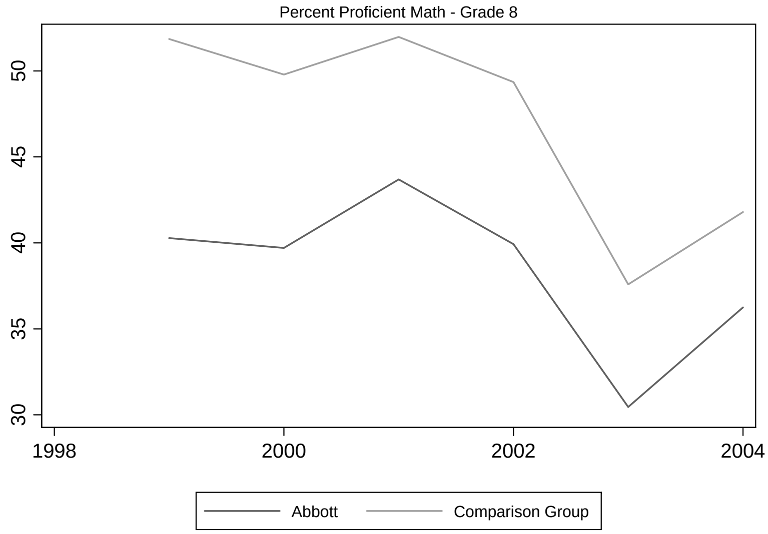
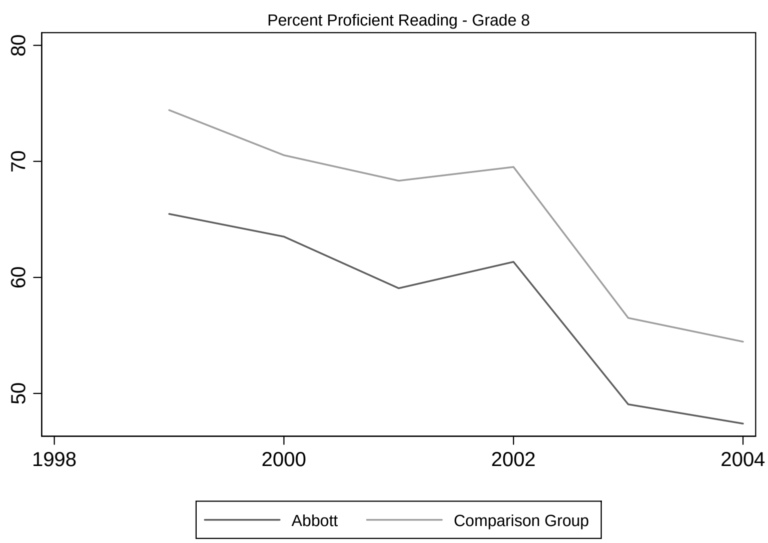


Figure 3.7: Average Proficiency Rates on Eighth Grade Reading Assessment by Abbott Status



3.7 Tables

Table 3.1: Effects of Abbott Policy on Education Expenditures

Instrumental Variable Specifications - Y=Expenditures per pupil

Model: Exp per pupil = α + βState Aid per pupil (IV) + γDistrict FE + δYear FE + ε

Entry in table is coefficient on State Funding for the Abbott*Post97 Indicator used as an IV.

		Estimate	Std Error
(1)	**Total Expenditures**	0.697	(0.110)**
	Expenditure Subcategories		
(2)	Current K-12 Expenditures	0.857	(0.097)**
(3)	Instructional Expenditures	0.483	(0.064)**
(4)	Support Services Exp	0.365	(0.041)**
(5)	Other Current Expenditures	0.011	(0.012)
(6)	Non-K-12 Expenditures	-0.004	(0.010)
(7)	Capital Outlays	-0.128	(0.076)
	Subcategories of Instructional or Support Services Expenditures		
(8)	Instructional Salaries	0.207	(0.042)**
(9)	Instructional Support	0.053	(0.009)**
(10)	Pupil Support	0.065	(0.010)**
(11)	Admin Support	0.006	(0.004)
(12)	Operations and Maintenance Support	0.025	(0.011)*
(13)	Transportation Support	-0.003	(0.006)
	Observations	676	

Robust standard errors in parentheses (clustered at the district level).
* significant at 5%; ** significant at 1%

Table 3.2: Means of Analysis Variables for Abbott and Comparison Districts

Means of Analysis Variables in 1997

	Abbott Districts					**Comparison Districts**				
District Level Community Characteristics	N	mean	sd	min	max	N	mean	sd	min	max
District Income per pupil (1,000$)	29	77.8	64.3	21.6	397.1	23	91.7	35.2	36.0	185.9
Equalized Housing Value (10,000$)	29	20.1	12.5	5.3	73.7	23	34.2	18.7	10.3	103.2
Effective Tax Rate	29	3.45	0.78	2.46	5.28	23	2.96	0.58	2.04	4.45
District Level School Characteristics										
Total Dist Enrollment	29	935.1	983.3	149.9	4419.2	23	338.9	182.4	101.5	704.3
11th Grade Enrollment	29	429.3	389.4	95.0	1953.0	23	228.0	118.8	62.0	516.0
12th Grade Enrollment	29	383.9	360.9	95.0	1770.0	23	206.0	107.8	65.0	440.0
Diplomas Granted	29	357.7	322.1	82.0	1511.0	23	199.6	102.8	70.0	426.0
Dropout Rate	29	1.9	0.9	0.4	4.8	23	1.9	1.4	0.0	6.6
Cumulative Promotion Index	29	65.0	22.7	31.2	123.1	23	74.2	11.7	47.8	93.5
Attendance Rate	29	92.3	1.2	89.9	95.0	23	93.5	1.0	91.3	94.9
Student Mobility Rate	29	4.2	1.5	1.5	8.5	23	3.7	1.4	1.5	7.5
Special Ed Rate	29	4.8	2.2	1.0	9.3	23	3.6	2.1	0.0	8.7
Pct LEP	29	9.9	9.0	0.4	34.0	23	5.1	4.8	0.0	16.6
Class Size	29	22.2	2.3	15.8	27.2	23	22.9	2.0	19.1	26.6
Student-Faculty Ratio	29	13.7	1.5	9.9	16.4	23	14.6	1.5	11.8	17.8
Student-Admin Ratio	29	308.3	64.0	203.2	526.0	23	305.6	65.5	200.7	436.1
District Level Testing Averages										
Math Scale Score	29	333.0	34.1	278.5	390.6	23	371.1	25.8	308.2	407.2
Math Proficiency Rate	29	64.8	16.7	38.7	92.5	23	81.1	11.7	50.3	96.6
Number Taking Exam- Math	29	333.2	316.8	67.0	1571.0	23	184.9	93.0	55.0	396.0
Reading Scale Score	29	322.7	25.8	281.0	371.0	23	355.2	19.6	312.8	389.9
Reading Proficiency Rate	29	62.1	13.4	40.0	86.6	23	79.0	10.3	57.2	95.4
Number Taking Exam- Reading	29	333.3	315.9	69.0	1567.0	23	185.7	93.3	55.0	399.0
Ave SAT Score - Math	29	433.4	37.7	370.0	528.0	23	472.0	25.5	410.0	539.0
Ave SAT Score - Verbal	29	421.0	40.8	362.0	522.0	23	463.4	25.9	410.0	515.0
25th Pctile Math SAT	29	374.8	34.3	310.0	470.0	23	406.1	26.9	350.0	470.0
75th Pctile Math SAT	29	490.0	42.1	420.0	580.0	23	536.3	30.0	460.0	605.0
25th Pctile Verbal SAT	29	358.0	41.9	298.0	460.0	23	397.0	30.0	350.0	460.0
75th Pctile Verbal SAT	29	483.9	43.9	420.0	580.0	23	528.9	25.2	480.0	580.0
Percent Taking SAT	29	53.7	11.2	27.0	71.0	23	62.6	9.3	47.0	83.0
Student Level Test Scores and Demographics										
Math Scale Score	9665	319.7	87.3	100	500	4272	369.0	77.7	100	500
Reading Scale Score	9665	312.7	78.0	100	500	4272	353.4	69.5	116	500
Title 1 Math	9665	0.09	0.29	0	1	4272	0.01	0.09	0	1
Title 1 Reading	9665	0.09	0.29	0	1	4272	0.01	0.11	0	1
Female	9665	0.55	0.50	0	1	4272	0.51	0.50	0	1
Black	9665	0.30	0.46	0	1	4272	0.20	0.40	0	1
Hispanic	9665	0.26	0.44	0	1	4272	0.16	0.37	0	1
Female*Black	9665	0.17	0.37	0	1	4272	0.10	0.31	0	1
Female*Hispanic	9665	0.14	0.35	0	1	4272	0.08	0.28	0	1

Table 3.3: Effect of Abbott Policy on 11th Grade Math Assessment: District-level Data

Effect of Policy on Math Scale Scores
Y= Math Scale Score (District Level)

	Panel A: Basic Specifications				Panel B: Add Controls			Panel C: Years Restricted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	Include District Fixed Effects			Include District Fixed Effects			Include District Fixed Effects		
Policy Dummy	-40.30	5.85	2.79		4.81	3.12		7.85	5.42	
	(8.77)**	(3.48)+	(3.50)		(3.45)	(3.53)		(3.18)*	(3.44)	
Post 1997 Dummy	40.18	7.10	7.10		5.04	5.05		1.61	1.65	
	(8.42)**	(1.85)**	(1.85)**		(1.93)*	(1.94)*		(1.65)	(1.67)	
Years getting Abbott $			1.53			0.86			1.25	
			(2.06)			(2.13)			(2.18)	
Abbott and Year=1998				2.91			2.29			5.08
				(2.84)			(2.85)			(2.77)+
Abbott and Year=1999				8.74			7.79			10.79
				(4.01)*			(3.89)+			(3.66)**
Abbott and Year=2000				5.88			4.40			7.85
				(5.39)			(5.55)			(5.37)
Year = 1998				7.98			6.10			2.65
				(2.01)**			(2.08)**			(2.01)
Year = 1999				5.30			3.39			-0.02
				(2.30)*			(2.27)			(2.01)
Year = 2000				8.07			5.71			2.36
				(2.57)**			(2.81)*			(2.55)
Total Dist Enrollment					0.11	0.11	0.11	0.08	0.08	0.08
(divided by 10)					(0.04)*	(0.04)*	(0.04)*	(0.05)+	(0.05)	(0.05)
Enrollment Sq					-0.16	-0.15	-0.15	-0.12	-0.11	-0.11
(divided by 100,000)					(0.05)**	(0.06)**	(0.06)**	(0.06)+	(0.07)	(0.07)
Constant	329.23	329.86	329.86	329.86	238.99	239.52	239.55	261.88	264.20	264.21
	(7.76)**	(0.80)**	(0.81)**	(0.81)**	(39.07)**	(39.76)**	(39.94)**	(42.77)**	(45.54)**	(46.02)**
Observations	416	416	416	416	416	416	416	260	260	260
R-squared	0.12	0.93	0.93	0.93	0.94	0.94	0.94	0.95	0.95	0.95

Standard errors in parentheses are clustered at the district level
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 3.4: Effect of Abbott Policy on 11th Grade Assessments: District-level Data

Effect of Policy on Math Proficiency and Reading Scores and Proficiency
All Columns include District Fixed Effects

