ABSTRACT

Title of dissertation: FOUR DECADES OF CHANGE IN U.S. PUBLIC EDUCATION: ESSAYS ON TEACHER QUALITY AND SCHOOL FINANCE

Sean Patrick Corcoran, Doctor of Philosophy, 2003

Dissertation directed by: Professor William N. Evans
Department of Economics

Several decades of research in the economics of education have shown that both the quality and quantity of school resources are important for student outcomes. In this dissertation, I present two essays that address changes in both the quality and quantity of resources available to public schools over the past four decades (1960 – 2000).

First, in chapter two I examine how the propensity for high test-scoring females to enter the teaching profession has changed over a forty-year period of occupational desegregation. While it has long been presumed that improved labor market opportunities for women have adversely affected the quality of teachers (over three quarters of whom are female), there is surprisingly little evidence measuring the extent to which this is true. In this essay, I combine data from five longitudinal surveys of high school graduates spanning the years 1957 to 2000 to evaluate this claim. I find that while

the test score ranking of the average new female teacher has fallen only slightly over this period, the likelihood that a female in the top decile of her high school class entered teaching has plummeted.

Next, in chapter three I examine the impact of rising within school district population heterogeneity and income inequality on local per-pupil expenditure and public school participation rates. Like the nation at large, the populations of school districts in the United States have become significantly more diverse, in (among other dimensions) racial and ethnic background, schooling, and income. Using a merged panel of school district demographics and financial data for 8,700 unified school districts over the 1970 to 2000 period, I look at the effects of this rising heterogeneity on the support for local public schools. I find that rising within-district income inequality is associated with greater per-pupil expenditure, a result consistent with a median voter model in which a lower tax price to the median voter results in greater per-pupil spending. Greater fractionalization in race and educational attainment appears to reduce per-pupil expenditure and increase enrollment in private schools.

FOUR DECADES OF CHANGE IN U.S. PUBLIC EDUCATION: ESSAYS ON TEACHER QUALITY AND SCHOOL FINANCE

by

Sean Patrick Corcoran

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2003

Advisory Committee:

Professor William N. Evans, Chair
Professor Ginger Z. Jin
Professor Wallace E. Oates
Professor Jennifer King Rice
Professor Robert S. Schwab

ACKNOWLEDGEMENTS

I would like to thank Dr. William Evans and Dr. Robert Schwab for their persistent support, guidance and encouragement throughout the writing of this dissertation. Without their belief in me, this work would not have been possible. I would also like to thank the participants of the University of Maryland graduate microeconomics seminar—in particular Drs. Ginger Jin, Seth Sanders, and Judith Hellerstein—for their insightful comments and suggestions. Chapter two of this dissertation also benefited from advice received at the "Teachers and Low Performing Schools" seminar at the November 2001 APPAM conference in Washington, D.C., and from two anonymous referees. I am indebted to Julie Topieski and Jennifer Godwin for introducing me to much of the data used in chapters two and three, respectively, and to Amy Harris for her efforts in compiling an earlier version of the panel used in chapter three. I also thank Elizabeth Kopits for her assistance with the Census GIS school district shapefiles. Finally, I thank the National Science Foundation and Russell Sage Foundation for supporting a large portion of this work. All responsibility for errors rests solely with me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

There are few publicly provided goods that command quite the same level of resources—and attention—as education. Expenditure on public education in the United States (at the primary and secondary level) accounted for roughly 4 percent of U.S. GDP and 21 percent of state and local government expenditure in 1999, according to the *Digest of Education Statistics* (2003) and the *Statistical Abstract of the United States* (2003). This level of commitment is not surprising—after all, the educational system is viewed not only as a path to individual self-sufficiency and social mobility, but also as an engine of economic growth, innovation, and competitiveness, a means of promoting civic and democratic values, and an important channel for reducing long-run inequalities in the population.[1]

At the same time, there are few goods as elusive as education. Despite nearly universal agreement over the importance of education in the United States, the nature of its production remains weakly understood. Since the seminal Coleman (1966) report, economists and other social scientists have sought to quantify the relationship between educational outputs and purchased inputs through the estimation of education

---

[1] Hoxby (2003) argues that the average poor child in the United States will receive much more in transfers through the education system over the course of her life (i.e. via federal and state redistribution) than she will through direct transfers (i.e. welfare payments).

production functions. This has not been a straightforward task—given its many roles, there has been little agreement over what the "output" of the educational process is, much less how it should be measured. Complicating matters further is the recognition that many inputs into the production of education—such as community and peer composition, family income, and parental involvement—are largely unobservable or otherwise outside of schools districts' control.

Research in the decades immediately following Coleman tended to focus on the relationship between the *quantity* of school resources, broadly defined, and student outcomes—i.e. the question of whether "money matters" in the production of education.[2] The short answer to this question appeared to be a resounding "no"—that is, the level of school expenditure per pupil alone seems to have little, if any, causal relationship with student test scores, graduation rates, labor market outcomes, or any other relevant measure of student performance.

Since that time, research into the production of education has shifted away from broader measures of school inputs (such as dollars) and toward the specific types, uses, or *quality* of inputs used in the education process. Clearly, an additional dollar of per-pupil expenditure can be put to countless uses—to reduce class sizes, to reduce class sizes for certain students, to recruit more talented teachers, to promote parental involvement, or to update technology or infrastructure (just to name a few examples), all of which may have varying effects on school performance. In identifying those inputs or applications of inputs that do in fact affect student outcomes, this literature has demonstrated that the relevant question for policymakers

[2] See Hanushek (1986, 1996) for a summary of this literature.

concerned about the performance of schools is not *whether* money matters, but *where* it matters most. While money is almost certainly not a sufficient condition for improved student outcomes, it is probably a necessary one. In my assessment of the modern education production function literature, money seems to matter, if spent correctly.

In this dissertation, I present two essays that address changes in both the quality and quantity of resources available to public schools over the past four decades. Specifically, in the first essay (chapter two) I investigate how changing labor market opportunities for women have affected the quality of graduates who have chosen to enter the teaching profession. While it has long been presumed that gender desegregation of the professions has adversely affected the quality of teachers (over three quarters of whom are female), there is surprisingly little evidence measuring the extent to which this is true. In this essay, I combine data from five longitudinal surveys of high school graduates spanning the years 1957 to 2000 to evaluate this claim. I find that while the relative test score ranking of the average new female teacher has fallen only modestly over this period, the likelihood that a female in the top decile of her high school class enters the teaching profession has plummeted.

In the second essay (chapter three), I examine how rising population heterogeneity and income inequality within school districts has affected the level of per-pupil expenditure and the fraction of school aged children enrolled in private schools. Like the nation at large, the populations of school districts in the United States have become significantly more diverse over the past several decades, in (among other dimensions) racial and ethnic background, schooling, and income.

Using a merged panel of school district demographics and financial data for 8,700 unified school districts over the 1970 to 2000 period, I look at the effects of this rising heterogeneity on the support for local public schools. I find that rising within-district income inequality is associated with *greater* per-pupil expenditure, a result consistent with a median voter model in which a lower tax price to the median voter results in greater per-pupil spending. Greater fractionalization in race and educational attainment appears to reduce per-pupil expenditure and increase enrollment in private schools.

These essays contribute to the existing literature in a number of ways. Recent empirical research has suggested that among purchased school inputs, teacher quality—particularly when measured with indicators of verbal and mathematical ability—is one of the few that have been consistently shown to have a positive and statistically significant relationship with student outcomes.[3] A related body of literature shows that college graduates entering the teaching profession in the 1970s and 1980s did not compare favorably with those choosing other professions.[4] There is little evidence, however, on how the verbal and mathematical abilities of those candidates who enter teaching have changed over a more extended period of labor market desegregation. This deficiency has largely been due to a lack of data—there are no datasets currently available that (a) collect data both before and after the great transformation of the labor market that began in the early 1960s, (b) contain measures

[3] See Wayne and Youngs (2003), Goldhaber (2002), or Rice (2003) for a survey of this research.

[4] See Manski (1985) and Hanushek and Pace (1995), for examples.

that can be considered reliable measures of teacher "quality," and (c) identify actual teachers and non-teachers with a relatively high degree of certainty. In chapter two of this dissertation, I address this problem by combining five different longitudinal surveys—two of which date prior to 1965—which track five cohorts of high school graduates through college and into the workforce (these datasets are described in detail in Appendix A). Each of these surveys contain measures of math and verbal ability, which allow me to place teachers and non-teachers in the skill distribution of their high school cohort (a population whose composition has remained relatively constant over the past forty years, particularly when compared with the population of college graduates). I am then able to assess how the propensity for women with high test score rankings—those individuals which research suggests would make the most effective teachers—to enter the teaching profession has changed across the five cohorts.

The second essay of this dissertation contributes to several strands of literature. First, a number of theoretical papers (Fernandez and Rogerson (1996) and Epple and Romano (1996) among them) have suggested that rising income inequality may be harmful to school spending. They characterize this problem as "the ends against the middle," where high-income households oppose spending on public schools because of their high demand for private schooling, and low-income households oppose spending because of their desire for lower taxes. Given these results, one might suspect that the persistent rise in income inequality observed over the 1970 to 2000 period has had a negative impact on the level of resources devoted to public education. By merging demographic data at the school district level from the 1970, 1980, 1990 and 2000 U.S. Census with financial data from the Census of Governments, I am able to observe how rising income inequality within school districts has been associated with per-pupil spending in practice. This necessitates the computation of within-district measures of income inequality for all school districts in all four census years, which we accomplish by using a novel maximum likelihood procedure to fit a distribution to grouped Census income data in each district (a methodology described in Appendix B). This chapter also contributes to a broader empirical literature that has looked at the relationship between population heterogeneity and public finance in general. This literature has typically found that greater heterogeneity is associated with reduced support for public programs.[5] While my results show that greater fractionalization in race and schooling within school districts correspond with lower per-pupil spending and higher rates of private school enrollment tend to support these findings, I find that rising heterogeneity in income may actually increase spending on public education. This may arise when greater heterogeneity in income represents a decrease in the tax price to the median voter.

This dissertation is organized as follows. In chapter two I begin with a description of changes in the labor market since 1960, and a discussion of some of the ways in which these changes may have affected the quality of incoming teachers (Section 2.2). The potential consequences of falling teacher quality are then considered through a review of the existing literature on the relationship between measured teacher characteristics and student performance (Section 2.3). After

[5] For example, Alesina, Baqir, and Easterly (1999) find that greater ethnic fractionalization within U.S. cities and counties is associated with lower spending on public infrastructure.

describing our methodology and data (Section 2.4), I present the results of a series of econometric estimates of the relationship between relative verbal and mathematical skills and entry into teaching across five cohorts of female high school graduates over the 1960 – 2000 period (Section 2.5). Allowing for the possibility that changes in labor market desegregation may have affected the quality of new *male* teachers, I also take a brief look at five cohorts of male high school graduates over the same period (Section 2.6).

Chapter three begins with the argument that despite improved opportunities to migrate between jurisdictions, local communities have remained quite heterogeneous in practice, a claim supported by recent empirical work. Given that perfect stratification of households across communities is rare, I explore some of the ways in which within-district population heterogeneity might affect support for public schools by presenting a simple median voter model of school spending. The model is extended to allow for heterogeneity across several dimensions, including income (Section 3.2). Our empirical strategy and data are described in Section 3.3. Then, in Section 3.4, I present OLS estimates measuring the effect of rising within-district income inequality and race and schooling heterogeneity on local school revenues per pupil and rates of private schooling. Allowing for the possibility that the local income distribution is endogenous to school spending, I also present instrumental variables estimates of the per-pupil revenues regressions, which I find are quite close to those produced using OLS. Chapter four concludes.

CHAPTER TWO

CHANGING LABOR MARKET OPPORTUNITIES FOR WOMEN AND THE QUALITY OF TEACHERS, 1957-2000

"The quality of teachers has been declining for decades, and no one wants to talk about it… We need to find a more powerful means to attract the most promising candidates to the teaching profession." - Harold O. Levy, chancellor of the New York City Public Schools in "Why the Best Don't Teach" (New York Times, September 9, 2000).