	Panel A: Y=Math Proficiency				Panel B: Y=Reading Scale Score				Panel C: Y=Reading Proficiency			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Policy Dummy	3.61	4.54			1.45	3.12			1.45	2.02		
	(1.63)*	(1.60)**			(2.38)	(2.43)			(1.21)	(1.18)+		
Post 1997 Dummy	1.35	0.24			-3.04	-2.80			-2.31	-2.37		
	(0.94)	(0.83)			(1.67)+	(1.78)			(0.84)**	(0.88)**		
Abbott and Year=1998			2.61	3.39			-3.74	-1.99			0.27	0.78
			(1.34)+	(1.42)*			(2.22)+	(2.56)			(1.08)	(1.26)
Abbott and Year=1999			4.70	5.60			3.22	4.89			1.59	2.11
			(1.82)*	(1.77)**			(2.24)	(2.28)*			(1.25)	(1.20)+
Abbott and Year=2000			3.57	4.72			4.71	6.27			2.45	3.08
			(2.64)	(2.66)+			(4.41)	(4.25)			(2.18)	(2.07)
Year = 1998			1.39	0.28			2.44	2.69			-1.02	-1.07
			(0.88)	(0.96)			(1.75)	(2.00)			(0.82)	(0.98)
Year = 1999			1.50	0.39			-5.35	-5.07			-2.17	-2.21
			(1.07)	(0.93)			(1.68)**	(1.78)**			(0.96)*	(0.97)*
Year = 2000			1.16	0.09			-6.08	-5.97			-3.76	-3.89
			(1.54)	(1.42)			(2.49)*	(2.49)*			(1.29)**	(1.27)**
Total Dist Enrollment	0.04	0.03	0.04	0.03	0.04	0.06	0.04	0.07	0.02	0.04	0.02	0.04
(divided by 10)	(0.02)*	(0.02)	(0.02)*	(0.02)	(0.03)	(0.04)+	(0.03)	(0.04)	(0.01)	(0.02)+	(0.01)	(0.02)+
Enrollment Sq	-0.06	-0.04	-0.06	-0.04	-0.05	-0.09	-0.05	-0.09	-0.03	-0.05	-0.03	-0.05
(divided by 100,000)	(0.02)*	(0.03)	(0.03)*	(0.03)	(0.03)	(0.05)+	(0.04)	(0.05)	(0.02)	(0.02)*	(0.02)	(0.03)+
Constant	28.15	37.97	28.21	38.57	287.73	270.86	286.89	268.41	48.88	31.18	48.28	29.61
	(17.37)	(19.81)+	(17.70)	(21.23)+	(24.21)**	(32.57)**	(25.17)**	(34.62)**	(12.40)**	(16.38)+	(12.72)**	(17.22)+
Observations	416	260	416	260	416	260	416	260	416	260	416	260
R-squared	0.93	0.94	0.94	0.94	0.94	0.95	0.94	0.95	0.94	0.95	0.94	0.95

Standard errors in parentheses are clustered at the district level
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 3.5: Effect of Abbott Policy on 11th Grade Assessments: Student-level Data

Effect of Policy on Individual Level Scale Scores
All Columns include District Fixed Effects

	Panel A: Math				Panel B: Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Policy Dummy	5.78	-5.33			1.58	-3.56		
	(3.54)	(4.83)			(2.62)	(3.76)		
Post 1997 Dummy	2.89	2.98			-1.79	-1.73		
	(1.63)+	(1.66)+			(1.62)	(1.60)		
Abbott and Year=1998			7.18	-4.17			-0.37	-5.31
			(3.14)*	(4.38)			(2.75)	(4.05)
Abbott and Year=1999			6.73	-4.42			2.00	-2.98
			(3.39)+	(4.20)			(2.27)	(3.30)
Abbott and Year=2000			3.36	-7.09			2.87	-1.85
			(5.91)	(6.74)			(4.98)	(5.39)
Year = 1997			2.16	1.73			1.35	1.08
			(2.94)	(3.37)			(1.99)	(2.23)
Year = 1998			3.67	3.24			3.57	3.32
			(2.53)	(2.73)			(2.38)	(2.50)
Year = 1999			3.68	3.83			-2.39	-2.30
			(2.36)	(2.49)			(2.00)	(2.05)
Year = 2000			4.62	4.49			-4.39	-4.44
			(2.97)	(3.14)			(2.64)	(2.65)
Title 1	-52.01	-53.85	-52.52	-54.00	-44.20	-42.77	-45.08	-43.11
	(16.23)**	(14.22)**	(16.01)**	(13.99)**	(12.61)**	(11.05)**	(12.34)**	(10.91)**
Female	-9.36	-9.46	-9.37	-9.46	3.62	4.70	3.61	4.69
	(1.03)**	(1.20)**	(1.02)**	(1.20)**	(1.00)**	(1.05)**	(1.00)**	(1.06)**
Black	-35.96	-43.36	-35.96	-43.30	-23.64	-28.40	-23.64	-28.24
	(8.15)**	(6.59)**	(8.18)**	(6.62)**	(6.11)**	(5.07)**	(6.14)**	(5.11)**
Hispanic	-20.89	-27.99	-20.97	-27.96	-17.00	-22.35	-17.05	-22.29
	(5.62)**	(4.68)**	(5.66)**	(4.70)**	(4.69)**	(4.04)**	(4.76)**	(4.06)**
Female*Black	8.22	7.42	8.26	7.45	5.60	5.39	5.67	5.47
	(1.75)**	(2.41)**	(1.74)**	(2.39)**	(1.13)**	(1.62)**	(1.12)**	(1.61)**
Female*Hispanic	-2.96	-2.95	-2.92	-2.94	-4.14	-3.22	-4.10	-3.17
	(1.34)*	(2.14)	(1.34)*	(2.13)	(1.12)**	(1.94)	(1.12)**	(1.95)
Title 1 * Treatment		6.19		5.81		-2.29		-3.40
		(8.70)		(8.69)		(6.98)		(6.69)
Female *Treatment		0.51		0.49		-2.73		-2.71
		(2.00)		(1.99)		(2.29)		(2.27)
Black*Treatment		18.69		18.59		11.86		11.44
		(5.68)**		(5.63)**		(4.67)*		(4.71)*
Hispanic*Treatment		17.07		16.90		12.57		12.31
		(5.07)**		(4.97)**		(4.53)**		(4.55)**
Female*Black*Treatment		0.95		0.95		0.38		0.37
		(3.22)		(3.18)		(2.54)		(2.51)
Female*Hispanic*Treatment		-0.33		-0.30		-1.82		-1.86
		(3.72)		(3.70)		(3.03)		(2.99)
Constant	355.87	359.97	354.82	359.08	335.45	337.32	334.83	336.73
	(4.35)**	(3.88)**	(4.62)**	(4.01)**	(3.39)**	(3.04)**	(3.50)**	(3.13)**
Observations	70352	70352	70352	70352	70325	70325	70325	70325
R-squared	0.21	0.21	0.21	0.21	0.16	0.17	0.17	0.17

Standard errors in parentheses are clustered at the district level
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 3.6: Effect of Abbott Policy on SAT Participation and Scores

Effects on SAT Participation and Performance
All Columns include District Fixed Effects

	Percent Taking SAT		Math Score		75th Percentile Math Score		25th Percentile Math Score		Verbal Score		75th Percentile Verbal Score		25th Percentile Verbal Score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Policy Dummy	-0.48		-4.25		0.24		-7.33		-0.92		-0.67		-2.45	
	(1.77)		(2.93)		(3.85)		(3.11)*		(3.03)		(3.92)		(3.23)	
Post 1997 Dummy	2.63		0.55		-0.27		3.89		-3.11		-2.57		2.79	
	(1.50)+		(2.52)		(3.34)		(2.67)		(2.58)		(3.43)		(2.27)	
Abbott and Year=1998		1.72		-5.81		-5.74		-12.59		-4.96		-4.20		-5.43
		(1.86)		(3.33)+		(5.24)		(3.76)**		(3.65)		(4.28)		(4.17)
Abbott and Year=1999		1.23		-8.09		-0.59		-10.03		-0.70		1.86		-5.14
		(2.73)		(4.31)+		(4.63)		(5.66)+		(3.63)		(4.94)		(3.75)
Abbott and Year=2000		-4.39		1.21		7.09		0.66		2.96		0.32		3.29
		(3.17)		(4.82)		(5.87)		(3.94)		(5.78)		(6.67)		(6.95)
Year = 1998		1.10		1.25		2.56		9.02		-0.79		-0.33		4.60
		(1.45)		(2.99)		(4.83)		(3.18)**		(3.29)		(3.55)		(3.95)
Year = 1999		2.77		1.60		-1.16		3.95		-2.87		-3.55		4.13
		(2.56)		(3.69)		(4.00)		(4.81)		(3.26)		(4.49)		(3.03)
Year = 2000		4.02		-1.25		-2.18		-1.33		-5.69		-3.80		-0.42
		(2.16)+		(3.17)		(4.13)		(2.95)		(3.49)		(5.02)		(3.24)
Constant	57.12	57.12	439.69	439.69	493.12	493.11	377.94	377.94	425.59	425.59	485.93	485.93	358.13	358.13
	(0.40)**	(0.40)**	(0.65)**	(0.66)**	(0.84)**	(0.85)**	(0.69)**	(0.69)**	(0.68)**	(0.68)**	(0.84)**	(0.85)**	(0.87)**	(0.88)**
Observations	312	312	312	312	312	312	312	312	312	312	312	312	312	312
R-squared	0.78	0.79	0.94	0.95	0.93	0.94	0.91	0.92	0.95	0.95	0.95	0.95	0.93	0.93

Standard errors in parentheses are clustered at the district level
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 3.7: Effect of Abbott Policy on Enrollment and Retention

Effect on Alternate District-Level Outcomes
All Columns include District Fixed Effects

	11th Grade Enrollment		G11 Enrollment as Fraction of Lagged G9 Enrollment		Cumulative Promotion Index		Attendance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Policy Dummy	-46.52		0.80		8.84		0.31	
	(26.11)+		(3.31)		(2.83)**		(0.15)*	
Post 1997 Dummy	13.12		-6.11		-8.08		0.48	
	(7.10)+		(2.64)*		(1.92)**		(0.10)**	
Abbott and Year=1998		-37.71		-0.44		4.82		0.52
		(30.05)		(3.56)		(3.81)		(0.15)**
Abbott and Year=1999		-54.94		0.16		7.56		0.11
		(37.15)		(3.95)		(4.84)		(0.25)
Abbott and Year=2000		-46.80		2.64		14.07		0.30
		(24.29)+		(4.51)		(5.75)*		(0.32)
Year = 1998		7.53		-4.02		-4.76		0.44
		(8.20)		(3.01)		(2.96)		(0.08)**
Year = 1999		19.12		-6.59		-11.26		0.46
		(10.89)+		(2.98)*		(3.85)**		(0.19)*
Year = 2000		12.52		-7.70		-8.12		0.54
		(6.54)+		(3.45)*		(4.29)+		(0.14)**
Constant	655.50	655.50	75.71	75.71	64.40	64.40	92.00	92.00
	(6.62)**	(6.65)**	(0.60)**	(0.61)**	(0.59)**	(0.58)**	(0.04)**	(0.04)**
Observations	416	416	416	416	416	416	312	312
R-squared	0.99	0.99	0.74	0.74	0.64	0.65	0.90	0.90

Standard errors in parentheses are clustered at the district level
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 3.8: Effect of Abbott Policy on Mediating Variables at the District-level

Effects on District-level Mediating Variables
All Columns include District Fixed Effects

	Class Size		Student-faculty Ratio		Student- Admin Ratio		Student Mobility Rate		Special Ed Rate		Pct LEP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Policy Dummy	-0.12		-0.44		-17.26		0.12		0.28		0.02	
	(0.49)		(0.21)*		(8.62)+		(0.23)		(0.31)		(0.07)	
Post 1997 Dummy	-0.50		-0.55		23.34		-0.12		0.32		-0.03	
	(0.46)		(0.16)**		(6.59)**		(0.12)		(0.25)		(0.04)	
Abbott and Year=1998		0.12		-0.28		-8.04		-0.08		0.35		0.12
		(0.50)		(0.22)		(11.51)		(0.17)		(0.29)		(0.12)
Abbott and Year=1999		-0.16		-0.30		-19.17		0.14		0.30		-0.02
		(0.57)		(0.27)		(13.62)		(0.28)		(0.45)		(0.09)
Abbott and Year=2000		-0.31		-0.74		-24.78		0.30		0.20		-0.04
		(0.74)		(0.36)*		(16.27)		(0.48)		(0.47)		(0.11)
Year = 1998		-0.32		-0.11		21.57		-0.16		0.02		-0.09
		(0.47)		(0.16)		(10.02)*		(0.12)		(0.22)		(0.11)
Year = 1999		-0.18		-0.65		22.06		-0.06		0.52		-0.02
		(0.51)		(0.22)**		(9.59)*		(0.19)		(0.36)		(0.04)
Year = 2000		-1.01		-0.89		26.45		-0.14		0.41		0.03
		(0.67)		(0.28)**		(13.66)+		(0.15)		(0.35)		(0.05)
Constant	22.86	22.86	14.22	14.22	298.80	298.80	4.11	4.11	4.15	4.15	9.24	9.24
	(0.09)**	(0.09)**	(0.06)**	(0.06)**	(2.18)**	(2.20)**	(0.07)**	(0.07)**	(0.08)**	(0.08)**	(0.02)**	(0.02)**
Observations	312	312	312	312	312	312	312	312	312	312	312	312
R-squared	0.74	0.76	0.81	0.86	0.81	0.81	0.82	0.82	0.85	0.85	1.00	1.00