2.1 Introduction

Teacher shortages and concerns over the quality of the teaching force have become perennial issues in the United States. With each passing year, school officials bemoan their inability to attract top candidates to teaching, and the debate over how best to attract and retain talented, better-qualified teachers seems to intensify. One popular explanation for this mounting frustration points to the remarkable gender desegregation of the labor market since 1960.[6] Schools that once found a captive labor pool in educated women are today forced to compete with more lucrative professions, with the best and brightest believed to be least likely to remain in teaching. This conjecture has been the impetus behind many suggested policy measures—salary increases, or relaxed testing and course requirements, to name a

---

[6] Temin (2002), for example, argues that our history of widening opportunities for women since 1960 has created conditions for the existence of multiple teacher pay-quality equilibria, and that U.S. schools are currently stuck in a low pay-low teacher quality equilibrium. Prior to 1960, he argues, multiple equilibria were not possible as women at all levels of ability were confined to a small number of occupations.

few—all intended to increase the attractiveness of teaching relative to other occupations.

Frustration over the quality of the teaching force comes amidst a growing body of evidence that shows that certain teacher characteristics—like their academic ability—are actually quite important for student achievement.[7] This evidence is particularly remarkable given the current lack of agreement over the importance of other measured inputs into the production of education, such as per-pupil expenditure and class size.[8] Recognizing the apparent importance of attracting quality teachers, the federal government in 2001 passed Title II of the Elementary and Secondary Education Act (ESEA)—a $2.9 billion program specifically targeted toward improving teacher quality.[9]

The hypothesis that gender desegregation of the professions has reduced the number of high-ability women choosing to teach is so frequently cited that it has virtually earned the status of common knowledge. However, there is surprisingly little empirical evidence on the extent to which occupational desegregation has in fact

---

[7] See, for example, Hanushek (1970, 1971), Ehrenberg and Brewer (1994, 1995), Ferguson and Ladd (1996), and Hanushek, Kain, and Rivkin (2000). Wayne and Youngs (2003) and Goldhaber (2002) review recent literature on the relationship between teacher characteristics and student achievement.

[8] Hanushek (1986, 1996) and Krueger (2002) provide contrasting views of this issue.

[9] In addition, local education agencies receiving money under Title I of ESEA must (beginning with the 2002-2003 school year) ensure that all teachers hired and supported using ESEA funds are "highly qualified," a criteria based largely on educational attainment and subject matter knowledge (demonstrated through state-administered tests—or, in the case of secondary teachers, through advanced certification or completion of an undergraduate or graduate degree in a specific subject matter).

affected teacher quality over time. This lack of evidence was noted in the conclusion to a recent paper by Podgursky, Monroe, and Watson (2001), who state:

> Economists have hypothesized a secular decline in teacher quality as a consequence of rising non-teaching earnings and job opportunities for high ability women. In this view, public schools benefited from the occupational crowding of women into teaching profession. Unfortunately, time-series data on teacher quality are not available to directly test this hypothesis.

In this chapter, we combine data from five longitudinal surveys of high school students spanning four decades (1957 to 2000), to see how the propensity for women with high verbal and mathematical skills (relative to their high school cohort) to teach has changed over time. Inclusion of data prior to 1970 allows us the opportunity to provide evidence as to how this relationship between academic ability and entry into teaching has changed over a period of vast gender desegregation of occupations.[10]

Our results show that examination of the entire distribution of teacher quality and its changes over time tells a much richer story than one could tell from measures of central tendency alone. While we detect only a modest decline in the *average* test score ranking of female entrants into the teaching profession over this period, the likelihood that a female near the *top* of the test score distribution—presumably the type of individual most likely to benefit from new opportunities in the labor market—has plummeted. We find that the probability a female from the top decile of her high

[10] We will be studying changes in the reduced-form relationship between academic ability and entry into teaching without explicitly investigating the channels through which this relationship may have been shaped. In other words, we will be observing the outcome of both demand and supply decisions in the market for teachers. Consequently, our results here should be interpreted as descriptive evidence on the academic ability of those individuals who actually identified themselves as teachers during this period. In a recent paper, Bacolod (2003) relates changes in relative teacher salaries to changes in the quantity and quality of teachers over time.

school class enters teaching (by approximately age twenty-six) has fallen from 17 percent in 1964 to under 5 percent in 1992, or just under 8 percent in 2000. While the probability that a female with at least a high school diploma entered teaching fell over the entire distribution over this period (with a slight increase as enrollment rose in the 1990s) the decline was much more pronounced at the top of the distribution.

We organize this chapter as follows. In Section 2.2, we begin with a discussion of changes in the labor market since 1960, and consider some of the ways in which these changes may have affected the quality of incoming teachers. Section 2.3 describes some of the existing research in the economics literature on teacher quality, discusses how other researchers have approached the rather nebulous measure of teacher "quality," and explains how our approach will differ from that of previous research. Section 2.4 describes our methodology and data. Section 2.5 presents our results on changes in the ability composition of new female teachers over time, and the results of a simple econometric model relating the academic ability of females to entry into teaching. We also take a brief look at the sample of male teachers in our data. In Section 2.6, we make some concluding remarks and suggestions for future research.

2.2 Background

As is well known, women's relationship with the labor market fundamentally changed during the forty-year period considered in this chapter. Labor force participation among women aged 25-34 nearly doubled between 1964 and 1992; at the same time the fraction of women 25-34 with (four-year) college degrees tripled (see Table 2.1). By comparison, the fraction of college educated young men increased by only 50 percent over the same period. These trends were likely aided in part by key legislative movements of the early 1960's and 70's that altered significantly the landscape of occupational opportunity for women.[11]

Some occupations also became considerably less gender segregated over this period.[12] Table 2.1 shows an index of gender representation in two professions—medicine and law—calculated as the percent female in each occupation divided by the overall female share of the labor force, using data on men and women aged 25-34 in the March Current Population Survey (CPS), 1968-1996. A value less than one indicates that women are *underrepresented* in the occupation (within this age range), relative to their representation in the labor force.[13] In the late 1960's this index ranged from 0.20 to 0.33 in medicine, and was virtually zero in law (0.08). In less than thirty years, the value of this index for both occupations had nearly reached one (0.79 and 0.94, respectively).

Despite striking shifts in the gender composition of professions such as medicine and law, the gender composition of new teachers has remained roughly constant. Women continue to dominate the teaching profession, comprising roughly 70 to 75 percent of teachers aged 25-34—a fraction virtually unchanged since 1964

[11] The Equal Pay Act of 1963 and Title VII of the Civil Rights Act of 1964 were among the most notable, replacing "protective labor laws" and explicitly outlawing employer discrimination on the basis of sex. See Lloyd and Niemi (1979) for a discussion.

[12] This is well documented in Blau, Simpson, and Anderson (1998), Beller (1992), and elsewhere.

[13] This and other methods of measuring occupational segregation and concentration are discussed in Siltanen, et. al. (1995).

(Table 2.1). The proportion is higher for elementary teachers (about 83 percent) than secondary teachers (50 percent), though the fraction has been steadily falling for elementary and rising for secondary teachers within this age range.

While teaching remains a predominantly female occupation, the occupation itself has significantly diminished in importance as a career path for female college graduates. Nowhere is this so evident as in the plummeting percentage of (working) female graduates who identify themselves as teachers (Figure 2.1). According to data from the March CPS, in 1964 over *half* of working female college graduates were teachers—by 1996, this percentage had fallen below 15 percent. While this drop in the fraction of graduates choosing to teach can be attributed largely to changes in the denominator (the enormous rise in college completion among women), one thing in this picture is clear: conditional on working, of those women who acquired a college education in the 1960's, most went into teaching; of those completing college today, most do not. This raises the question—how has the composition of those women who *do* choose to teach changed over time?

Occupational desegregation and the movement of females into high-skill professions such as medicine and law do not necessarily imply a reduction in the pool of talent available for teaching, if the pool of college graduates is expanding. While Table 2.1 indicates a dramatic increase in the pool of female college graduates, the growth in teachers over the same period was not nearly as fast. The total stock of teachers rose by only 89 percent between 1960 and 1996, rising the most between 1960 and 1970 (46 percent) to accommodate an enrollment boom, and rising slowly thereafter. Meanwhile, the fraction of young women with degrees tripled, with the most talented more likely than ever to earn a degree. Figure 2.2 shows college completion rates among five cohorts of female high school graduates, conditional on a high school standardized test score decile.[14] While in 1964 only 33 percent of women in the top three deciles completed college, by 2000 75 percent of women scoring in these deciles were finishing college. It remains to be seen whether this influx of talented females allowed the teaching profession to maintain a certain level of quality, or whether these new graduates increasingly selected into occupations other than teaching.

2.3 Prior Literature on Teacher Quality

Concern over teacher quality is not new. Weaver (1983) recounts a nearly 100-year history in the United States of attempts to raise teacher standards. Much of the research on teacher quality in recent years has focused on 1) what "quality" exactly refers to, and how it should be measured[15], 2) how those entering teaching compare along various dimensions to their non-teacher peers, 3) whether teacher characteristics are in fact important for student achievement, and 4) how best to raise the quality of the teaching force without drastically reducing the supply of teachers or the diversity of the profession.

[14] The data used in Figure 2.2 will be described in greater detail in Section 2.4. For now, this figure is presented solely for illustrative purposes.

[15] Interest in this question has been considerably renewed with the ESEA ("No Child Left Behind") requirement that teachers be "highly qualified." Debate over what this term means, and whether or not federal definitions of "highly qualified" teachers are in fact indicators of "quality," continues.

Ideally, teacher quality would be measured as a multi-dimensional vector of all teacher characteristics that are positively associated with outputs of the educational process. This vector likely includes many attributes—like patience, creativity, or communication skills, for example—that are unobservable or otherwise difficult to measure. Recognizing the shortcomings of the approach, researchers have focused instead on such measurable characteristics as degree attainment, certification status, teaching experience, school selectivity and test scores in this vector of teacher "quality." Others, like Loeb and Page (2000), Lakdawalla (2001), and Stoddard (2003) interpret relative teacher salaries as indicators of teacher quality. Of course, these latter interpretations depend on a willingness to accept relative wages as a measure of teacher quality—as some authors have pointed out, the relationship between teacher salaries and teacher quality (and student achievement) is a tenuous one.[16]

In this paper, we use a measure of academic ability—scores from a standardized test administered in high school—as our measure of "quality." We acknowledge the obvious limitations of using a one-dimensional measure of quality—to think that a single test score can capture all of the factors that make for an effective teacher would be naïve at best—but it is hard to believe that measures, say, of a teacher's verbal and mathematical ability are *not* an important dimension of teacher quality. After all, these are tests of skills that teachers are expected to cultivate in their

[16] See Hanushek, Kain, and Rivkin (1999), Manski (1987), and Ballou and Podgursky (1997). Loeb and Page (2000) argue that cross-sectional studies of the relationship between teacher salaries and student achievement that do not take into account alternative wage opportunities are misspecified. They find that once relative wages are controlled for, that teacher salaries are modestly related to student outcomes.

own students. And there is a growing literature that finds that such measures—particularly teachers' verbal scores—are among the most important determinants of student achievement.[17] Ehrenberg and Brewer (1995), for example, find that a one-half standard deviation increase in the verbal aptitude score of white female teachers would have raised the synthetic gain scores of white elementary students in the 1966 Coleman Report data by roughly 4 to 8.5 percent. Likewise, Ferguson and Ladd (1996) find that a one standard deviation increase in teachers ACT composite scores (in the state of Alabama) would have resulted in a 0.10 standard deviation increase in student reading scores from 3rd to 4th grade (comparable to about one-half the black-white test score gap in urban areas during this time period). What a teacher's test score measures is less clear, but this literature has illustrated that these scores do indeed capture *something* (whether specific skills or general intelligence) that is important in explaining the academic achievement of their students.

Discovery of the apparent importance of teachers' academic ability in the production of education comes at the same time a number of other studies find that the academic ability of teachers and aspiring teachers did not compare favorably to that of their peers in the 1970's and 80's. These conclusions have often been made through comparison of the mean SAT or ACT scores of examinees intending to major in education to those who do not. Weaver (1983), for example, reports that the average prospective education major ranked at about the 37th percentile of all SAT verbal test takers in 1972, a ranking that remained unchanged through 1980. Ballou and

[17] See, for example, Hanushek (1970, 1971), Ferguson and Ladd (1996), Ferguson (1998), Ehrenberg and Brewer (1994, 1995), Hanushek, Kain, and Rivkin (2000), Wayne and Youngs (2003) and Goldhaber (2002).

Podgursky (1997) find that the average SAT score of prospective education majors fell at about the 45th percentile, but that this ranking improved considerably by 1992. While these comparisons are informative, they may be misleading as not all SAT/ACT takers attend college, and those who do may change majors, or may never enter the teaching force. These studies also tend to neglect other potentially interesting aspects of the distribution of teacher quality, beyond the mean.

Vance and Schlechty (1982) were one of the first to make use of a longitudinal study to compare the academic ability of teachers and non-teachers. The availability of longitudinal data permitted the authors to actually follow students into the workforce, avoiding the problems inherent in using SAT scores of intended education majors.[18] Their tabulations on college graduates from the National Longitudinal Study of the High School Class of 1972 (NLS-72) indicate that teachers identified in 1979—particularly those who had expressed an intention to continue teaching—came disproportionately from the bottom two quintiles of the SAT score distribution.

Several other cross-sectional studies of college graduates have found a negative relationship between academic ability and the likelihood of entering teaching. Manski (1987), in a test of how increased teacher salaries might impact the quality of the teaching force, found a negative, statistically significant relationship between SAT/ACT scores and entry into teaching among working college graduates in the NLS-72. Hanushek and Pace (1995) and Vegas, Murnane and Willett (2001) obtain

[18] Gitomer and Latham (1999) are able to link SAT and ACT scores to practicing teachers by linking test score data to prospective teachers who elected to take the Praxis I or II licensing exam. Unfortunately, this data is only available for those who took the licensing exam between 1994 and 1997.

similar results among college graduates using data from High School and Beyond. Podgursky, Monroe and Watson (2001) also find the same relationship in Missouri state administrative data.

The picture emerging from this literature is that the test scores of college graduates choosing teaching as a profession during the 1970's and 80's did not compare favorably to those of their peers. In light of the extended trend of occupational desegregation described in Section 2.2, however, we would like to know whether or not this relationship between academic ability and entry into teaching has in fact *worsened* over time. The literature on this question is quite limited, and virtually all papers to date have considered only college graduates over a fairly short time period.[19] Murnane, et. al. (1991) and Bacolod (2003) are among the few to address the question of changes in teacher quality over time. Each use college graduates from the National Longitudinal Surveys of Young Men, Women, and Youth and find that the percentage of graduates (of any IQ) entering teaching fell over the period 1967-1989, with a greater decline among those with high IQ scores.

Here, we hope to improve on the existing literature by bringing together for comparison many of the datasets used in the cross-sectional studies mentioned above (and incorporating some older ones), measuring teacher quality with a variable shown to have important effects on student achievement, and by taking care to keep the underlying sample constant over time. Our analysis will cover a significantly longer

---

[19] Pavalko (1970), using the Wisconsin Longitudinal Study, is an exception. Dividing the high school class of 1957 into three ability groups, he finds that at that time in Wisconsin, teachers were drawn disproportionately from the *higher* third of the IQ distribution.

time period than that considered by Murnane et. al. and Bacolod (encompassing a period of more rapidly changing labor market characteristics), and in most cases our sample of graduates is significantly larger. Our methodology and data will be described in further detail in Section 2.4.

2.4 Methodology and Data

2.4.1 Methodology

To provide a picture of the changing relationship between academic ability and entry into teaching over time, we combine data from five longitudinal surveys, two of which are comprised of individuals who graduated from high school prior to 1965. We first look (in Section 2.5.1) at how the test score ranking of the average new female teacher has changed over time, and examine changes in the overall composition of new female teachers. We then estimate for individuals in each high school cohort the likelihood of entering the teaching profession (by approximately age twenty-six), conditional on measured academic ability. Each of these methods provides a perspective on how school districts' ability to attract high test-scoring individual into the teaching profession has changed over time. While our primary focus is on women (Section 2.5.1 and 2.5.2), we also perform the same analysis on our sample of men (Section 2.5.3).

The econometric model we estimate is as follows:

(1) $\Pr(Y_{it} = 1 | x_{it}, z_{it}) = f(\alpha_t + \beta_t x_{it} + \gamma_t' z_{it})$,

where Y_{it} is an indicator equal to one if person i (in cohort t) identifies herself as a teacher, f is the logistic function and x_{it} is a measure of person i's academic ability. In some specifications, we add a vector of student and/or parental characteristics (z_{it}), but our primary interest here is in the "total relationship" between test score and entry into teaching, not necessarily the "partial relationship." Loosely, we are interested in whether or not there have been changes in the gradient $\partial f / \partial x$—a measure of the relationship between underlying academic ability and entry into teaching in a given year—that would have affected the distribution of ability among new teachers.

In each cohort, we include those individuals with at least a *high school* diploma as our sample. As we described in the introduction, college completion among females tripled over our sample period. Were we to use college graduates as a reference group, our interpretation of changes in the gradient $\partial f / \partial x$ could be biased if college completion has increased differentially across ability groups over time.[20] A simple example may help to illustrate this idea. Suppose the population is fixed at 500 "low ability" and 500 "high ability" individuals. Each year, school districts recruit 60 low ability and 40 high ability individuals into the teaching profession (i.e. assume the skill distribution of teachers is unchanged over time). In the first year, assume 100 low ability and 300 high ability individuals graduate from college (a total of 400, with 25 percent of all graduates of low ability and 75 percent of high ability). Based on our assumption of the skill distribution of new teachers, and assuming that only college graduates can become teachers, 60 percent of low ability college graduates became teachers (60 of 100) and 13.3 percent of high ability college graduates became

[20] Vegas, Murnane, and Willett (2001) is one of the few studies that begins with a sample of high school graduates. However, like most of the empirical work described in Section 2.3, they do not look at changes in teacher quality over time, but rather the impact of certain individual characteristics on various sequential decisions (high school completion, college graduation, and finally entry into teaching).

teachers (40 of 300)—i.e. a high ability college graduate is 46.7 percentage points less likely to become a teacher than a low ability college graduate (loosely, our gradient $\partial f / \partial x$ for year one). If we normalize this marginal effect by the overall mean (the 25 percent of all college graduates who become teachers in year one), then the elasticity of entry to teaching with respect to ability is -1.87 (-46.7/25). Now, in year two, assume 200 low- and 350 high-ability individuals complete college (a total of 550, with 36 percent of college graduates of low ability and 64 percent of high ability). In this year, our proportions change to 30 percent of low ability graduates and 11.4 percent of high ability graduates entering the teaching profession—a difference of only 18.6 percentage points. With an overall fraction of 16.6 percent of college graduates entering teaching, this implies an elasticity of -1.02. While the composition of the teaching force has not changed (it remains 60/40), a comparison of the marginal effects (or elasticities) across years might lead us to believe that the negative relationship between ability and teaching has weakened from year one to year two.[21] We clearly would not want to interpret these effects of a changing population composition as changes in the quality of teachers.

A look at Figure 2.2 suggests that increases in college completion rates were not in fact constant across the ability distribution. While the absolute increase in the fraction of females completing college was greater for high scoring females, the increase was proportionately larger for students at the bottom of the distribution—low

[21] It can be shown in this example that when the percentage increase in college completion rates is equal across ability groups that the elasticity of entry into teaching with respect to ability remains constant (though the marginal effect will still be affected).

scoring females were over five times more likely to complete college in 2000 vs. 1964, compared to twice as likely for high scoring females. Assuming that $\partial f / \partial x$ is negative among college graduates (as was suggested by the cross-sectional studies discussed in Section 2.3), one might estimate $\partial f / \partial x$ to be *less negative* among college graduates in later years simply as a result of the influx of more low-ability students into the sample. Similarly, one could overestimate a change in $\partial f / \partial x$ if individuals with higher test scores were more represented in later year samples. Restriction of the sample to *working* college graduates could add yet another layer of this kind of sample selection bias, by the same reasoning.

We avoid these potential problems by using a sample of all (working and non-working) high school graduates—a sample we believe to be much more stable over this period. While the high school graduation rate among women also rose (from 68 percent to 86 percent in Table 2.1—an increase driven largely by gains in high school completion among black women), the growth was not nearly as dramatic.[22] Admittedly, any change in the ability composition of high school graduates could bias our results in a similar manner. Unfortunately, there is no way using our data to detect the extent of change in the composition of high school graduates, but we believe that any bias induced by such changes is smaller than that potentially induced by limiting our analysis to college graduates.

[22] In some specifications of (1), we include race dummies and report predicted probabilities for white women only.

2.4.2 Data

Our data consists of five separate cohorts of high school graduates from five longitudinal surveys—the Wisconsin Longitudinal Study (WLS) for the class of 1957, Project Talent (Talent) for the classes of 1960-64, the National Longitudinal Study of the High School Class of 1972 (NLS-72), the sophomore cohort of High School and Beyond (HSB) for the class of 1982, and the National Education Longitudinal Study of 1988 (NELS) for the class of 1992. These five studies are alike in that they each include a detailed survey of the student during their senior year, all require students to participate in a battery of aptitude tests, and all conduct numerous follow-up surveys after high school. Together, these surveys provide us five distinct pictures of the relationship between academic ability (as measured in high school) and occupational choice among high school graduates over four decades.[23]

In each survey, we take as our sample women who have graduated from high school, have a test score available, and responded to a selected follow-up survey. To allow for comparison across surveys and avoid life-cycle effects on occupation choice, we choose the follow-up survey conducted when most respondents were approximately twenty-six years of age. These follow-up surveys were conducted in

[23] Detailed information about these longitudinal studies can be found in Appendix A. The Department of Education surveys (NLS-72, HSB, and NELS) intentionally oversampled certain minority or socioeconomic groups. These datasets each provided sampling weights, which we use where appropriate. Unweighted results are provided in section five of Appendix A. Note, the NELS data was recently added to this chapter, upon its release in Spring 2003. Consequently, less scrutiny has been given to its accuracy. Appropriate caution should be used when interpreting results from this data.

1964 (for WLS respondents), 1971-74 (Talent), 1979 (NLS-72), 1992 (HSB), and 2000 (NELS). Descriptive statistics for each sample cohort are provided in Table 2.2.[24]

Neither of our 1960's-era datasets is ideal. WLS, consisting exclusively of white high school graduates in Wisconsin is clearly not a nationally representative sample, but it is large—over 4,600 women participated in the 1964 follow up. By contrast, Project Talent is designed to be nationally representative, but publicly released data contains one-third as many observations as WLS. As a test of the comparability of the WLS data, we contrasted the 1936-42 birth cohort (approximately the same cohort as our WLS respondents) of Wisconsin-born females with white females born outside of Wisconsin in the 1970 1% and the 1980 and 1990 5% Census Public Use Micro Samples (PUMS; see Table 2.3). Nothing in Table 2.3 suggests that white Wisconsin women of this cohort look markedly different from white women of the same birth cohort born outside of Wisconsin.[25] In the most relevant year (1970), when this cohort was approximately thirty years of age, we see little difference in the educational attainment, labor force participation, or fraction teaching between the two groups, with the exception being the fraction earning masters' degrees, which is almost one percentage point higher among females born outside of Wisconsin. Earnings and wages are slightly higher among females outside of Wisconsin (by about

[24] Not many observations were lost in restricting the sample to follow-up participants with test scores. The lowest rate of response among female high school graduates (with test scores available) across the four surveys was 86.5 percent (WLS)—see Table 2.2.