Standard errors in parentheses are clustered at the district level
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 3.9: Effect of Abbott Policy on NAEP Scores

Average NAEP Scale Scores for Central Cities and Urban Fringe in New Jersey

	4th Grade Reading			4th Grade Math			8th Grade Math		
Year	**Central city**	**Urban fringe**	**Gap**	**Central city**	**Urban fringe**	**Gap**	**Central city**	**Urban fringe**	**Gap**
1992 [1]	194	228	**34**	202	231	**30**	243	277	**34**
	(3.9)	(1.6)		(4.1)	(1.6)		(3.2)	(1.7)	
1994 [1]	190	225	**36**						
	(5.6)	(1.6)							
1996 [1]				213	229	**16**			
				(6.1)	(1.9)				
2003	201	228	**27**	219	241	**21**	255	283	**29**
	(5.5)	(1.3)		(5.6)	(1.0)		(3.7)	(1.3)	
2005	209	224	**15**	233	245	**12**	260	285	**26**
	(3.8)	(1.4)		(4.7)	(1.1)		(8.7)	(1.4)	

Standard deviations are in parentheses.
1 Accommodations were not permitted for this assessment.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 2003 and 2005 Reading Assessments and 1992, 1996, 2003 and 2005 Math Assessments. Downloaded from http://nces.ed.gov/nationsreportcard/naepdata/
NOTE: The NAEP scale ranges from 0 to 500.

3.8 Data Appendix

The data sources and definitions of main variables are described in more detail in this section.

3.8.1 National Center for Education Statistics Data

Enrollment All enrollment figures are for October 15th of the given school year. These data are downloaded from the Common Core of Data website at http://nces.ed.gov/ccd/index.asp.

Cumulative Promotion Index (CPI) The CPI is an alternative measure for a graduation rate. It uses two years of consecutive data for all high school grades to estimate the probability that a student entering 9th grade will graduate in four years with a diploma. This measure was developed by Christopher Swanson at the Urban Institute. For more information on the CPI see http://www.urban.org/url.cfm?ID=410843.

My CPI measure is calculated from annual enrollment data downloaded from NCES. The formula appears below.

CPI = $(10thGrEnroll_t/9thGrEnroll_{t-1})$* $(11thGrEnroll_t/10thGrEnroll_{t-1})$* $(12thGrEnroll_t)/11thGrEnroll_{t-1})$*$(Diplomas_t/12thGrEnroll_{t-1})$

3.8.2 New Jersey Department of Education Data

School Report Card The School Report Cards are published annually by the NJ-DOE and include, among other variables, school-level data on enrollment, staffing, expenditures, and results on student assessments. A longitudinal historical file covers the school years 1994-95 to 2001-02. These files do not, however, include data on all items back to 1995, as many items were only collected in later years. The items used in this paper are defined below.

These data are found at http://education.state.nj.us/rc/.

Average class size is total enrollment divided by the number of classrooms for elementary school and by the number of English classes for secondary schools.

Limited English proficient (LEP) is the number of students in LEP programs divided by total enrollment.

Student mobility rate is the number of students who left or entered the district during the school year (but after the Oct 15 enrollment count) divided by total enrollment.

The SAT data are reported to New Jersey by the College Board. The NJDOE data report card documentation does not include information on how the participation rates are calculated.

Attendance rate is the average percent of students in attendance each day. It is the sum over the school year of all students present each day divided by total possible student-days (total enrollment*number of school days).

Student/Administrator ratio is total fall enrollment divided by the number of administrators in full-time equivalents (FTEs).

Student/Faculty ratio is total fall enrollment divided by the sum of full-time equivalent classroom teachers and educational support personnel assigned to the district in October.

NJ Assessment Data The 11th grade assessment that I use for the primary analysis is the High School Proficiency Test (HSPT); testing using the HSPT began in 1991 and continued through 2000. This was replaced in 2002 by the High School Proficiency Assessment (HSPA).

The state reports the average scale score and fraction of students in each district who pass each state assessment in the annual school report card data. An electronic version of these data is available starting with the 1997-98 school year. The figure reported in the report card historical files is the cohort pass rate, which incorporates multiple administrations of the test for students who fail the first time. Because I cannot calculate this number from the cross-sectional microdata I use data from only the fall administration of the test for first-time test takers.

I extend the 11th grade assessment series back in time using paper copies of the results published by the state in earlier years. I have these district-level scores for the tests given in 1993 to 2000, which gives me 5 pre-reform data points and 3 post-reform years.

I also obtained restricted student-level results for this exam from the New Jersey Department of Education for 1995 and 1997-2000. I use the same election criteria for my analysis that the state uses for reporting the aggregated scores. Only first-time 11th grade test takers are included; special education and limited English proficient students are excluded.

3.8.3 National Assessment of Educational Progress (NAEP) data

The NAEP data used in this paper were downloaded using the NAEP Data Explorer at
http://nces.ed.gov/nationsreportcard/naepdata/.

CHAPTER IV

On the Optimal Allocation of Students and Resources in a System of Higher Education

4.1 Introduction

Approximately 77% of college students in the United States attend public institutions, where total annual expenditures now exceed $190 billion (NCES, 2005a,b). Despite the magnitude of public involvement in higher education, and despite the enormous body of research on the economics of education, economists have not established a normative model of how students and resources should be allocated in a system of public higher education. The aim of this paper is to provide a simple and tractable framework for the analysis of this issue and to derive several intuitive results regarding the optimal allocation of students and resources.

Given a distribution of student ability and a limited pool of resources for higher education, we model the social planner's decision to establish schools and populate them with students and resources. Our model is driven by two simple assumptions: (1) that there is complementarity between resources and student ability in producing educational outcomes; and (2) that there is a fixed cost to establishing a school. We show that these assumptions produce a tiered structure that sorts students by skill and results in discontinuous spending and educational output per student for essentially identical students at the margin between schools.

The existence of a fixed cost creates economies of scale, both for individual schools and for the whole system. Because they can tailor educational spending more closely to student quality, university systems serving a larger population produce more output per student, holding constant total resources per student. Improved tailoring raises aggregate social welfare, but it does not benefit all students (i.e., it is not Pareto improving). In particular, the lowest ability students at each school will lose when an additional school is introduced into the system, because they will be dropped into a lower tier. Larger systems with more schools will provide a college education to a larger fraction of the population, and they will feature a wealthier and more selective flagship school, at the optimum.

The principal contribution of this paper is that these results, which are broadly consistent with observed stylized facts, can be derived from a very simple model. Most states have a hierarchy of postsecondary institutions exhibiting markedly different levels of resources. The most obvious example is California, where the state's Master Plan for Higher Education clearly lays out a three tier structure comprised of the University of California system, the California State University system, and the California Community Colleges. Even without an explicit plan, most states have a flagship public university, some number of other four year institutions, and a system of local community colleges. Students are distributed among these schools largely on the basis of their measured academic ability, and discontinuous levels of public spending per student in each tier are strongly and positively associated with average ability. On average across the country, instructional expenditure per student in public universities that grant doctorates is more than twice that in community colleges. The difference in total expenditures directly relevant to education is higher.[1]

[1]Based on authors' calculations of 2004 data from the Integrated Postsecondary Education Data System.

Further, Winston (1999) shows that spending per student and subsidy per student are generally increasing in student quality in the U.S across all universities.

Our paper is closely related to existing work on systems of educational provision at the elementary and secondary level, as well as earlier work on competitive (as distinct from planned) systems of higher education. All of the earlier work that we are aware of contains explicit consideration of peer effects. Arnott and Rowse (1987) study an elementary and secondary system from a planning perspective similar to ours. They consider the allocation of students to various classes within a school, where students vary in ability and classrooms have a level of resources per student that applies to all students in the class. Their principal finding is that any type of partition is possible, depending on the strength of peer effects.[2]

Peer effects are a central feature of other related work. Rothschild and White (1995) analyze competitive outcomes in higher education with peer effects and demonstrate the potential for efficient private provision. Epple and Romano (1998) construct a model of private and public secondary schools in order to analyze the effects of voucher reforms. Epple, Romano, and Sieg (2003) and Epple, Romano, and Sieg (2006) consider a model of higher education in which universities compete on quality and university differentiation is driven by exogenous endowment differences.

The driving force in our model is complementarity in production. The notion that complementarity leads to positive assortative matching (tiers, in this case) is hardly new. This is the underlying mechanism in the marriage market model of Becker (1973), for example. In the education literature, Arnott and Rowse (1987), Bénabou (1996), Epple and Romano (1998), Epple et al. (2003) and Epple et al. (2006) all

[2]Effinger and Polborn (1999) work with a model that, on the surface, appears similar to ours. They begin, however, by assuming that there are two different schools and that some students are innately better served at the "lower" school. They then solve an allocation problem, under the assumption that attendance at one school versus the other affects wages in a market with imperfect information.

derive at least some results that resemble our tiered structure.[3] In addition to assuming complementarity between resources and ability, this earlier work considers peer effects, and, in most cases, outcomes are influenced by the distribution of income.

Our model can be interpreted as a simplified case of much of this earlier work. As is often the case, simplification yields both benefits and costs. We show that it takes only two strong (but not unreasonable) assumptions to generate an optimal system that is broadly consistent with the stylized facts of state higher education systems. By eliminating peer effects, we demonstrate that complementarity, along with fixed costs, is sufficient to make a tiered system optimal. Our simple model also allows us to consider some issues that do not appear in the prior literature. In particular, our results on resource discontinuities, our analysis of the optimal number of universities, and our discussion of the selectivity of schools are new, even as they are latent in earlier work.

We cannot, however, claim to have captured everything that might be important in our simple model. As indicated by previous research, markets, peer effects and associations between income and ability are important in higher education systems. Transportation costs (and therefore spatial considerations) and the political economy of education finance are also surely influential. Abstracting from these concerns enables us to isolate the role of complementarity and to build intuition, but it also eliminates consideration of the ways in which these factors may reinforce or counteract our findings.

We also believe that our emphasis on a planner's perspective is of value. In Epple and Romano (1998), Epple et al. (2003) and Epple et al. (2006), schools compete with each other by maximizing quality. While this is a reasonable approach, we think

[3]A strand of the literature that studies continuous optimization problems has touched on the implications of complementarity between student ability and student quality. Fernandéz and Galí (1999) is an example. That paper differs significantly from our analysis by considering a continuum of pre-existing schools of exogenous quality.

that for the case of public universities in state systems, it is more natural to consider the planner's problem of maximizing total output across a set of schools. While both approaches lead to similar mathematical results on sorting, we think that framing the problem from a planner's perspective is of heuristic value.

We build up our model incrementally throughout the remainder of the paper. In section 2, we introduce the key elements of the model. Section 3 describes the solution when the optimal number of schools is fixed. Section 4 extends the analysis to consider the optimal number of schools. Section 5 concludes. All proofs are relegated to the appendix.

4.2 A Description of the Model

We model a social planner's problem. The planner takes as given the distribution of students, the amount of resources available, the education production function, and the fixed cost of establishing a school. The planner chooses the number of universities, selects which students attend each university, and decides how many resources to give each university.

We model educational output as a function of student ability and resources per student at the student's school. We assume that ability and resources are complementary. Our assumptions about the distribution of students will be innocuous, but our assumptions about the curvature of the education production function are key to our results.

We believe a planner's problem is attractive both because it is relatively simple and because it is a good approximation to the real world, where the vast majority of students attend public institutions. In our model, students are not explicitly decision-makers. If they were, however, they would all have unanimous preferences to attend schools with higher resources per student, because this is the only dimension along which universities vary. The planner can therefore use selective admissions to produce the desired allocation. In other words, a planner with control over admissions policy can satisfy all incentive constraints.