[25] Keep in mind that large samples as those we have used in Table 2.3 will tend to find statistically significant differences, even when the difference itself is quite small.

7 to 8 percent) in 1970, likely explained in part by their higher average educational attainment and greater tendency to reside in MSAs. Wider differences between these cohorts appear later in life (1980 and 1990), particularly in the fraction completing higher degrees. While we are unable to claim WLS to be a nationally representative sample, we can have some confidence that our sample—at least at the age when most decisions about entering the teaching profession are made—looks much like the larger population of white females in the U.S. With the inclusion of Project Talent, we hope to have a fairly representative picture of female high school graduates in the late 1950's and early 60's.

Among other variables, we have for each student a measure of academic ability, and a self-reported occupation at approximately age twenty-six. As our measure of academic ability, we combine the raw scores from the math and verbal portions of a test battery administered to each of the five cohorts during their senior year.[26] We then compute a centile ranking and standardized score based on each student's placement in the distribution of all high school graduates of the same gender (suggesting that our ability measure should be interpreted as a measure of *relative*—

[26] See Appendix A, Section A.3 for information about the specific content of each test battery, which does differ somewhat across surveys. The exam given to the Wisconsin Longitudinal Study participants (the Henmon-Nelson Test of Mental Ability) differed significantly from the others in that it was designed as an IQ test, and thus may measure different aptitudes. The High School and Beyond cohort completed the test battery during the sophomore and senior years—for comparability, we use the raw score from the test battery administered during the senior year.

not absolute—math and verbal aptitude). Table 2.2 provides the mean centile rank and standard score of all high school graduates responding to each follow-up survey.[27]

While the tests of math and verbal abilities administered to each cohort do differ from each other, most are quite similar in content to other standardized tests administered in high school, like the SAT or ACT. Indeed, among those students for whom we have both a test score and an SAT or ACT score, the correlation between these scores is quite high (0.850 for the NLS SAT scores, and 0.841 and 0.863 for the NLS and HSB ACT scores, respectively). As an additional, admittedly rough test of the comparability of these test scores across surveys, we used logistic regression analysis to examine the relationship between centile rankings or standardized scores and entry into medicine among men across four of the five surveys.[28] With the fraction of male high school graduates eventually entering medicine remaining roughly constant since 1970, and having no reason to believe that the relationship between cognitive ability and entry into the medical profession has changed much since 1960, we would expect (if our test scores measure similar aptitudes) to see a fairly consistent relationship between these two variables over time.[29] And as Table

---

[27] Means differ slightly from fifty and zero due to varied response to the follow-up surveys. Note that centile ranks were determined using *all* female high school graduates for whom we have a test score available, not just those who responded to the follow-up survey.

[28] Too few of the NELS participants were classified as physicians to include the NELS in this comparison.

[29] Using data from the March Current Population Survey, the fraction of male high school graduates age 25-34 who report themselves to be physicians or surgeons has remained very close to 0.5—0.6 percent, with no apparent upward or downward trend, over the period 1970-1996.

2.4 shows, this is indeed the case. In all cases, the coefficient on test score is positive and statistically significant (evidence that the qualities measured by these scores are strongly associated with entry into a high-skilled profession such as medicine), and the marginal effect (also shown as a percentage of the sample mean, i.e. as an elasticity) of test score on entry into medicine among men remains roughly constant across the four surveys.

Our occupation variable has a fairly broad definition here and should be interpreted as "profession" or "line of work," as the individuals in our sample report occupations, regardless of whether or not they are currently working. A broader definition is actually quite useful in this context, as it has been suggested by Polachek (1981) and Flyer and Rosen (1997) that women who expect to spend more time out of the labor force may select into occupations like teaching because of their flexibility.[30] If this is indeed the case, comparing teachers to non-teachers in a sample of *working* women may lead to biased results.[31] Note that teachers identified in our data may be

---

[30] In our WLS sample, close to 35 percent of the women identifying themselves as teachers were out of the labor force at the time of the 1964 follow-up survey. Project Talent is unique in that most women out of the labor force did *not* report an occupation. We identified 58 "housewives" in the Project Talent survey who held teaching *certificates*—an indication that they might be teaching, if they were in the labor force at the time of the survey. However for consistency with the other surveys, we did not count these individuals as teachers. Incidentally, 18 of the 58 housewives (31 percent) with teaching certificates were ranked in the top *decile* of the test score distribution.

[31] Most empirical work on teacher quality to date restricts analysis to working women only. One caveat associated with our use of a broader definition of occupation is that it suggests our results on teachers apply to those women who identify themselves as teachers, not necessarily those employed as teachers at the time of the survey. If some low-scoring women report themselves as teachers and are not in the labor force because they could not find work as a teacher, for example, then our results—as a

elementary or secondary teachers, and may work (or have worked) for public or private school districts. Unfortunately, our data does not allow us to make an accurate distinction between these classes of teachers.

Tables 2.5a and 2.5b report the top ten most frequently reported occupations by female high school and college graduates in four of our five surveys. As expected, we find evidence here of significantly less concentration in the occupations held by women in later years, particularly among those with a college education. Whereas 49 percent of female college graduates in the WLS sample were teachers in 1964 (compare to 52 percent in Figure 2.1), only 11.8 percent were teachers in the 1992 HSB follow-up (compare to 15 percent in Figure 2.1). In 1992, female college graduates were more likely to be in management (14 percent) or clerical work (a broad category, 17 percent) than in teaching.

2.5 Results on New Entrants into Teaching

2.5.1 Descriptive Evidence on the Changing Ability of New Female Teachers

Table 2.6 presents some preliminary findings on the female teachers we identified in our five cohorts. While not surprisingly their mean centile rank and standard score lies consistently above that of the average high school graduate (fifty and zero, by definition), we find the centile rank of the average new female teacher falling about 3 percentage points, from the 67^{th} (or 69^{th} if using Project Talent) to the

---

general statement about practicing teachers—would bias the average ability of teachers downward.

64th percentile—about 5.2 percent—over this time period.[32] As centile rankings mask information about the tails of the distribution, we also compared the mean standard scores of teachers across these cohorts. Here, the downward trend in the mean among female teachers is much more stark—a fall from 0.60 standard deviations above the mean female high school graduate (or 0.65 in Talent) to 0.46 from 1964 to 2000—conservatively, a 23 percent drop.

Figure 2.3 takes a closer look at the distribution of new female teachers across decile groups. Perhaps the most striking trend in the composition of new female teachers observed here is the steadily declining share of new female teachers who scored in the top decile of their high school cohort. While in 1964 over 20 percent of young female teachers fell in the top decile of their high school class (over 25 percent when considering Project Talent), this number had fallen to 11 percent by 2000.[33] This drop in the fraction of new female teachers scoring in the top decile can be reconciled with the modest decline in the mean centile rank observed over the 1964 – 2000 period through examination of changes in the other nine deciles. As Figure 2.3 shows, there was a similar decline in the fraction of teachers who scored in the lowest decile of the test score distribution—from 3 percent to under 1 percent by 2000 (with a unexplained rise in 1992). Approximately 31 percent of teachers scored in the 2nd –

---

[32] The magnitude of this decline in the centile rank of the average teacher may in fact be underestimated, if the ability composition of *high school* graduates has changed over time (and if the increase in the high school graduation rate is driven by students at the lower end of the distribution). But again, we are unable to test this hypothesis in this data.

[33] Keep in mind that this is the top decile of the *high school graduate* distribution—a sizable population, not a small 'elite' as the top decile of the college graduate distribution might be considered.

6th deciles in 1964; this proportion grew monotonically over our high school cohorts after 1971-74, with nearly 41 percent of the teachers in our sample falling in these deciles in 2000. The fraction of teachers in the 7th, 8th, and 9th deciles fluctuated over these five cohorts with few discernible trends. It does appear, however, that the fraction of teachers who scored in the 7th (and possibly 9th) deciles of their high school class has been steadily rising over this period (although the jump in the 9th decile appears only in our last cohort).

Our results so far suggest that—while the academic ability of the average new female teacher has remained relatively unchanged over this period relative to the average high school graduate (when comparing mean centile ranks)—teachers are much less likely to come from the top decile of their high school class in 2000 vs. 1964.[34] They also appear to be much less likely to come from the bottom decile of the test score distribution in later years (although the proportion of teachers from this decile is relatively much smaller). We investigate these findings in further detail in Section 2.5.2.

2.5.2 Econometric Evidence on the Changing Propensity to Teach—Females

In the last section, we found that the academic ability of the *average* new female teacher—measured as one's placement in the distribution of female high

---

[34] It should be emphasized again that the changes in test score rankings of new teachers are changes in the *relative* academic ability of new teachers, i.e. relative to other high school graduates. It has been noted that although the relative ability of new female teachers has been falling over time, the absolute skill level of new female teachers may be rising, if parental education is an input into cognitive ability (observe in Table 2.6 the steady increase in average parental education for these teachers). I thank Darius Lakdawalla for being the first to make this observation.

school graduates—has fallen somewhat since 1964. Here, we find that this decline can be largely explained by a marked decrease in the propensity to teach among women at the highest end of the test score distribution—presumably those who stood to benefit most from changes in occupational opportunity over this period.

To investigate how the relationship between academic ability and entry into teaching has changed over time, we estimated equation (1) separately for each cohort, including at first a continuous measure of academic ability (the student's centile ranking or standardized score), and a control for the respondent's age.[35] We then allow test scores in enter in a nonlinear fashion into our model—specifically, we implement decile dummies as our measure of ability. Finally, we included other covariates that might explain differential rates of entry into teaching—most importantly, parental education and race.[36] Unfortunately, the range of additional covariates available in our five datasets varied widely from one dataset to the next, limiting our ability to make consistent comparisons across cohorts.

Tables 2.7a and 2.7b provide coefficient estimates from our initial specifications. Table 2.7a implements centile rankings as our measure of academic

---

[35] At first, we are primarily interested in the total relationship between entry into teaching and academic ability. Other authors, such as Murnane, et. al. (1991) include a cognitive ability measure as one of several regressors, which captures the partial relationship between teaching and ability. I thank an anonymous referee for noting that the total relationship is of foremost interest in this chapter. We include a control for age due to some variation (especially in Project Talent) in the respondents' age when we observed them.

[36] Parental education was not provided in Project Talent. Talent did identify the respondent's race, but there were too few ethnic minorities identifying themselves as teachers to include race dummies in this regression. In regressions where race dummies are included, whites are the omitted group. Of course, race dummies were not relevant in the WLS regressions.

ability; while 2.7b uses standard scores (sampling weights have been used in the NLS-72, HSB, and NELS regressions). The marginal effects of test score ranking on the likelihood of entering teaching—calculated as $\partial f / \partial x_{it}$ in equation (1)—and their respective elasticities (i.e. the marginal effect normalized by the mean of y_{it}) are also computed for a female high school graduate of average age, both at the mean test score and one standard deviation above the mean (or at the 85th centile in Table 2.7a).[37] Standard errors for all coefficients, marginal effects, and elasticities are shown in parentheses.

As one would expect in a sample of graduates with a minimum of a high school degree, we find in Tables 2.7a-b that test scores and entry into teaching have a positive, statistically significant relationship across all five cohorts. What is more interesting in these tables is the substantial weakening of this relationship between ability and entry into teaching since 1964. On both an absolute and elasticity basis, the effect of a one point increase in centile rank (or standard score) on the probability that a female high school graduate enters teaching has fallen markedly over this period, with the decrease much more prominent when standard scores are used as a measure of ability (as would be expected, given our results in Section 2.5.1). As observed in Table 2.7a, the elasticity of a one-centile increase in test score calculated at the mean white female has dropped from 0.022 in 1964 (with a 95% confidence interval of [0.018, 0.026]), to 0.015 in 2000 (with a 95% confidence interval of [0.011,

[37] Elasticities—or, the *percentage* increase in the likelihood of choosing teaching as an occupation for a unit change in test score rank—are more comparable here than marginal effects, as the mean of the dependent variable falls somewhat over the five cohorts. We use elasticities to make all of our cross-cohort comparisons in this chapter.