We consider a utilitarian social welfare function, so that the social planner seeks to maximize the aggregate level of educational output. Distributional considerations could be modeled by giving the social planner a preference for equality of outcomes (e.g., a concave social welfare function), equality of expenditure (e.g., a loss function for school quality disparities), or equality of access (e.g., a loss function for selective admissions). Courant, McPherson, and Resch (2006) argue that distributional considerations may affect the design of higher education because higher education may be a useful instrument for smoothing preexisting differences in welfare. We acknowledge that distributional concerns are interesting, but we focus on the utilitarian case to maintain simplicity and because we believe it is a good characterization of higher education (as opposed to primary and secondary education), where selective admissions prevail and there is typically no presumption of education for all.[4]

We do not consider peer effects. They are not needed to obtain any of our results, and omitting them simplifies the model and makes the mechanics more transparent. Previous research has included both peer effects and expenditures per student as inputs to education (Arnott and Rowse, 1987; Epple and Romano, 1998; Effinger and Polborn, 1999; Epple et al., 2003), clouding the issue of what drives the model. Our results show that complementarity is sufficient for educational sorting.

The education produced by an individual student is denoted by $h(x, r)$, where x

[4]There is a class of concave social welfare functions that we could employ without changing any qualitative results. This leads to limited additional insight, at the cost of significant additional notation. A sufficient condition for our results to hold is that the transformation $U(h(\cdot))$ be supermodular and concave, where $h(\cdot)$ is the education production function and $U(\cdot)$ is the social welfare function. The conditions on $U(\cdot)$ and $h(\cdot)$ that ensure this have no obvious economic interpretation.

is that student's ability and r is the resources per student at the student's school. Ability follows a continuous, differentiable cumulative distribution, $F(x)$, with a probability density function denoted by $f(x)$ and a finite support bounded by $\underline{x}$ and $\overline{x}$.[5]

To establish a school, the planner pays a fixed cost, θ, and then purchases the variable input into education. We assume that all students at a school receive the same resources per student. In effect, educational resources at the school are a congestible public good. For any level of total resources that a school provides, the level of resources per student depends only on the number of students in the school.

We assume that the education production function, $h(x, r)$, is continuous, twice differentiable, and increasing and concave in each argument. Logically, output should be increasing in ability and resources. We also suppose that it is concave in each element. As a normalization, we assume that students with zero resources produce zero output, and we restrict the domain of h to weakly positive values of resources. Finally, we assume that the education output function exhibits complementarity. This may also be called supermodularity, and it is equivalent to a positive cross partial derivative.

$$h = h(x, r)$$
$$h_1 > 0 \qquad h_{11} < 0$$
$$h_2 > 0 \qquad h_{22} < 0$$
$$h_{12} > 0 \quad h(x, 0) = 0$$

Only complementarity should be a controversial assumption. Complementarity means that, at any given level of resources per student, higher ability students produce more when given a marginal increase in resources. While it is not obvious that this is true in all cases, we find it to be a plausible assumption. We note also that it is pervasive

[5]A finite support is not necessary generally, but it will be required when we later assume a uniform distribution.

in the literature (Arnott and Rowse, 1987; Epple et al., 2003).

The planner will choose to set up K universities, indexed by $k = 1, \ldots, K$. The planner must pay θK in fixed costs from the total available resources T. What remains, R, the resources net of the fixed costs, is partitioned among the schools. We denote the proportion of R allocated to school k as ρ_k. The planner must also partition the distribution $F(x)$ between schools. For each value of x, the planner allocates a proportion of the distribution to each school, denoted by $p_k(x)$. The total measure of students is denoted by S. The measure of students at a school is denoted by s_k and is equal to $S \int p_k(x) f(x) dx$. Thus, the resources per student at a school, r_k, may be written as $\frac{\rho_k R}{s_k}$.

The planner simultaneously chooses the number of universities and the partition of students and resources. It is useful, however, to write the planner's problem when the number of schools is fixed as a sub-problem. We denote the global value function as V, and the value function when K is fixed as W:

$$
\begin{aligned}
V(T, S, \theta) &= \max_{K} W(T, S, \theta, K), \text{ where} \\
W(T, S, \theta, K) &= \max_{\{\rho_\mathbf{k}\}, \{\mathbf{p_k(x)}\}} S \int_{\underline{x}}^{\overline{x}} h\left(x, \frac{\rho_1 R}{s_1}\right) p_1(x) f(x) dx + \\
&\quad ... + S \int_{\underline{x}}^{\overline{x}} h\left(x, \frac{\rho_K R}{s_K}\right) p_K(x) f(x) dx \\
\text{s.t.} \quad & \theta K + R \leq T \\
& \sum_{k=1}^{K} \rho_k \leq 1 \qquad \text{(P1)} \\
& \rho_k \geq 0 \; \forall \; k \\
& \sum_{k=1}^{K} p_k(x) \leq 1 \; \forall \; x \\
& p_k(x) \geq 0 \; \forall \; k, x
\end{aligned}
$$

Each integral of program P1 represents a school. The output of a school is the integral of individual student outputs with resources equal to r_k, integrated over $p_k(x)f(x)$, the distribution of students assigned to the school.

The first and second constraints are the planner's budget constraint. The third disallows "negatively funded" schools. The fourth and fifth restrict the planner's partition, disallowing negative assignments, while permitting the planner to not educate some students.

In choosing the optimal number of schools, the planner balances the burden of paying the additional fixed costs for more schools against the inefficiency of sending very different types of students to the same school. The planner allocates students and resources, which implicitly sets the resources per student at each school.

First order conditions for this problem can, at least in principle, be established using variational methods. In the interest of clarity, we shall instead demonstrate that the problem can be reduced to a more tractable form.

4.3 The Optimal Allocation When the Number of Universities is Fixed

We begin by isolating the allocation decision, taking the number of schools as fixed. With a fixed number of universities, the planner's solution is a mapping from the set of students and resources into universities. One class of partitions of the type space involves grouping the highest ability types together in one school, then grouping the next highest ability types in a second school, and so on. We call this a monotonic partition.

Definition 1. *A partition is monotonic if and only if, for least and greatest elements $\underline{x}_k$ and $\overline{x}_k$ in each school, a student x is assigned to school k if and only if $\underline{x}_k \leq x \leq \overline{x}_k$.*[6]

[6]Alternatively, this could be stated as, for least and greatest elements $\underline{x}_k$ and $\overline{x}_k$ in each school, $p_k(x) = 0$ if

Any partition that results in one school having both higher and lower ability students than another school cannot be monotonic. Any partition that puts two students of the same type into different schools cannot be monotonic. Supermodularity (complementarity) of the underlying education production function is a sufficient condition to make the optimal partition monotonic.

Proposition IV.1. *If $h(x, r)$ is complementary (supermodular), then the optimal partition of students is monotonic.*[7]

Supermodularity is sufficient to generate educational sorting, even when there are no peer effects. Imagine, instead, a non-monotonic partition between two schools. The allocation can be improved by replacing a lower ability student with a higher ability student in the school with more resources per student.

Corollary 1. *In any optimal monotonic partition, any school that has higher ability students than another school will also have higher resources per student.*

Resources and ability are complements. This immediately leads to the conclusion that universities with higher ability students should have more resources per student.

Proposition IV.1 tells us the shape of the optimal solution, allowing us to rewrite program P1. The planner sets an admissions policy by determining the lowest ability type admitted to each school.[8] We denote the highest type assigned to school k by

$x < \underline{x}_k$ or $x > \overline{x}_k$ and $p_k(x) = 1$ if $\underline{x}_k < x < \overline{x}_k$.

[7]All proofs are in the appendix.

[8]There will be no gaps between the lowest type in one school and the highest type in the next school; otherwise total output could be increased by giving a higher ability student the place of a student at the lower school.

a_k, with a_0 denoting the lowest type at the lowest school.[9]

$$
\begin{aligned}
V(T,S,\theta) &= \max_K W(T,S,\theta,K), \text{ where} \\
W(T,S,\theta,K) &= \max_{\{\rho_\mathbf{k},\mathbf{a_k}\}} S\int_{a_0}^{a_1} h\left(x,\frac{\rho_1 R}{s_1}\right) f(x)dx + \\
&\quad ... + S\int_{a_{K-1}}^{a_K} h\left(x,\frac{\rho_K R}{s_K}\right) f(x)dx \\
\text{s.t.} \quad & \theta K + R \leq T \qquad \text{(P2)} \\
& 0 \leq a_0 \leq ... \leq a_k \leq a_{k+1} \leq ... \leq a_K \leq 1 \\
& \sum_{k=1}^{K} \rho_k \leq 1 \\
& \rho_k \geq 0 \ \forall \ k
\end{aligned}
$$

Program P2 has one fewer constraint than program P1, and the suboptimization problem for W is a standard static optimization problem. One can easily construct a Lagrangean and characterize the first-order conditions for any given K. In principle, the planner can find the optimal allocation for each value of K that is feasible, then choose the best among these.

We find significant heuristic value in further simplifying the problem. First, we assume that the distribution of student ability is uniform on $[0,1]$. This simplifies notation, but does not substantively affect any interpretations. Second, we normalize S to 1. Third, for the remainder of this section only, we assume that the number of schools is fixed at two. Again, this substantially clarifies the tension in the model, and all of the following results are easily translatable to other values of K.

Under these additional assumptions, the planner chooses one value of ρ, the proportion of resources to be allocated to the lower school, and two cut-off conditions, the lowest ability type admitted to the lower school, a, and the lowest ability type

[9] I.e., $p_k(x) = 1$ for $x \in [a_{k-1}, a_k]$ and $p_k(x) = 0$ otherwise.

admitted to the higher school, b.

$$\max_{a,b,\rho} \; H(a,b,\rho) = \int_a^b h\left(x, \frac{\rho R}{b-a}\right) dx + \int_b^1 h\left(x, \frac{(1-\rho)R}{1-b}\right) dx$$

$$\text{s.t.} \quad 0 \le a \le b \le 1 \tag{P3}$$

$$0 \le \rho \le 1$$

First-order necessary conditions for an interior solution follow from the unconstrained optimization problem. At an interior optimum, the Lagrange multipliers on the inequality constraints are all zero. Only one constraint, $a = 0$, can ever bind at the optimum.[10]

$$H_\rho = \frac{R}{b-a} \int_a^b h_2\left(x, \frac{\rho R}{b-a}\right) dx - \frac{R}{1-b} \int_b^1 h_2\left(x, \frac{(1-\rho)R}{1-b}\right) dx = 0 \tag{4.1}$$

$$H_a = -h\left(a, \frac{\rho R}{b-a}\right) + \frac{\rho R}{(b-a)^2} \int_a^b h_2\left(x, \frac{\rho R}{b-a}\right) dx = 0 \tag{4.2}$$

$$\begin{aligned} H_b = & h\left(b, \frac{\rho R}{b-a}\right) - h\left(b, \frac{(1-\rho)R}{1-b}\right) - \frac{\rho R}{(b-a)^2} \int_a^b h_2\left(x, \frac{\rho R}{b-a}\right) dx \\ & + \frac{(1-\rho)R}{(1-b)^2} \int_b^1 h_2\left(x, \frac{(1-\rho)R}{1-b}\right) dx = 0 \end{aligned} \tag{4.3}$$

Another way to write these first-order conditions is to substitute s_k and r_k back into the equations.

$$H_\rho = \frac{R}{s_1} \int_a^b h_2\left(x, r_1\right) dx - \frac{R}{s_2} \int_b^1 h_2\left(x, r_2\right) dx = 0 \tag{4.1b}$$

$$H_a = h\left(a, r_1\right) - \frac{r_1}{s_1} \int_a^b h_2\left(x, r_1\right) dx = 0 \tag{4.2b}$$

$$H_b = -h\left(b, r_1\right) + h\left(b, r_2\right) + \frac{r_1}{s_1} \int_a^b h_2\left(x, r_1\right) dx - \frac{r_2}{s_2} \int_b^1 h_2\left(x, r_2\right) dx = 0 \tag{4.3b}$$

Equation 4.1 states that the full marginal output of a dollar spent at either school must be the same in equilibrium. Educational production per student depends not

[10] The other inequality constraints, $a = b$, $b = 1$, $\rho = 0$, and $\rho = 1$ all imply that one university is empty and unused. This cannot be optimal. Whenever the fixed cost has been paid, the optimal allocation uses all available schools to tailor resources per student. In the Cobb-Douglas case, which we explore in detail below, with $\underline{x} = 0$, $a = 0$ will not bind because the lowest student produces zero.

only on the total budget of a school, but also on the number of students over which this budget is spread; the key metric is resources per student. The price of an additional unit of resources per student in a school is equal to the size of the school. Rearranging 4.1 yields:

$$\frac{\text{Price of } r_2}{\text{Price of } r_1} = \frac{s_2}{s_1} = \frac{1-b}{b-a} = \frac{\int_b^1 h_2\left(x, \frac{1-\rho}{1-b}\right) dx}{\int_a^1 h_2\left(x, \frac{\rho}{b-a}\right) dx} = \frac{\text{Marginal Effect of } r_2}{\text{Marginal Effect of } r_1}$$

Equation 4.2 describes the condition for the lowest ability person who receives education. The first term represents the contribution to education made by the marginal person when he or she is admitted. The second term represents the reduction in education of those already at the school, due to congestion, when an additional student is added. When the marginal person is added to the school, holding the school's total resources fixed, the level of resources per student falls (at a rate of $\frac{\rho R}{(b-a)^2}$), and this causes a decrease (in the amount of $h_2(x, \rho R/(b-a))$) in production for each student in the school. Thus, at the optimum, the direct contribution of the marginal student just offsets the reduction that student causes by congesting resources.

Equation 4.3 describes a similar condition for the marginal student between schools. Suppose the decision is made to send the best person from the lower school to the upper school. Their direct contribution rises by the amount $h(b, \frac{(1-\rho)R}{1-b}) - h(b, \frac{\rho R}{b-a})$, as a result of attending a school with higher resources per student (recall from corollary 1 that the higher school will have more resources per student at the optimum). This gain is exactly equal to the net crowding effect. The other two terms in 4.3 are the combined marginal benefit in the lower school of moving the student and the combined marginal loss to the students at the upper school from increased congestion.

This marginal student faces a discontinuity. He or she would produce discretely more at the upper school. Because of the complementarity between resources and student ability, the students at the better school enjoy more resources per student. The top person in the lower tier is almost exactly the same as the lowest person in the upper tier in terms of ability, but there is a discrete gap in their educational outcomes.

4.4 The Optimal Number of Universities

The above analysis characterized the optimal allocation, taking the number of schools as fixed. The planner must also choose the optimal number of schools. This analysis is less straightforward because the problem is discrete.[11] We can develop some intuition by looking at the case where $\theta = 0$, which again allows the use of standard calculus. When there is no fixed cost, the optimal solution is to tailor the resources per student to each ability type, with the resources per student rising in student ability.

Proposition IV.2. *If there are no fixed costs ($\theta = 0$), the optimal solution features a unique level of funding (a unique school) for each student ability that is funded at a positive level. The optimal amount of resources per student, $r(x)$, is an increasing function with $r'(x) = \frac{-h_{21}(x,r(x))}{h_{22}(x,r(x))} > 0$.*

When there are no fixed costs, the planner tailors education quality specifically for each ability level. The proof of proposition IV.2 solves a basic control problem. The solution demonstrates that resources will be rising in student ability, and that the rate of this increase will depend on the curvature of h. Greater complementarity increases the slope of resources as ability rises. Greater concavity in the value of resources will dampen the relationship.

[11] One may wish to appeal to discrete optimization tools such as integer programming to solve such a problem. Unfortunately, integer programming techniques, such as cutting plane methods, are not applicable because they require first solving the case where variables are not constrained to be integers. This will not work here because the objective function is not defined for non-integer values of K.

When there are fixed costs, the planner can provide only finite tailoring, which implies that almost all ability types will receive resources different from the infinite school optimum. This creates both winners and losers. Figure 4.1 shows a hypothetical resources per student function for the no fixed cost case and the same function when $K = 2$. For $K = \infty$, $r(x)$ must be increasing. The area trapped by $r(x)$ will represent the net resources available, R, when the distribution of ability is uniform and measure 1. It is possible, but not necessary, that all students receive some education in this system.

Suppose that $K = 2$ is the constrained optimum (which will be the case for some values of θ). Some measure of students at the bottom may receive no education in the constrained case. Students between a and b attend the lower tier school, and students above b attend the upper tier school. The lowest ability students at each school receive more funding than they would in the case with perfect tailoring. In general, an increase in the ability to tailor resources will increase total educational output, but it will not be Pareto improving. The lowest students at each school will lose if the number of schools increases.

When there are fixed costs, the planner must balance the benefits of tailoring against the costs of setting up new universities. A graphical representation of the planner's global choice provides further intuition. Figure 4.2 shows several hypothetical curves in total resources versus total educational production space. These curves are the $W(T, S, \theta, K)$ value functions from program P2, for several values of K, with S and θ held constant. Each curve shows how total output changes as total resources rises, holding fixed the number of schools, and the measure and distribution of students. These curves are increasing and concave. The global optimum, $V(T, S, \theta)$, is the upper envelope of these curves.

A number of comparative statics can be visualized as the expansion or compression of figure 4.2. Holding T constant, a fall in θ compresses the graph. Each W curve shifts horizontally to the left, and curves at a higher K shift more. As a consequence, systems with lower θ will be more productive.

Proposition IV.3. *The average product of resources,* $\frac{V}{T}$*, of the optimal system is rising in the measure of students when resources per student is held constant and falling in the fixed cost per school.*

The second part of proposition IV.3 is rather obvious. Average productivity rises when more resources are available for education and fewer are required for paying the fixed cost. The first part follows from the implied economies of scale. Larger systems will spread the fixed cost over more students, allowing more money to be used as an input.

Figure 4.2 also provides insight into how the optimal number of universities is chosen. For the given value of T, the planner will choose the highest curve. Each W curve begins on the T-axis at $T = \theta K$. For T above that point, W is increasing and concave. If each W satisfies the single-crossing property, then the optimal number of schools must be rising in T. Currently, we are unable to prove (or disprove) that the single-crossing property is satisfied without any additional assumptions, though our intuition is that the property will hold for a fairly broad class of functions. This property holds in the Cobb-Douglas case, and we can prove several further results with this functional form.[12]

Proposition IV.4. *If* $h(x, r) = x^{\alpha} r^{\beta}$*, with* $0 < \alpha, \beta < 1$*, then the optimal number of schools,* K^**, is weakly rising in total resources,* T*.*

[12] The assumption of Cobb-Douglas can be slightly relaxed to an assumption that $h(x, r) = \alpha r^{\beta} g(x)$ without changing the proof used. We suspect that this property is true for a broader set of h functions, but the current proof uses the multiplicative separability of Cobb-Douglas, which is a relatively strong assumption.

Proposition IV.4 implies that richer systems should have more schools, thereby achieving better tailoring. Note, again, that better tailoring does not mean that all students benefit. Within each school, there are students who receive more resources than they would if perfect tailoring were feasible. Thus, there will be losers from an increase in total system resources if the addition of resources causes a rise in the number of schools. In particular, some portion of the lowest ability students at any given school will experience a decrease in educational quality when K rises. Increases in T are not, therefore, necessarily Pareto-improving, even if resources are dropped exogenously into the system.

Proposition IV.4 is closely tied to two additional comparative static results, which relate the optimal number of schools to the fixed cost of establishing a school and to the size of a system.

Proposition IV.5. *If $h(x, r) = x^{\alpha} r^{\beta}$, then the optimal number of schools, K^*, is rising in the measure of students when resources per student is held constant and falling in the fixed cost per school.*

The intuition behind proposition IV.5 is clear from figure 4.2. A reduction in θ shifts all the curves to the left. Curves with higher K values shift more. Thus, the diagram is contracted, and the cut-off points all move to the left. Holding T constant, K^* must weakly rise as a result. Raising the measure of students, while keeping total resources per student constant, has the same effect on the cut-off points.

University systems that serve a larger population, therefore, should be superior in several ways. Even if they are not richer per student, they should have more universities. They should do a better job of tailoring educational quality to students, and they should produce more per dollar of resources and more per student.

Larger university systems will also serve a greater fraction of the ability distribution, and they will feature more selective flagship universities.

Conjecture 1. *If $h(x,r) = x^\alpha r^\beta$, then the selectivity of the top university will be rising in K, holding R constant.*

This remains a conjecture, because we have been unable to prove this for the general case, but there are reasons to believe that the claim is true. First, it is clearly true in the limit. As the number of schools approaches infinity, the top school will become arbitrarily selective. Second, we have investigated this claim numerically, assuming that ability is distributed uniformly on $[0,1]$. Our assumptions require that $0 < \alpha < 1$ and $0 < \beta < 1$. We performed a grid search over these intervals with a .1 width, for $K = 1$ through $K = 5$.[13] For each α, β pair we numerically located the optimal cut points for each value of K and checked that the selectivity of the top school is rising in K. This procedure revealed no counterexamples. These examples, of course, do not prove the conjecture. Note, however, that even if there is some set of values for α, β and K that generate a counterexample to the claim, the predicted relationship will likely still emerge in the real world. A similar result about the low end of the distribution holds in simulations. Systems with more universities will serve a larger fraction of the distribution.[14]

We selected additional parameter values and extended the search up to $K = 10$. Two examples are provided in figure 4.3 for illustration. When $K = 1$, the lowest type admitted to the top school is the same as the lowest type admitted to the bottom school. As K rises, the lowest type admitted to the top school (A_K) also rises. The limiting argument suggests that as $K \to \infty, A_K \to 1$, giving a sense

[13] The numerically estimated solution is very sensitive to starting values when the parameters are near 0 or 1, which necessitates an extra layer of search. We checked many values close to 0 and 1, and we performed a finer grid search over the middle of the parameter space (from .2 to .8) where starting value sensitivity is reduced.

[14] An alternative approach is to calibrate the model. We prefer the grid search primarily because we do not believe there is a reliable way to calibrate α and β. Since we find no contradictions to our claim throughout the entire parameter space, we feel that the grid search is more comprehensive than a calibrated example, which would focus on a single pair of α and β.

of how these curves would project forward. A corresponding shape exists for the lowest student admitted to the bottom school, with this value approaching 0 in the limit. The corresponding curves have a similar shape for each of the large number of parameter value pairs that we have examined.

These results suggest that schools in states with a larger number of universities should have more selective flagships. Descriptive data from the Integrated Postsecondary Education Data System on university characteristics in 2001 support this hypothesis.[15] Figure 4.4 plots the 75th percentile of the combined SAT scores for students at each state's flagship university against the number of two- and four-year public universities in that state. It is clear from the graph that states which have more institutions (better tailoring) feature a more selective flagship university. Figure 4.5 plots the percentage of applicants admitted by the flagship university against the number of public universities in that state. The data again suggest that larger university systems have more selective flagships.

4.5 Extensions and Conclusions

The purpose of this paper is to provide a framework for analyzing the optimal allocation of students and resources within a system of higher education. Our hope is that future research will enrich the model and test its implications.

Our model does not include tuition.[16] At the optimum, the social cost of moving a student from their assigned university to a better one is the change in their educational output minus the net crowding effect. If individuals experience a private gain from educational output, there will be some measure of students at any university for whom the private gain from a university upgrade will outweigh the total social

[15] These data are available at http://nces.ed.gov/ipeds/.

[16] The existing literature primarily considers tuition policies that enable ability screening for schools maximizing quality (e.g., Epple and Romano (1998); Epple et al. (2003, 2006)).

cost. This suggests that there are gains to be made by allowing students to pay for an upgrade.

Similarly, students might be willing to pay a premium to attend a university out of state. If the social planner's objective function includes only the education of in-state residents, the optimal tuition policy will be to admit out-of-state students as long as their tuition, at the margin, exceeds the current resources per student at a school. In general, the introduction of tuition policy will make the total amount of system resources an endogenous variable.

Our model also makes empirical predictions about the relationship between the number of universities in a system and the selectivity of its flagship university and about the effects of introducing an additional university to a system. When new universities are introduced, our model suggests that some types of students will experience a reduction in educational quality, while others will experience an increase. At the same time, overall educational output, and the marginal value of additional revenue, should rise with the introduction of a new university. Our hope is that future research will utilize variation in the fixed cost (e.g., land grants, changes in federal support) and the size of the population (e.g., migration, the Baby-Boom) to test and further refine our findings.