0.019]), or about a third of its size (compare columns (2) and (7)).[38] These effects are again a bit stronger when using standard scores as our measure of academic ability. Table 2.7b indicates the elasticity of a one standard deviation increase in test score calculated at the mean white female has dropped considerably—from 0.640 (with 95% confidence interval of [0.522, 0.757]) to 0.488 (95% confidence interval of [0.356, 0.620]), or about 0.26 percentage points.[39] The point estimate using Project Talent of 0.79 shows an even greater decline than that suggested by WLS. While given our samples we are unable to reject the hypothesis that these elasticities differ from 1964 to 2000 at the 95% level (we can do so at the 90% level), we nevertheless find the sharp decline in this strength of this relationship between test scores and entry into teaching to be striking.

Results thus far suggest that the positive relationship between academic ability and entry into teaching among female high school graduates has indeed weakened over time. Our stronger results with the use of standard scores also suggest that individuals near the tails of the test score distribution may be particularly important in explaining this trend. For this reason, and to allow for the possibility that test scores may in fact be nonlinearly related to the likelihood of entering teaching, we estimate (1) again for each cohort, but instead implement decile dummies as our measure of ability.

---

[38] This result differs little when the elasticity is computed at the 85th percentile vs. the mean centile.

[39] Note that a movement of one standard score will be a much greater movement (in magnitude) than one centile rank. Thus the marginal effects and elasticities will correspondingly be larger in magnitude when using standard scores. We are more interested here in the *change* in these effects than the magnitude, per se.

Table 2.8 presents the results of this estimation (the lowest decile is the omitted group). We again include a control for the respondent's age, and use sampling weights in all available cases. Here, most of the coefficients on the decile dummies are positive and statistically significant (in this context, statistical significance simply means that being in decile *j* makes one more likely to be a teacher than if she were in the lowest decile).

As marginal effects and elasticities like those provided in Tables 2.7a-b are less meaningful under this specification, we have calculated the (average) predicted probability that a female in decile *j* becomes a teacher, for all five cohorts (see Table 2.9). For females in most deciles the probability of being identified as a teacher fell roughly in half from 1964 to 1992, with a modest rise between 1992 and 2000 (compare columns (1), (4) and (5) in Table 2.9). We find however, much larger drops in this probability for females in the top three deciles from 1964-1992, or the top, 8th, and bottom deciles from 1964-2000 (the introduction of the NELS cohort changes our findings somewhat). Columns (6)-(10) of Table 2.9 normalize these predicted probabilities by the overall fraction of each cohort that we identified as teachers. Values greater than one indicate that the probability that a female high school graduate from that decile becomes a teacher is higher than average; values less than one indicate the opposite. Here these trends are clearer—women in the top decile are much less likely to become teachers, relative to the average, in later years vs. earlier years. The opposite trend is true for deciles near the bottom of the distribution (with the lowest decile being a notable exception).

Figure 2.4 plots these predicted probabilities. This picture illustrates particularly well the nonlinear relationship between test score and entry into teaching, and the comparatively larger reduction in the predicted probability of becoming a teacher among those at the top of the ability distribution.

We repeated the same analysis above with the inclusion of race dummies, parental education and other covariates, like socioeconomic status (where available). The results were virtually unchanged, and are not presented here.[40]

If women near the top of their high school classes are increasingly less likely over this period to enter teaching, what career paths were they pursuing? We tabulated the top ten occupations reported by females ranking in the top decile of their class, in four of our five cohorts (NELS is excluded), and the results are striking, if not unexpected. Whereas close to 20 percent of females in the top decile in 1964 chose teaching as a profession (teaching was the most frequently reported occupation among this group in 1964), only 3.7 percent of top decile females were teaching in 1992. Top scoring women in our 1992 cohort were much more likely to be working as computer specialists (5.9 percent), accountants (6.0 percent), or managers (15.1 percent).[41] Top decile females were almost as likely to be lawyers and judges (3.2 percent) as teachers.

---

[40] In the analog to Figure 2.4, we computed the predicted probability of entering the teaching profession for a *white* female with average parental education, conditional on test score decile.

[41] Because we have not conditioned on labor force participation ("none" and "housewife" are counted as occupations), these tabulations should be unaffected by changes in female labor force participation. Interestingly—even in 1992—the most frequently reported occupation among top decile females was clerical work (17.2 percent).

2.5.3 Evidence on New Male Teachers

An intriguing possible side effect of the gender desegregation of occupations and the movement of talented women into high-cognitive ability occupations is the substitution of high skilled *men* into teaching. As men lost a virtual monopoly on certain professions, finding themselves in competition with capable women for these positions, some may have opted for a career in teaching. While our sample sizes are obviously much smaller for male teachers, we summarize some of our findings on male teachers in Table 2.10.

Our results in Table 2.10 are quite interesting, if only suggestive. Across these five cohorts, we find that the average academic ability of *male* teachers rose from 1964-2000 by 6.6 percent (18.2 percent through 1992), or 28.2 percent (100 percent through 1992) if using standardized scores. This increase in the average ability of male teachers also appears to be driven by those at the top of the ability distribution, as the next panel in Table 2.10 indicates. While (as with women) the probability that a male high school graduate entered teaching fell over this period for most decile groups (again with a slight increase between 1992 and 2000), here the decline in probability is much *less* dramatic for those in the top decile. While most other decile groups saw a decline in the likelihood of entering teaching of 35-75 percent from 1964 to 2000, this reduction was only 29 percent for those in the top decile (the difference is much more stark when comparing HSB to earlier cohorts, but the large fraction of top decile men identified as teachers in HSB compared to adjacent cohorts casts some doubt over this finding).

Again, appropriate caution should be used with the results in Table 2.10. Our sample of male teachers was naturally much smaller than that of females (as few as 62 male teachers in the case of HSB), and the representation of males (particularly top decile males) in the general population of teachers may be small enough that such a trend may not be economically significant. In addition, male teachers are much more likely to be secondary teachers than elementary teachers, suggesting that if anything, these results are most relevant for male secondary level teachers. Nevertheless, we find our results on men to be intriguing, and worthy of additional study with more suitable data.

2.6 Conclusions and Further Research

Despite the wealth of cross-sectional studies that have examined the characteristics of college graduates choosing teaching as a career, there has been little empirical evidence on how these characteristics—particularly the academic ability of these teachers—have changed over time. We believe—in light of the vast occupational desegregation witnessed during the past four decades—that it is of great interest to understand how this desegregation may have affected the recruitment of highly skilled women into teaching.

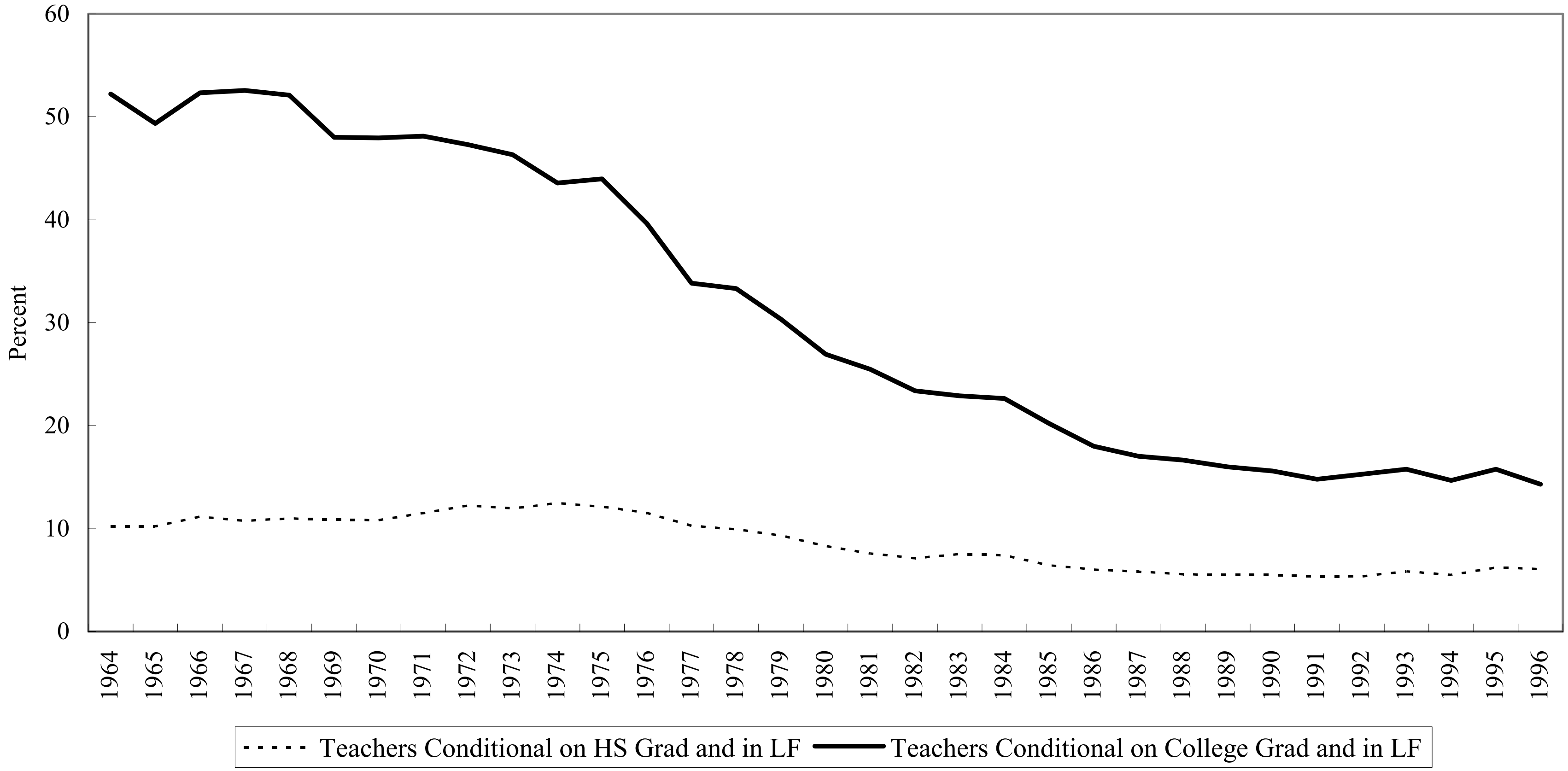
In this chapter, we have found—at least among these five cohorts of high school graduates—a slight but detectable decline in the relative academic ability of the average new female teacher, when ability is measured as one's centile rank on a standardized test of verbal and mathematical aptitude. The magnitude of this decline is greater when measuring ability with standardized scores. In addition, we find that examination of the entire ability distribution of new teachers is more informative than trends in central tendency alone. Consistent with the earlier findings of Murnane et. al. (1991) and Bacolod (2003) on college graduates in the NLSY, we find that the likelihood of entering teaching as a profession by age 25-26 has fallen for all ability levels of high school graduates (with a modest increase between 1992 and 2000, likely due to rising enrollment and class size reduction programs), but this probability has fallen much more for females at the top of the distribution—i.e. females in the top decile of their high school class are much less likely to become teachers in 2000 vs. 1964. While our sample sizes of men are much smaller, we detect the opposite trend among men.

While we are confident that these results are indicative of trends in teacher quality among entering female teachers, and believe that our estimates provide some initial measures of the extent to which females of high verbal and math ability are less likely to enter teaching, we acknowledge some shortcomings inherent in our data. First, while our five datasets all have the advantage of being longitudinal in nature (allowing us to compare respondents to others in their high school class *and* be confident of their ultimate occupational choice), they provide snapshots of only five particular moments in time. Second, the specific exams given to students and the other covariates available varied from one survey to the next. Consequently, we cannot be completely certain that the standardized test scores used here measure exactly the same skills or aptitudes, nor can we consistently control for as many other interesting covariates across the surveys as we would wish. In addition, our test score rankings are only measures of *relative*—not absolute—academic ability, and thus we are unable to make inferences about changes over time in the *level* of verbal and mathematical skills among new teachers. Finally, despite the importance of the academic ability of teachers to student success, we acknowledge again that test scores are but one measure of teacher "quality," and do not claim to have captured in this measure the myriad of important skills, talents, and character traits that make for a great teacher.