In this paper, we focused on a deliberately simple model. Nevertheless, it captures a number of key features about the provision of public higher education. In particular, our model offers a normative explanation for a tiered university system, within which higher ability students receive more resources. It highlights the tradeoff inherent in tailoring education quality to student ability. It also provides a model for understanding the optimal number of universities in a system, and makes suggestions about how university systems should vary.

4.6 Appendix

Proposition IV.1. *If $h(x, r)$ is complementary (supermodular), then the optimal partition of students is monotonic.*

Proof: Fix the number of schools and the resources in each school.[17] Suppose the optimal partition is not monotonic. Call the two schools that violate monotonicity 1 and 2, and, without loss of generality, assume 1 has the higher resources per student. If monotonicity fails, then $\exists\ y \in 1 < z \in 2$. The proposed solution produces $h(y, r_1) + h(z, r_2)$. Switching the two students yields $h(y, r_2) + h(z, r_1)$. And,

$$h(y, r_2) + h(z, r_1) > h(y, r_1) + h(z, r_2) \Leftrightarrow$$

$$h(z, r_1) - h(z, r_2) > h(y, r_1) - h(y, r_2)$$

which is a definition of supermodularity, since $z > y$ and $r_1 > r_2$. **QED.**

Corollary 1. *In any optimal monotonic partition, any school that has higher ability students than another school will also have higher resources per student.*

Proof: Suppose that the optimal partition is monotonic, with students in 1 being higher ability than students in 2, but with $r_2 > r_1$. By supermodularity, swapping any two students between schools raises output. **QED.**

Proposition IV.2. *If there are no fixed costs ($\theta = 0$), the optimal solution features a unique level of funding (a unique school) for each student ability that is funded at a positive level. The optimal amount of resources per student, $r(x)$, is an increasing function with $r'(x) = \frac{-h_{21}(x,r(x))}{h_{22}(x,r(x))} > 0$.*

Proof: Part I: Suppose that there is a school with positive resources and two or more distinct ability types. Then there exists some school with both y and z with

[17] Clearly, if any two schools provide the same resources per student, there would be economies of scale gains to merging the schools. Thus, we can proceed as if the resources per student differs at each school.

$z > y$. Let s denote the total measure of students at the school. Without loss of generality, suppose that the measure of each type is the same, $s/2$. Then, the output of the proposed optimum can be written:

$$V = \frac{1}{s}h\left(y, \frac{\rho R - \varepsilon}{s}\right) + \frac{1}{s}h\left(z, \frac{\rho R + \varepsilon}{s}\right) \tag{4.4}$$

where $\varepsilon = 0$ and total funding at the school is ρR. We show that it is optimal to set $\varepsilon > 0$, which is equivalent to separating y and z into two different schools:

$$\frac{\partial V}{\partial \varepsilon} = -h_2\left(y, \frac{\rho R - \varepsilon}{s}\right) + h_2\left(z, \frac{\rho R + \varepsilon}{s}\right) > 0$$

The last inequality follows directly from supermodularity and contradicts the optimality of the proposed solution.

Part II: The infinite school problem may be written as a control problem:

$$\max_{r(x)} \int_0^1 h(x, r(x))dx$$
$$\text{s.t.} \quad \int_0^1 r(x)dx = T$$

It can easily be shown that the Hamiltonian leads to a degenerate solution with $h_2(x, r(x)) = \beta \in \mathbb{R}$. This is an implicit function, and the conditions of the implicit function theorem are satisfied because $h_{22}(\cdot) \neq 0$. The implicit function theorem yields the final result, which we sign from our assumption of concavity and complementarity:

$$r'(x) = \frac{-h_{21}(x, r(x))}{h_{22}(x, r(x))} > 0.$$

QED.

Proposition IV.3. *The average product of resources, $\frac{V}{T}$, of the optimal system is rising in the measure of students when resources per student is held constant and falling in the fixed cost per school.*

Proof: Fix T and S. Lowering θ relaxes the resource constraint for any value of K. $V(T,S,\theta)$ must therefore rise. Since T is fixed, $\frac{V}{T}$ must rise.

A rise in the measure of students when resources per student is held constant means that S rises but $\frac{T}{S}$ is fixed. We can write this as a γ proportional change, with $\gamma > 1$. Denote the value function as $W(T,S,\theta,K)$. For any value of K, output per dollar of total resources can be written:

$$
\begin{aligned}
\frac{W(\gamma T, \gamma S, \theta, K)}{\gamma T} &= \max_{\rho_k, a} \frac{\gamma S}{\gamma T} \sum_{k=1}^{K} \int_{a_k-1}^{a_k} h\left(x, \frac{\rho_k(\gamma T - \theta K)}{\gamma s_k}\right) f(x)dx \\
&= \max_{\rho_k, a} \frac{S}{T} \sum_{k=1}^{K} \int_{a_k-1}^{a_k} h\left(x, \frac{\rho_k(\gamma T - \theta K)}{\gamma s_k}\right) f(x)dx \\
&> \max_{\rho_k, a} \frac{S}{T} \sum_{k=1}^{K} \int_{a_k-1}^{a_k} h\left(x, \frac{\rho_k(T - \theta K)}{s_k}\right) f(x)dx \\
&= \frac{W(T,S,\theta,K)}{T}
\end{aligned}
$$

The second equality uses the envelope theorem, which tells us that the optimal cut-points will not change when the parameters are varied in small amounts. The inequality uses the fact that $\frac{\partial}{\partial\gamma}\frac{\rho_k(\gamma T - \theta K)}{\gamma s_k} = \frac{\rho_k \theta K}{\gamma s_k} > 0$. Since this is true of any K, it must be true for the optimal K. **QED.**

Proposition IV.4. *If* $h(x,r) = x^\alpha r^\beta$*, with* $0 < \alpha, \beta < 1$*, then the optimal number of schools,* K^**, is weakly rising in total resources,* T*.*

Proof: Define $W(T,S,\theta,K)$ to be the constrained solution, when the number of universities is fixed at K, and denote the derivative of $W(\cdot)$ with respect to T by

W_T. In the Cobb-Douglas case, we can relate W and W_T:

$$\begin{aligned}
W(T,S,\theta,K) &= \max_{\rho_k,a} S \sum_{k=1}^{K} \int_{a_{k-1}}^{a_k} x^\alpha \left(\frac{\rho_k R}{(a_k - a_{k-1})S}\right)^\beta dx \\
&= \max_{\rho_k,a} S \left(\frac{R}{S}\right)^\beta \sum_{k=1}^{K} \int_{a_{k-1}}^{a_k} x^\alpha \left(\frac{\rho_k}{a_k - a_{k-1}}\right)^\beta dx \\
&= \max_{\rho_k,a} S^{1-\beta}(T-\theta K)^\beta \sum_{k=1}^{K} \int_{a_{k-1}}^{a_k} x^\alpha \left(\frac{\rho_k}{a_k - a_{k-1}}\right)^\beta dx \\
W_T(T,S,\theta,K) &= \max_{\rho_k,a} \beta S^{1-\beta}(T-\theta K)^{\beta-1} \sum_{k=1}^{K} \int_{a_{k-1}}^{a_k} x^\alpha \left(\frac{\rho_k}{a_k - a_{k-1}}\right)^\beta dx
\end{aligned}$$

The vector of ρ and a values that maximize the objective function will also maximize the marginal value of resources, W_T, since W_T is an affine transformation of W: $W_T(T,S,\theta,K) = \frac{\beta}{T-\theta K} W(T,S,\theta,K)$. Now, consider any two numbers of universities, with $K > \hat{K}$:

$$\begin{aligned}
W_T(T,S,\theta,K) &= \frac{\beta}{T-\theta K} W(T,S,\theta,K) \\
&> \frac{\beta}{T-\theta K} W(T,S,\theta,\hat{K}) \\
&> \frac{\beta}{T-\theta \hat{K}} W(T,S,\theta,\hat{K}) \\
&= W_T(T,S,\theta,\hat{K})
\end{aligned}$$

Since the derivative of W with respect to T is higher the higher is K, the family of W functions will satisfy the single crossing property in the $T-W$ plane. For each $K > \hat{K}$ and $T > \hat{T}$, $W(\hat{T},S,\theta,K) > W(\hat{T},S,\theta,\hat{K}) \Rightarrow W(T,S,\theta,K) > W(T,S,\theta,\hat{K})$. As is illustrated in figure 4.2, W will be zero up until $T = \theta K$. So, W functions with higher K values start rising at a later point.

$V(T,S,\theta)$, the optimum when K is a choice variable, is the upper envelope of the family of W functions in figure 4.2. Because of the single crossing property, this upper envelope must lie on a W for a weakly higher K as T rises. Thus, K^* is rising in T. **QED.**

Proposition IV.5. *If $h(x,r) = x^\alpha r^\beta$, then the optimal number of schools, K^*, is rising in the measure of students when resources per student is held constant and falling in the fixed cost per school.*

Proof: Define $T^*(i,j)$ as the T that solves $W(T^*(i,j), S, \theta, i) = W(T^*(i,j), S, \theta, j)$, as in figure 4.2. Define $Q(K) = \max_{\rho_k, a} \sum_{k=1}^{K} \int_{a_{k-1}}^{a_k} x^\alpha \left(\frac{\rho_k}{a_k - a_{k-1}}\right)^\beta dx$. For the Cobb-Douglas case, as is shown in the proof of proposition IV.4, $W(T, S, \theta, K) = S^{1-\beta}(T - \theta K)^\beta Q(K)$. Therefore, $T^*(i,j)$, which will be unique if it exists by the single-crossing property from proposition IV.4, solves

$$S^{1-\beta}(T^*(i,j) - \theta i)^\beta Q(i) = S^{1-\beta}(T^*(i,j) - \theta j)^\beta Q(j) \tag{4.5}$$

Without loss of generality, assume $i < j$. Totally differentiate equation 4.5 with respect to T and θ:

$$\begin{aligned}&(dT^*(i,j) - id\theta)\beta S^{1-\beta}(T^*(i,j) - \theta i)^{\beta-1}Q(i)\\ &= (dT^*(i,j) - jd\theta)S^{1-\beta}(T^*(i,j) - \theta j)^\beta Q(j)\end{aligned}$$

Rearrange:

$$\begin{aligned}\frac{dT^*(i,j)}{d\theta} &= \frac{j\beta S^{1-\beta}(T^*(i,j) - \theta j)^\beta Q(j) - i\beta S^{1-\beta}(T^*(i,j) - \theta i)^{\beta-1}Q(i)}{\beta S^{1-\beta}(T^*(i,j) - \theta j)^\beta Q(j) - \beta S^{1-\beta}(T^*(i,j) - \theta i)^{\beta-1}Q(i)}\\ &= \frac{jW_T(T^*(i,j), S, \theta, j) - iW_T(T^*(i,j), S, \theta, i)}{W_T(T^*(i,j), S, \theta, j) - W_T(T^*(i,j), S, \theta, i)}\\ &> 0\end{aligned}$$

The last two steps follow directly from the analysis in the proof of proposition IV.4. This shows that all cut-off points rise when θ rises. This implies that, when T is held constant, a rise in θ must weakly decrease the number of cut-off points passed with total resources T, which is equivalent to a weakly falling K^*.

For the size result, note that $W(\gamma T, \gamma S, \theta, K) = (\gamma S)^{1-\beta}(\gamma T - \gamma\frac{\theta}{\gamma}K)^\beta Q(K) = \gamma W(T, S, \frac{\theta}{\gamma}, K)$. Since γ does not depend on any of the parameters, the cut-off points for the $\gamma W(T, S, \frac{\theta}{\gamma}, K)$ system are equivalent to the cut-off points for the $W(T, S, \frac{\theta}{\gamma}, K)$ system. Thus, an increase in the measure of students, holding constant total resources per student, which is equivalent to choosing $\gamma > 1$, is equivalent in its effect on K^* to a reduction of θ to $\frac{\theta}{\gamma}$. Since K^* was proved above to be weakly falling in θ, it must be weakly rising in the measure of students. **QED.**

4.7 Figures

Figure 4.1: Resources per Student as a Function of Student Ability

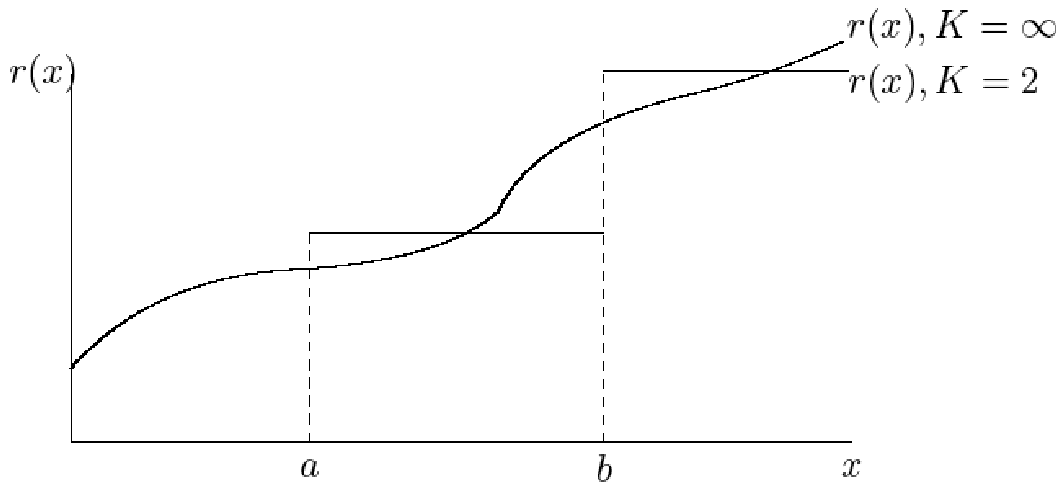


Figure 4.2: Total Output versus Total Resources for Several Values of K

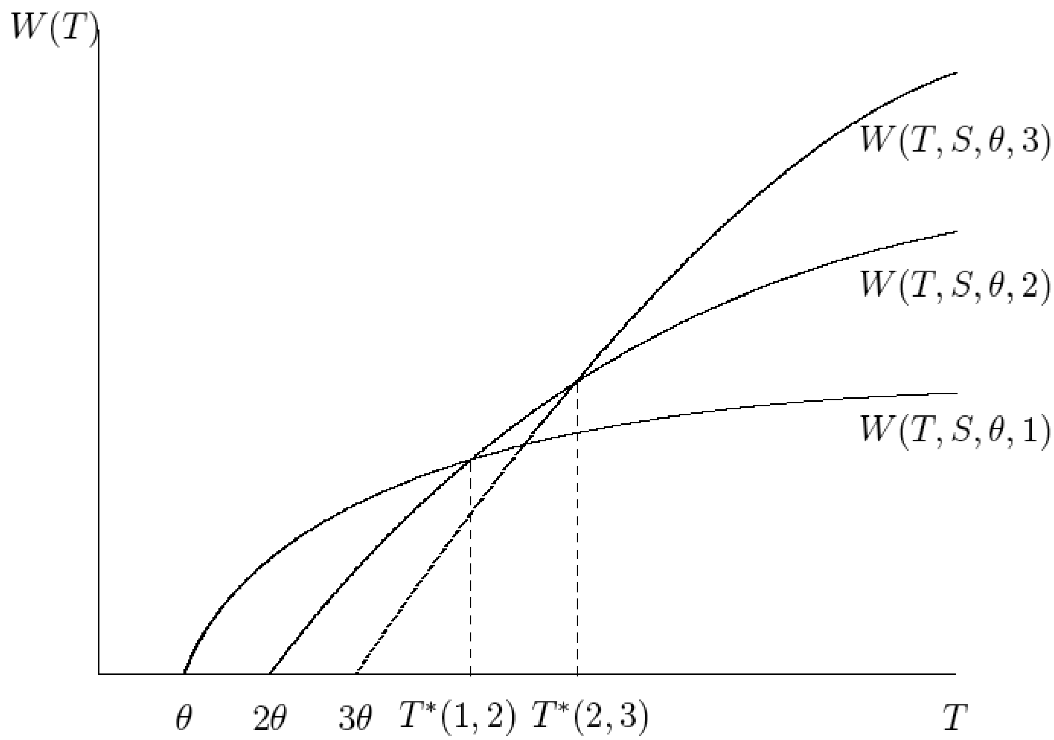


Figure 4.3: Numerically Estimated Optimal Values of A_0 and A_K for Selected α and β

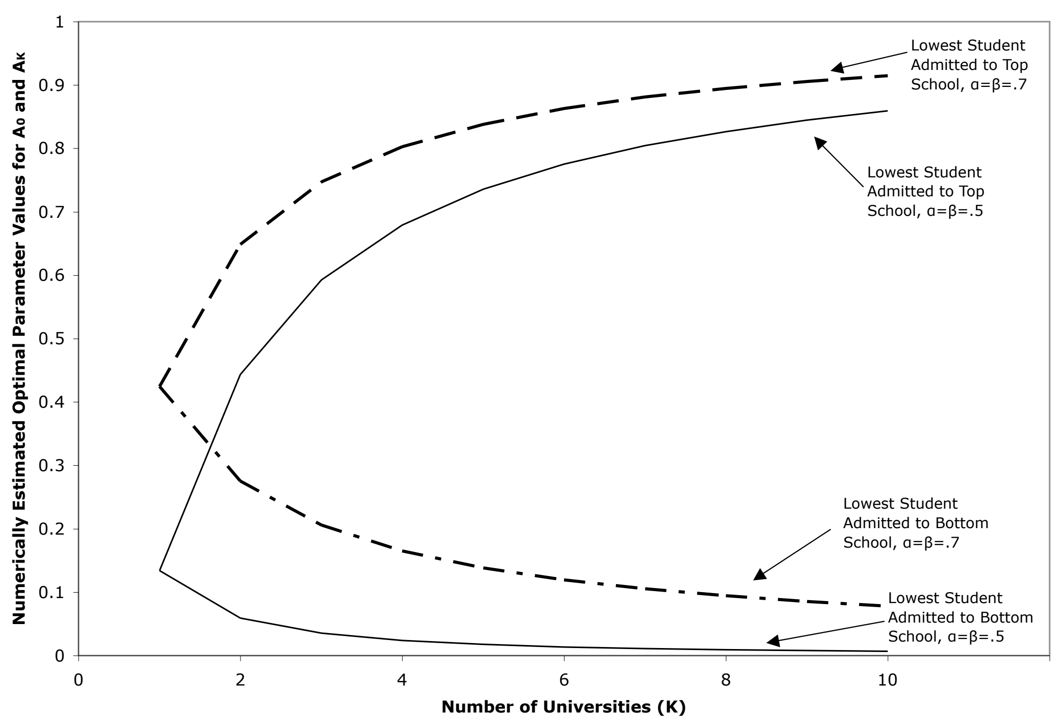
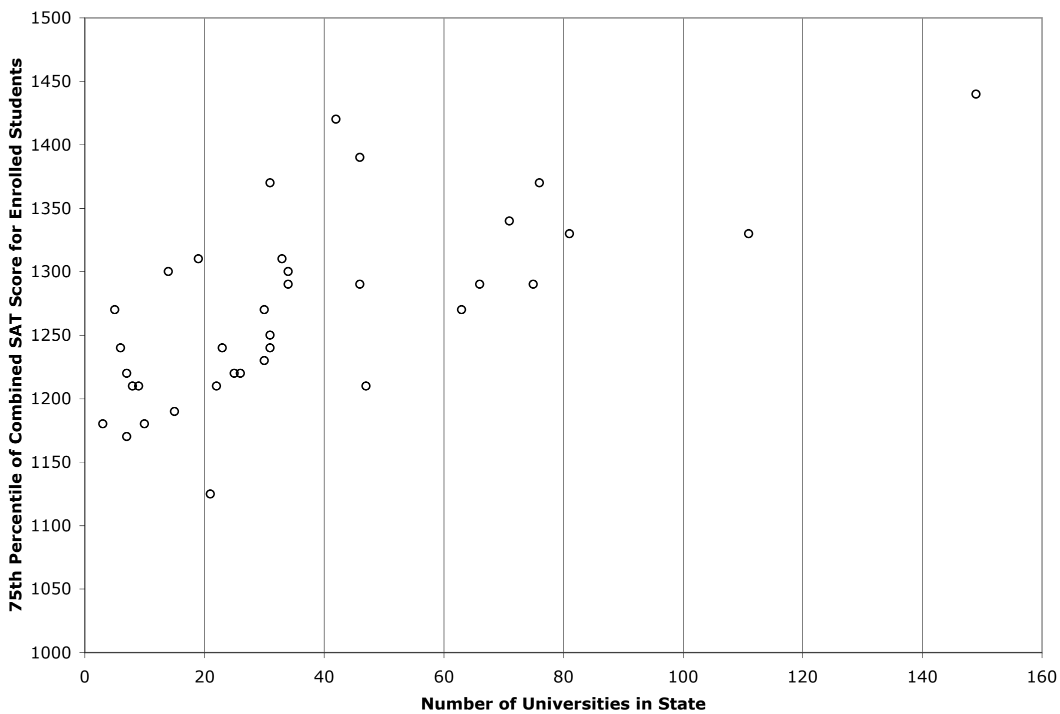
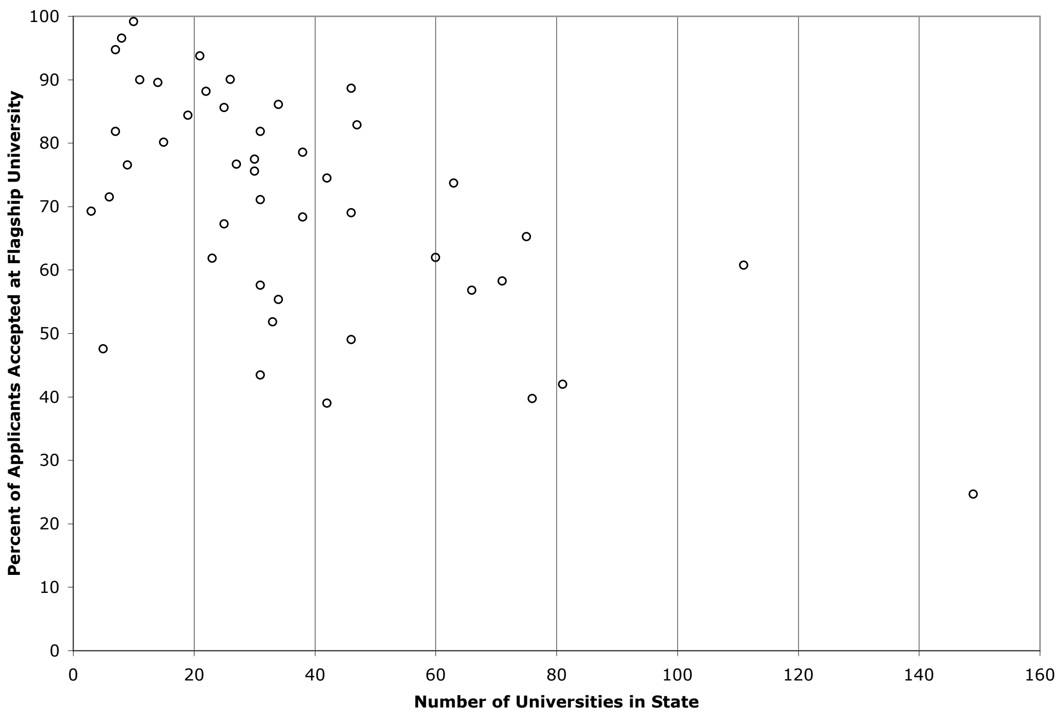


Figure 4.4: Flagship SAT Scores versus Number of Universities in State



Data include all 36 states with available SAT scores from the 2001 Integrated Postsecondary Education Data System. A regression of 75th Percentile SAT scores at the state flagship on the number of colleges in a state yields the following estimates: SAT = 1213 (14) + 1.52 (.29) * Number of Schools + error, where standard errors are in parentheses.