If our results can be applied to the wider population of new teachers in the United States, a given student in 2000 (conditional on having a female teacher) could expect to find a teacher who is—on average—of only slightly lower academic ability than a given student in 1964. However, that student is much less likely to find a teacher of the highest academic ability than a student in 1964. Further, given recent research on the sorting of teachers across schools within states and school districts—the likelihood that a student in low income or predominately black school encounters a teacher of the highest academic ability is likely even lower.[42] For the casual observer, these results will surprise few (as suggested by our leading quote from Chancellor Levy). However, we hope that our quantitative results here will serve as an impetus for further research into changes in teacher quality over time (our curious finding on the quality of male teachers in particular calls out for more work and better data). If the significant loss of women in the top decile—those who likely stood to benefit most from occupational desegregation—is indicative of a wider trend, then these findings should be of interest to parents, researchers and policymakers alike.

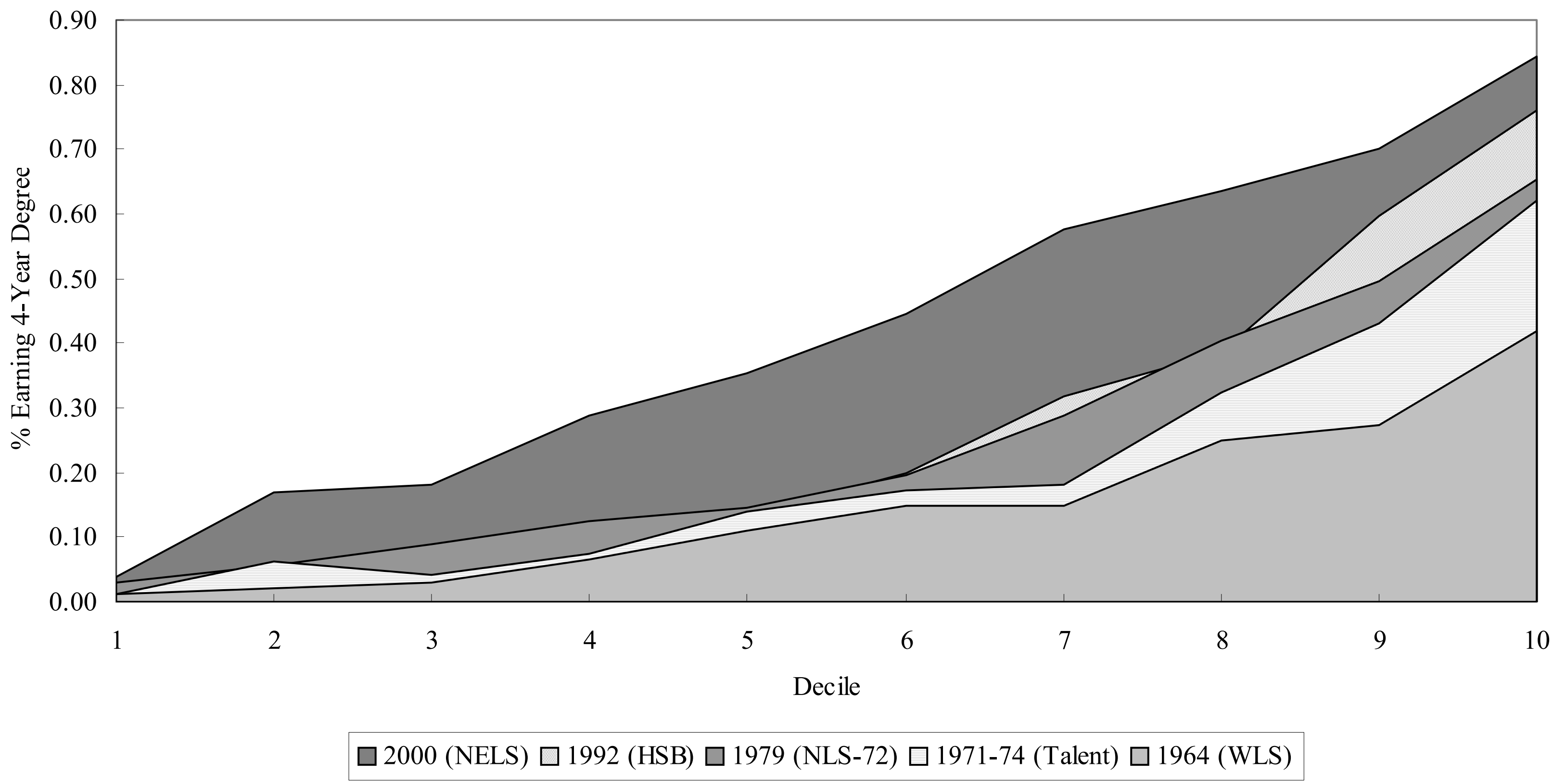
[42] See Lankford, Loeb and Wyckoff (2001).

Figure 2.1: Females 25-34--Percent Teachers



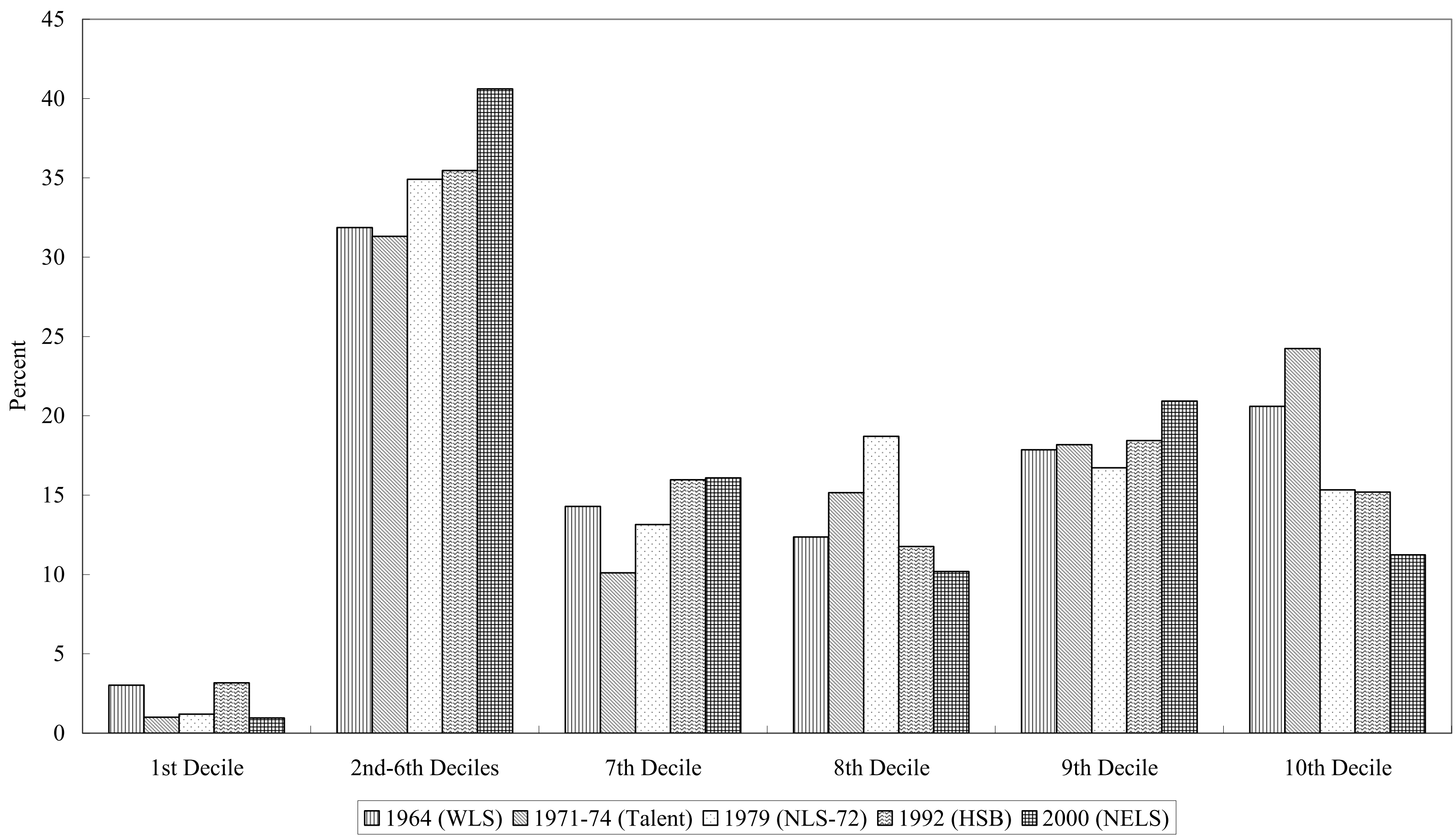
Source: March Current Population Survey, 1964-1996

Figure 2.2: Female College Completion Rates by Test Score Decile, 1964 - 2000



Source: Author's calculations

Figure 2.3: Distribution of Teachers Across Decile Groups 1964-2000



Source: Author’s calculations.