Figure 4.5: Flagship Admissions Acceptance versus Number of Universities in State



Data include all 45 states with available admissions data from the 2001 Integrated Postsecondary Education Data System. A regression of fraction admitted to the flagship on the number of colleges in a state yields the following estimates: Fraction Admitted = 107 (15) * -1.00 (.20) Number of Schools + error, where standard errors are in parentheses.

CHAPTER V

Conclusion

Together, the three essays in this dissertation cover aspects of the distribution and value of resources through all levels of public education in the United States.The Abbott case has very positive and very negative conclusions. New Jersey succeeded in dramatically increasing the resources available to urban schools but failed to put any forethought into how they would gauge success. The state had the opportunity to learn a tremendous amount about the costs and benefits of various educational programs and interventions. The results in the first two essays show that Abbott districts increased spending and that the achievement of some students improved dramatically, but the data do not exist to determine which programs or expenditures led to the measured increases. The third essay, starts at the opposite end of the spectrum. Starting with just a budget and knowledge of how student ability and resources interact, we show how to allocate students and resources across schools to produce the most educational output. All three essays lead to more unanswered questions and it is my hope that I, and others, will continue to work toward answering them.

Bibliography

Abbott v. Burke, 119 N.J. 287 (Abbott II), 1990.

Abbott v. Burke, 136 N.J. 444 (Abbott III), 1994.

Abbott v. Burke, 149 N.J. 145 (Abbott IV), 1997.

A Conversation on New Jersey School Funding: Past, Present and Future, 2006. Public Forum held at Education Testing Service in Princeton, NJ, on May 23, 2006.

Robinson v. Cahill, 67 N.J. 35, 1975.

James Andreoni. Toward a theory of charitable fund-raising. *Journal of Political Economy*, 106(6):1186–1213, December 1998.

Richard Arnott and John Rowse. Peer group effects and educational attainment. *Journal of Public Economics*, 32(3):287–305, April 1987.

Stephen J. Bailey and Stephen Connolly. The flypaper effect: Identifying areas for further research. *Public Choice*, 95:335–361, 1998.

Gary S. Becker. A theory of marriage: Part 1. *Journal of Political Economy*, 91(4): 813–846, July - August 1973.

Kelly Bedard and William O. Brown Jr. The allocation of public school expenditures, 2000. Claremont Colleges Working Papers, number 2000-16.

Roland Bénabou. Equity and efficiency in human capital investment: The local connection. *The Review of Economic Studies*, 63(2):237–264, April 1996.

Julian R. Betts. Which types of public school spending are most effective? New evidence on the school quality debate. University of California at San Diego, Economics Working Paper Series 95-03, Department of Economics, UC San Diego, January 1995.

David F. Bradford and Wallace E. Oates. Towards a predictive theory of intergovernmental grants. *The American Economic Review*, 61(2):440–448, 1971.

Jeanne Brooks-Gunn. Do you believe in magic?: What we can expect from early childhood intervention programs. *Social Policy Report*, 17(1):3–14, 2003.

Center on Education Policy. Choices, changes, and challenges: Curriculum and instruction in the NCLB era. A report in the series From the Capital to the Classroom: Year 5 of the No Child Left Behind Act, July 2007.

Charles T. Clotfelter, Elizabeth Glennie, Helen F. Ladd, and Jacob L. Vigdor. Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. NBER Working Paper 12285, June 2006.

David K. Cohen, Stephen W. Raudenbush, and Deborah Loewenberg Ball. Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25 (2):119–142, 2003.

Paul N. Courant and Susanna Loeb. Centralization of school finance in michigan. *Journal of Policy Analysis and Management*, 16(1):114–136, 1997.

Paul N. Courant, Michael McPherson, and Alexandra M. Resch. The public role in higher education. *National Tax Journal*, 59(2):291–318, June 2006.

Digest of Education Statistics. Table 158. Revenues for public elementary and secondary schools, by source of funds: Selected years, 1919-20 through 2003-04, 2006.

William Duncombe and John Yinger. How much more does a disadvantaged student cost? *Economics of Education Review*, 24:513–532, 2005.

Matthias R. Effinger and Mattias K. Polborn. A model of vertically differentiated education. *Journal of Economics*, 69(1):53–69, February 1999.

Dennis Epple and Richard E. Romano. Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, 88(1):33–62, March 1998.

Dennis Epple, Richard Romano, and Holger Sieg. Peer effects, financial aid and selection of students into colleges and universities: An empirical analysis. *Journal of Applied Econometrics*, 18(5):501–525, September - October 2003.

Dennis Epple, Richard Romano, and Holger Sieg. Admission, tuition, and financial aid policies in the market for higher education. *Econometrica*, 74(4):885–928, July 2006.

Bari Anhalt Erlichson. Comprehensive school reform in new jersey:waxing and waning support for model implementation. *Journal of Education for Students Placed at Risk*, 10(1):11–32, 2005.

William N. Evans, Sheila E. Murray, and Robert M. Schwab. Schoolhouses, courthouses, and statehouses after Serrano. *Journal of Policy Analysis and Management*, 16(1):10–31, 1997.

Martin Feldstein. The effect of a differential add-on grant: Title I and local education spending. *The Journal of Human Resources*, 13(4):443–458, 1978.

Ronald F. Ferguson. Can schools narrow the black-white test score gap? In Christopher Jencks and Meredith Phillips, editors, *The Black-White Test Score Gap*. Brookings Institution Press, 1998.

Raquel Fernandéz and Jordi Galí. To each according to...? Markets, tournaments, and the matching problem with borrowing constraints. *The Review of Economic Studies*, 66(4):799–824, October 1999.

William A. Firestone, Margaret E. Goertz, Brianna Nagle, and Marcy F. Smelkinson. Where did $800 million go? The first year of New Jersey's Quality Education Act. *Educational Evaluation and Policy Analysis*, 16(4):359–373, Winter 1994.

William A. Firestone, Margaret E. Goertz, and Gary Natriello. *From Cashbox to Classroom: The Struggle for Fiscal Reform and Educational Change in New Jersey*. Teachers College Press, New York, 1997.

Ronald C. Fisher and Leslie E. Papke. Local government responses to education grants. *National Tax Journal*, 53(1):153–168, March 2000.

Ellen Frede, Kwanghee Jung, W. Steven Barnett, Cynthia Esposito Lamy, and Alexandra Figueras. The abbott preschool program longitudinal effects study (apples). http://nieer.org/resources/research/APPLES.pdf, June 2007.

Margaret E. Goertz and Gary Natriello. Court-mandated school finance reform: What do the new dollars buy? In Helen F. Ladd, Rosemary Chalk, and Janet S. Hansen, editors, *Equity and Adequacy in Education Finance: Issues and Perspectives*. National Academy Press, Washington, D.C., 1999.

Timothy J. Goodspeed. The relationship between state income taxes and local property taxes: Education finance in New Jersey. *National Tax Journal*, 51(2):219–238, June 1998.

Nora Gordon. Do federal grants boost school spending? Evidence from Title I. *Journal of Public Economics*, 88(9-10):1771–1792, 2004.

Eric A. Hanushek. The failure of input-based schooling policies. *The Economic Journal*, 113(485):F64–F98, 2003.

Eric A. Hanushek, editor. *Courting Failure: How School Finance Lawsuits Exploit Judges' Good Intentions and Harm our Children*. Hoover Institution Press, Stanford, CA, 2006.

Eric A. Hanushek. School resources and student performance. In Gary Burtless, editor, *Does Money Matter?: The Effect of School Resources on Student Achievement and Adult Success*, pages 43–73, 1996.

Eric A. Hanushek, John F. Kain, and Steven G. Rivkin. Why public schools lose teachers. *The Journal of Human Resources*, 39(2):326–354, 2004.

James R. Hines and Richard H. Thaler. Anomalies: The flypaper effect. *Journal of Economic Perspectives*, 9(4):217–226, 1995.

Caroline M. Hoxby. All school finance equalizations are not created equal. *Quarterly Journal of Economics*, 116(4):1189–1231, 2001.

William Hussar and William Sonnenberg. Trends in disparities in school district level expenditures per pupil, 2000. Also available as `http://nces.ed.gov/pubs2000/2000020.pdf`.

Christopher Jencks and Meredith Phillips, editors. *The Black-White Test Score Gap*. Brookings Institution Press, Washington, D.C., 1998.

Alan B. Krueger. Economic considerations and class size. *The Economic Journal*, 113(485):F34–F63, 2003.

Cynthia Esposito Lamy, Ellen Frede, Holly Seplocha, Janis Strasser, Saigeetha Jambunathan, Jo Anne Junker, and Ellen Wolock. Giant steps for the littlest children: Progress in the sixth year of the abbott preschool program. May 2005.

Susanna Loeb and Marianne E. Page. Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *The Review of Economics and Statistics*, 82(3):393–408, 2000.

Gordon MacInnes. Abbott districts convocation: Closing the gap with non-Abbotts in fourth grade, 2005. Also available as `http://www.nj.gov/njded/abbotts/announce/convocation.shtml`.

Paul A. Minorini and Stephen D. Sugarman. School finance litigation in the name of educational equity: Its evolution, impact, and future. In Helen F. Ladd, Rosemary Chalk, and Janet S. Hansen, editors, *Equity and Adequacy in Education Finance: Issues and Perspectives*. National Academy Press, Washington, D.C., 1999.

Sheila E. Murray, William N. Evans, and Robert M. Schwab. Education-finance reform and the distribution of education resources. *American Economic Review*, 88(4):789–812, 1998.

NCES. *Digest of Education Statistics 2004*. Government Printing Office, U.S. Department of Education, National Center for Education Statistics, 2005a.

NCES. *Projections of Education Statistics to 2014*. Government Printing Office, U.S. Department of Education, National Center for Education Statistics, 2005b.

New Jersey Department of Education. Comparative spending guide, 1999-2006. Downloaded at `http://www.state.nj.us/njded/guide/`.

New Jersey Department of Education. Historical report card data, 1995-2002. Downloaded at `http://education.state.nj.us/rc/2002/NJ-CSV-Files620.zip`.

New Jersey Department of Education. DFG report, 2004. Also available as `http://www.nj.gov/njded/finance/sf/dfg.pdf`.

New Jersey Department of Education. State summary: Grade 8 early warning test (EWT), 1991, 1993-1996.

New Jersey Department of Education. State aid summary, 1998-2006. Also available as `http://www.state.nj.us/njded/stateaid/`.

Maria Newman. Court-Ordered Windfall Sends Schools on a Shopping Spree. *The New York Times*, March 29, 1998.

Deborah Reed, Kim S. Rueben, and Elisa Barbour. *Retention of New Teachers in California*. Public Policy Institute of California, San Francisco, CA, 2006.

Monica Rhor. A school feels impact from parity funding. *The Philadelphia Inquirer*, page B01, December 25, 1998.

Michael Rothschild and Lawrence J. White. The analytics of the pricing of higher education and other services in which the customers are inputs. *Journal of Political Economy*, 103(3):573–586, June 1995.

The New York Times. A truce in New Jersey's school war, February 9, 2002.

Petra E. Todd and Kenneth I. Wolpin. On the specification and estimation of the production function for cognitive achievement*. *The Economic Journal*, 113(485): F3–F33, 2003.

Martin R West and Paul E. Peterson, editors. *School Money Trials: The Legal Pursuit of Educational Adequacy*. Brookings Institution Press, Washington, D.C., 2007.

P.K. William, D. McLaughlin, M.C. Glander, and W.J. Fowler Jr. CCD data file: Longitudinal school district fiscal-nonfiscal detail file, fiscal years 1990 to 2002 (NCES 2006-320), 2006a. Downloaded at `http://nces.ed.gov/edfin/links/findata.asp`.

P.K. William, D. McLaughlin, M.C. Glander, and W.J. Fowler Jr. Documentation of the NCES longitudinal school district fiscal-nonfiscal detail file, fiscal years 1990 to 2002 (NCES 2006-320), 2006b. Also available as `http://nces.ed.gov/edfin/pdf/2006320.pdf`.

Gordon C. Winston. Subsidies, hierarchy and peers: The awkward economics of higher education. *Journal of Economic Perspectives*, 13(1):13–36, Winter 1999.

John Yinger. State aid and the pursuit of educational equity: An overview. In John Yinger, editor, *Helping Children Left Behind*. The MIT Press, 2004.