Figure 2.4: Predicted Probability of Becoming a Teacher, by Decile

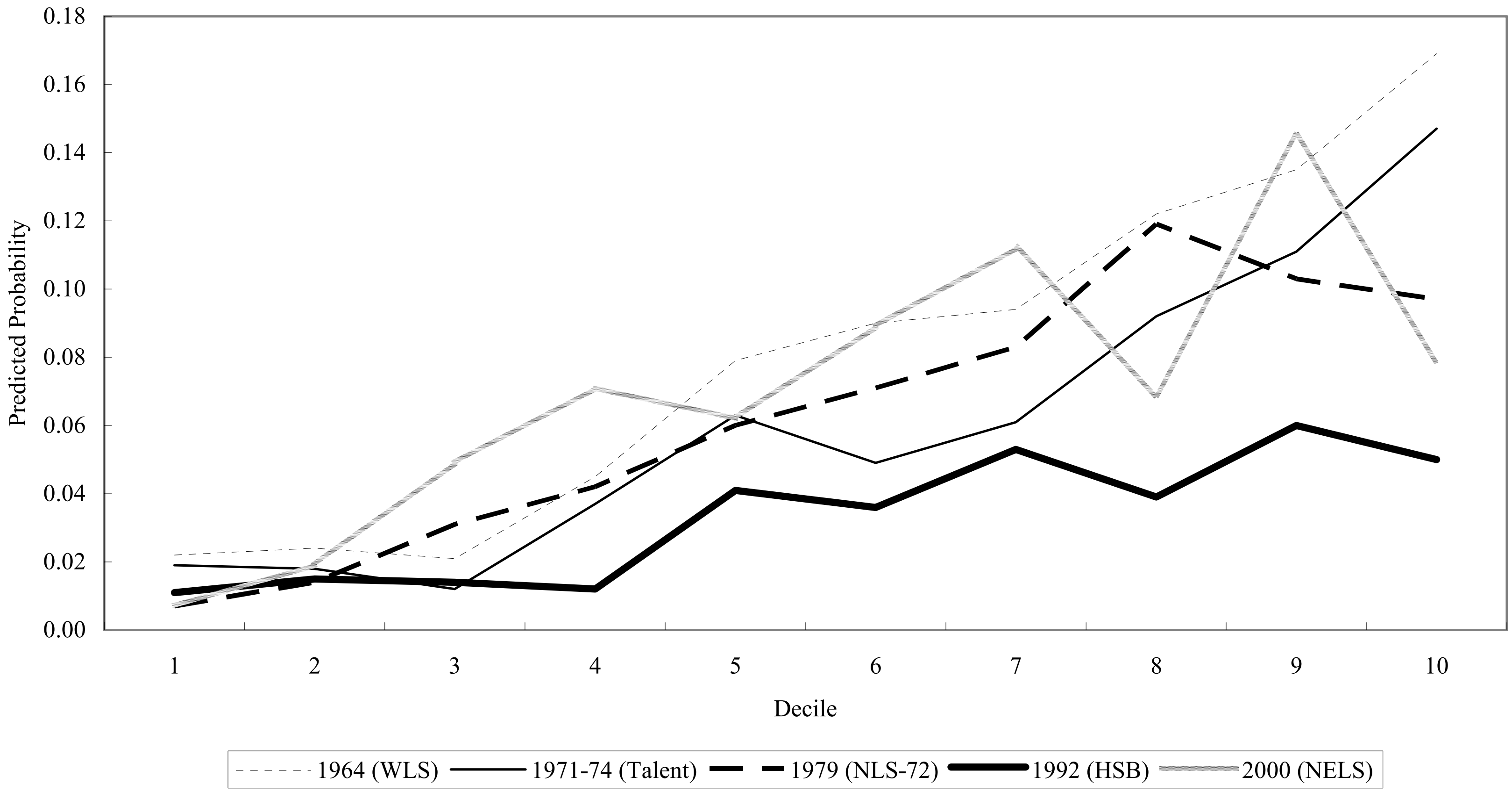


Table 2.1: Labor Market Characteristics, Selected Years, 1964-1996

	1964	1970	1980	1990	1996
% Young Females (age 25-34) with:					
At least a high school degree	67.4	73.3	84.3	86.5	86.5
At least four-year college degree	8.9	12.0	20.7	22.8	26.1
In the labor force	37.2	45.7	65.4	73.2	74.0
Index of gender representation (age 25-34)					
Physicians	0.33	0.20	0.47	0.69	0.79
Lawyers	0.08	0.20	0.44	0.60	0.94
Public elementary teachers (millions)	1.0	1.3	1.4	1.7	1.9
Public secondary teachers (millions)	0.6	1.0	1.1	1.1	1.2
Public school enrollment (millions)	39.0	52.2	49.7	53.8	59.5
% Young teachers (age 25-34) that are female					
All	--	70.0	72.4	74.2	76.3
Primary	--	85.1	84.6	86.0	82.8
Secondary	--	46.6	54.4	53.0	57.2

Source: March Current Population Survey 1964-1996, except teacher counts, taken from the Statistical Abstract of the United States, 2000. First teacher count figure for 1960, first enrollment is from 1965. Gender representation index is calculated from data beginning in 1968.

Table 2.2: Descriptive Statistics for Five Cohorts of Female High School Graduates

	WLS	Talent	NLS72	HSB	NELS
High school graduation year(s)	1957	1960-63	1972	1982	1992
Follow-up survey year(s)	1964	1971-74	1979	1992	2000
Sample size[1]	4,609	1,634	6,751	5,389	4,284
Response rate to follow-up survey (%)[2]	86.5	100.0	88.5	90.4	-
Average age in years at follow-up	25.0	26.8	26.1	27.5	26.1
Race/Ethnicity (%):					
White	--	88.1	82.1	74.5	74.6
Black	--	6.4	9.6	12.5	10.5
Hispanic	--	1.3	3.2	10.9	10.6
Other	--	4.2	5.0	2.1	4.4
% With at least 2 year of college	28.6	29.7	40.6	35.6	59.9
% With at least 4 years of college	14.5	20.5	24.7	26.0	42.3
% Working full or part-time	38.9	48.7	72.1	74.1	86.8
Mean, mother's education (years)	10.3	--	12.4	12.6	13.1
Mean, father's education (years)	9.7	--	12.8	13.0	13.6
% Currently married (HS graduates)	82.9	80.2	60.8	58.5	46.4
% Currently married (college graduates)	65.3	77.8	46.7	50.5	40.3
Mean, centile rank[3,4]					
High school graduates	50.0	50.4	50.4	50.7	50.5
College graduates	73.4	74.9	72.4	76.2	67.4
Mean, standard score					
High school graduates	0.019	0.000	-0.001	0.010	0.000
College graduates	0.820	0.843	0.742	0.914	0.579

[1] Sample consists of those individuals who (a) have completed high school by the given follow-up year, (b) responded to the follow up, and (c) had a valid test score. Sampling weights were used in generating means in NLS-72, HSB, and NELS surveys.

[2] Response rate is calculated as the percent of all high school graduates (with test scores) responding to the follow-up survey. The public use version of Project Talent was designed such that all included students were respondents to the follow-up survey in 1971-74.

[3] Centile ranks are based on students' placement in the distribution of all graduates of the same gender in their high school class. For the NLS-72, HSB, and NELS surveys, centile ranks were assigned using an algorithm incorporating base-year weights.

[4] 'College graduates' refers to completion of a four-year bachelor's degree.

Table 2.3: Comparison of White Female High School Graduates Born in Wisconsin versus Other Women, 1936-1942 Birth Cohorts, 1970, 1980 and 1990 PUMS

	Females, ages 28-34 in 1970, born in:			Females, ages 38-44 in 1980, born in:			Females, ages 48-54 in 1990, born in:		
	WI	All other areas[1]		WI	All other areas[1]		WI	All other areas[1]	
Sample size	1,494	51,174		7,475	271,924		7,939	295,978	
% Unemployed	3.85	3.75		3.55	3.73		2.74	3.14	
% Not in labor force	56.5	56.4		29.2	32.3	**	22.8	26.8	**
Teachers (% among high school grads)	10.1	9.1		6.8	7.1		6.8	7.0	
Elementary teachers	6.6	6.3		4.8	5.4	**	5.6	5.5	
Secondary teachers	3.2	2.5		1.4	1.3		0.6	0.9	**
Bachelor's degree (%)[2]	16.5	16.7		16.8	18.5	**	17.9	19.8	**
Master's degree (%)[2]	1.4	2.2	**	3.9	4.7	**	5.95	8.0	**
Real wage/salary income (1990 $)[3]	15,262	16,463	**	15,345	16,677		18,455	20,897	**
	(8,509)	(8,647)		(9,919)	(10,367)		(13,884)	(15,008)	
Real weekly wage (1990 $)[3]	310.19	333.80	**	307.54	334.64	**	369.35	418.24	**
	(174.90)	(175.52)		(198.49)	(207.48)		(277.02)	(300.42)	
% Living in MSA	59.7	64.9	**	69.2	69.4		56.4	63.5	**
% Living in same state as birth	67.5	57.9	**	65.6	56.4	**	65.4	55.6	**

** Identifies a statistically significant difference in sample means. Person weights were used in the 1990 samples. The 1970 and 1980 samples were designed to be self-weighting.

[1] Females from all other areas may include foreign born.

[2] In the 1970 and 1980 samples, we assume those with at least four years of college have a bachelor's degree, and those with at least six years of college have a master's degree.

[3] In computing mean wages or income, only those earning at least $1000/year and working at least 40 weeks/year were included.

Table 2.4: Logit Estimates of the Probability of Male High School Graduates Entering Medicine

Sample	WLS		NLS-72		HSB	
Dependent variable	Physician in 1975		Physician in 1986		Physician in 1992	
Mean of dependent variable	0.011		0.023		0.008	
Mean age of respondent	43.0		32.9		27.4	
Sample size	4,331		4,358		4,834	
Variable	(1)	(2)	(1)	(2)	(1)	(2)
Constant	-8.075	-5.347	-7.323	-4.760	-8.567	-5.671
	(0.696)	(1.374)	(0.502)	(0.206)	(0.790)	(0.286)
Centile rank	0.052		0.051		0.055	
	(0.008)		(0.006)		(0.010)	
Standardized score		1.374		1.473		1.330
		(0.198)		(0.167)		(0.198)
Marginal effect of test score at:						
Mean	0.0002	0.0065	0.0004	0.0124	0.0002	0.0045
95th percentile or +1 SD	0.0013	0.0249	0.0023	0.0512	0.0011	0.0169
Elasticity of marginal effect at:						
Mean	0.0521	1.3676	0.0506	1.4603	0.0546	1.3250
95th percentile or +1 SD	0.0509	1.3487	0.0486	1.4197	0.0537	1.3125

Standard errors in parentheses. For the WLS, NLS-72 and HSB surveys, we used data from the 1975, 1986 and 1992 follow-up respectively.

Table 2.5: Most Frequently Reported Occupations, Four Cohorts of Females with at Least a High School Degree

A: Females with a High School Degree

	WLS -- 1964		Talent – 1971-74		NLS-72 – 1979		HSB – 1992	
Rank	Occupation	%	Occupation	%	Occupation	%	Occupation	%
1	None/housewife	27.0	Housewife	45.1	Clerical	35.4	Clerical	34.2
2	Clerical	15.9	Secretary	4.4	Service worker	11.9	Service worker	16.3
3	Stenographer/typist	12.8	Elementary teacher	2.7	Manager	8.7	Manager	11.0
4	Teacher	9.7	Registered nurse	2.2	Operative (non transport)	7.1	Sales/worker	5.3
5	Nurse/professional	4.6	Student	1.8	Teacher	6.7	None/housewife	3.7
6	Bookkeeper	2.9	Bookkeeper	1.6	Sales	6.1	Operative (non transport)	3.6
7	Office machine operator	2.1	Factory assembly	1.4	Nurse, dietician, therapist	5.0	Teacher	3.5
8	Waitress	1.9	Clerk/typist	1.2	None/housewife	3.0	Craftsman	2.6
9	Beautician	1.9	None	1.1	Health technician	2.1	Health technician	2.2
10	Medical/dental technician	1.8	Clerical – other	1.0	Craftsman	2.0	Nurse, dietician, therapist	2.2

B: Females with at least a Four-Year Degree

	WLS -- 1964		Talent – 1971-74		NLS-72 – 1979		HSB – 1992	
Rank	Occupation	%	Occupation	%	Occupation	%	Occupation	%
1	Teacher	48.8	Housewife	34.1	Teacher	24.6	Clerical	17.3
2	None/housewife	14.2	Elementary teacher	13.1	Clerical	16.4	Manager	13.9
3	Student	3.9	Teacher – other	3.5	Manager	11.5	Teacher	11.8
4	Social worker	3.4	Student	2.6	Nurse, dietician, therapist	9.2	Sales	8.9
5	Stenographer/typist	3.1	Registered nurse	2.3	Sales	4.9	Service worker	6.7
6	Nurse, professional	3.1	Secondary teacher	2.3	Service worker	4.3	Accountant	5.4
7	Medical/dental technician	2.5	Social worker	2.0	Writer/artist	3.2	Nurse, dietician, therapist	5.1
8	College professor	2.2	Sec. teacher/P.E.	1.5	Health technician	2.6	Computer specialist	4.2
9	Therapist/healer	2.1	Music teacher	1.5	Social worker	2.1	Writer/artist	4.1
10	Musician/music teacher	1.5	Pre-K/K teacher	1.2	Accountant	2.0	Health technician	2.3

Project Talent differed from other surveys in that women who were out of the labor force did not report an occupation. Consequently, many of the women who reported "housewife" in Project Talent would have specified a profession in any of the surveys. Also, Project Talent did not use Census classifications of occupations, as the other surveys did, reducing comparability between occupation titles in this table.

Table 2.6: Descriptive Statistics, New Female Teachers

	WLS	Talent	NLS-72	HSB	NELS
High school graduation year	1957	1960-63	1972	1982	1992
Follow-up survey year	1964	1971-74	1979	1992	2000
Number of teachers[1]	369	99	431	219	302
% of all female high school graduates	8.0	6.1	6.8	4.1	7.0
% of all female college graduates	55.2	30.4	24.4	13.3	16.3
Race of teacher (%)					
White	--	89.9	86.5	80.2	83.8
Black	--	7.1	8.4	10.7	5.3
Hispanic	--	--	1.4	7.9	8.0
Other	--	3.0	3.7	1.1	2.8
% With at least 4 years of college	79.1	98.0	96.2	78.2	92.7
% Currently working	64.9	100.0	88.2	78.0	94.5
% Currently married	66.7	73.9	55.4	60.2	50.1
Mean, mother's education (years)	11.6	--	13.2	13.7	14.0
Mean, father's education (years)	11.3	--	13.5	14.5	14.7
Ability measures:					
Mean centile rank[2]	67.2	69.5	66.4	64.8	63.7
Mean standard score[3]	0.60	0.65	0.55	0.50	0.46

[1]Teachers were identified using the following codes: WLS (#51 – Teachers, NEC"), Project Talent (#400-433 – all elementary and secondary teachers), NLS-72 (#142-145 – "Elementary Kindergarten, Secondary, and Teachers except college and university, nec), and NELS (#23 – "Educator, K-12"). Occupations reported by the sophomore cohort in HSB were not codified using any numeric system, so we hand-coded each respondent's occupation.

[2]For the NLS-72, HSB, and NELS surveys, centile ranks were assigned using an algorithm incorporating base-year weights. Mean centile ranks calculated using follow-up weights.

[3]Scores were standardized using base-year survey weights; mean standardized scores calculated using follow-up weights.

Table 2.7a: Logistic Regressions of Teacher Entrance Models for Females with at least a High School Degree (I)

Dependent variable	WLS Teacher in 1964	Talent Teacher in 1971	NLS-72 Teacher in 1979	HSB Teacher in 1992	NELS Teacher in 2000
Sample mean	0.080	0.061	0.068	0.041	0.070
Sample size	4,609	1,634	6,751	5,389	4,284
	(1)	(2)	(3)	(4)	(5)
Intercept	2.694	-0.052	2.273	0.010	0.821
	(3.535)	(2.105)	(3.260)	(5.541)	(4.453)
Centile score	0.024	0.028	0.021	0.018	0.017
	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)
Age	-0.262	-0.163	-0.239	-0.160	-0.166
	(0.141)	(.080)	(0.124)	(0.198)	(0.170)
Marginal effect of centile score at:					
Mean centile	0.0014	0.0012	0.0011	0.0005	0.0009
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
85^{th} centile	0.0029	0.0028	0.0020	0.0009	0.0014
	(0.0004)	(0.0006)	(0.0003)	(0.0002)	(0.0003)
Elasticity of centile score at:					
Mean centile	0.022	0.027	0.020	0.017	0.015
	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)
85^{th} centile	0.021	0.025	0.019	0.017	0.015
	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)
Log-likelihood	-1197.0	-347.3	-1505.0	-758.7	-1059.7
Pseudo r-squared	0.060	0.070	0.047	0.029	0.030

Standard errors in parentheses. Sampling weights were used in columns (3), (4), and (5). See Appendix A for unweighted versions of these regressions. Marginal effects and elasticities calculated for a female high school graduate of average age. Missing values for the age variable were replaced with the sample mean; a dummy variable was included in the regression taking the value of one for all individuals for whom this replacement was made.

Table 2.7b: Logistic Regressions of Teacher Entrance Models for
Females with at least a High School Degree (II)

Dependent variable	WLS Teacher in 1964	Talent Teacher in 1971	NLS-72 Teacher in 1979	HSB Teacher in 1992	NELS Teacher in 2000
Sample mean	0.080	0.061	0.068	0.041	0.070
Sample size	4,609	1,634	6,751	5,389	4,284
	(1)	(2)	(3)	(4)	(5)
Intercept	3.805	1.746	3.018	1.165	1.148
	(3.520)	(2.103)	(3.265)	(5.413)	(4.489)
Standardized score	0.685	0.807	0.643	0.493	0.516
	(0.063)	(0.121)	(0.055)	(0.090)	(0.071)
Age	-0.259	-0.178	-0.227	-0.168	-0.147
	(0.141)	(0.079)	(0.126)	(0.195)	(0.172)
Marginal effect of standardized score at:					
Mean score	0.042	0.036	0.032	0.014	0.027
	(0.003)	(0.005)	(0.003)	(0.002)	(0.004)
One standard deviation above mean	0.073	0.072	0.055	0.022	0.042
	(0.009)	(0.015)	(0.007)	(0.005)	(0.008)
Elasticity of standardized score at:					
Mean score	0.640	0.769	0.609	0.479	0.488
	(0.060)	(0.118)	(0.053)	(0.088)	(0.067)
One standard deviation above mean	0.602	0.727	0.581	0.470	0.471
	(0.053)	(0.106)	(0.049)	(0.085)	(0.063)
Log-likelihood	-1200.1	-347.5	-1504.4	-759.2	-999.0
Pseudo r-squared	0.058	0.070	0.047	0.029	0.045

Standard errors in parentheses. Sampling weights were used in columns (3), (4), and (5). See Appendix A for unweighted versions of these regressions. Marginal effects and elasticities calculated for a female high school graduate of average age. Missing values for the age variable were replaced with the sample mean; a dummy variable was included in the regression taking the value of one for all individuals for whom this replacement was made.

Table 2.8: Logistic Regressions, Who Enters Teaching,
Females with at least a High School Degree

Dependent variable	WLS Teacher in 1964	Talent Teacher in 1971	NLS-72 Teacher in 1979	HSB Teacher in 1992	NELS Teacher in 2000
Sample mean	0.080	0.061	0.068	0.041	0.070
Sample size	4,609	1,634	6,751	5,389	4,284
	(1)	(2)	(3)	(4)	(5)
Intercept	2.693	0.273	0.427	-0.261	-1.968
	(3.585)	(2.113)	(3.455)	(5.980)	(4.765)
10th decile	2.131	2.350	2.560	1.517	2.395
	(0.333)	(0.476)	(0.420)	(0.539)	(0.533)
9th decile	1.869	1.962	2.636	1.723	3.094
	(0.335)	(0.486)	(0.416)	(0.533)	(0.536)
8th decile	1.749	1.736	2.801	1.268	2.262
	(0.347)	(0.496)	(0.415)	(0.547)	(0.536)
7th decile	1.461	1.270	2.395	1.584	2.793
	(0.341)	(0.528)	(0.422)	(0.529)	(0.533)
6th decile	1.428	1.095	2.245	1.177	2.542
	(0.346)	(0.552)	(0.423)	(0.553)	(0.528)
5th decile	1.289	1.344	2.055	1.323	2.142
	(0.367)	(0.529)	(0.424)	(0.561)	(0.543)
4th decile	0.696	0.757	1.678	0.112	2.302
	(0.366)	(0.588)	(0.446)	(0.641)	(0.570)
3rd decile	-0.103	-0.373	1.381	0.217	1.900
	(0.471)	(0.824)	(0.464)	(0.637)	(0.562)
2nd decile	0.087		0.609	0.337	0.933
	(0.432)		(0.510)	(0.642)	(0.592)
Age	-0.258	-0.161	-0.201	-0.152	-0.111
	(0.142)	(0.080)	(0.132)	(0.212)	(0.179)
Log-likelihood	-1192.5	-346.3	-1489.8	-753.0	-979.8
Pseudo r-squared	0.064	0.073	0.056	0.037	0.064

Standard errors in parentheses. Sampling weights were used in columns (3), (4), and (5). See Appendix A for unweighted versions of these regressions. Missing values for the age variable were replaced with the sample mean; a dummy variable was included in the regression taking the value of one for all individuals for whom this replacement was made.

Table 2.9: Predicted Probabilities of Entering Teaching as an Occupation,
Females with at least a High School Degree

	Predicted Probabilities					Predicted Probabilities as Proportion of the Sample Mean				
	WLS	Talent	NLS-72	HSB	NELS	WLS	Talent	NLS-72	HSB	NELS
Sample mean	0.080	0.061	0.068	0.041	0.070	0.080	0.061	0.068	0.041	0.070
Decile of test score	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
10th	0.169	0.147	0.096	0.057	0.079	2.11	2.41	1.41	1.39	1.13
9th	0.135	0.111	0.109	0.054	0.145	1.69	1.82	1.60	1.32	2.07
8th	0.122	0.092	0.117	0.046	0.069	1.53	1.51	1.72	1.12	0.99
7th	0.094	0.061	0.089	0.062	0.112	1.18	1.00	1.31	1.51	1.60
6th	0.090	0.049	0.079	0.041	0.089	1.13	0.80	1.16	1.00	1.27
5th	0.079	0.063	0.068	0.047	0.062	0.99	1.03	1.00	1.15	0.89
4th	0.045	0.037	0.048	0.024	0.071	0.56	0.61	0.71	0.59	1.01
3rd	0.021	0.012	0.029	0.021	0.049	0.26	0.20	0.43	0.51	0.70
2nd	0.024	0.018	0.018	0.022	0.019	0.30	0.30	0.27	0.54	0.27
1st	0.022	0.019	0.001	0.017	0.007	0.28	0.31	0.02	0.42	0.10

Values in columns (1) – (5) are the average predicted probability of entering the teaching profession, by decile, for a female with at least a high school degree. These columns correspond to reported (weighted) results in Table 2.8. See Appendix A for the predicted probabilities, based on unweighted regression results. Values in columns (6) – (10) are the predicted probabilities in columns (1) – (5), normalized by the sample mean for each cohort.

Table 2.10: Characteristics of New Male Teachers, 1964-2000

	WLS 1964	Talent 1971-74	NLS-72 1979	HSB 1992	NELS 2000
Sample size (males with at least a high school degree)	4,379	1,527	6,444	4,984	3,855
Number of teachers	169	69	135	62	87
Mean centile ranking	60.6	59.6	60.9	71.6	64.6
Mean standard score	0.39	0.30	0.37	0.78	0.50
Average predicted probability of teaching					
1st-2nd deciles	0.013	0.016	0.009	0.002	0.007
3rd-4th deciles	0.042	0.033	0.013	0.006	0.020
5th-6th deciles	0.038	0.059	0.025	0.009	0.024
7th-8th deciles	0.038	0.049	0.025	0.009	0.025
9th-10th deciles	0.063	0.067	0.027	0.025	0.038
Average predicted probability divided by sample mean					
1st-2nd deciles	0.328	0.363	0.438	0.197	0.323
3rd-4th deciles	1.067	0.731	0.585	0.554	0.885
5th-6th deciles	0.976	1.306	1.182	0.898	1.070
7th-8th deciles	0.985	1.084	1.186	0.883	1.150
9th-10th deciles	1.624	1.485	1.247	2.490	1.733

Predicted probabilities are calculated in the same manner as those in Table 2.9. Probabilities have been computed for males with at least a high school degree, of average age. Sampling weights were used in columns (3), (4), and (5). See Appendix A for information about the weights used in these columns.

CHAPTER THREE

POPULATION HETEROGENEITY, INCOME INEQUALITY, AND THE SUPPORT FOR PUBLIC EDUCATION

3.1 Introduction

At the same time that public school districts are struggling to attract the most talented candidates into the teaching profession (shown in chapter two), they are also facing a different kind of challenge—rising heterogeneity within their own constituencies. These demographic changes could have significant implications for the future of public education in the United States. For example, in the presence of human capital spillovers, greater interaction between families of different skills, incomes, races, or language abilities through the education system may do much to reduce long-run inequalities in income and educational attainment in the population, as Benabou (1996), Schwab and Oates (1991) and others have suggested. On the other hand, if this increased heterogeneity is accompanied by significant disagreement over the level or use of school resources, some school district residents may respond by reducing their support for public schooling, or by withdrawing from the system altogether.

Heterogeneity within school districts has risen markedly since 1970.[43] Districts have not only become more racially and ethnically diverse, but also more

[43] We show this to be true along a number of dimensions in a sample of 8,700 unified public school districts, in Section 3.3.

heterogeneous in age, educational attainment, native language and immigrant status, religious preference, home ownership rates, and wealth. Rising inequality in income is another important dimension of heterogeneity within school districts—at the national level, household income inequality rose over 16 percent between 1969 and 1999 (as measured by the Gini coefficient); about three quarters of all U.S. school districts experienced a rise in income inequality to some degree over this period.[44] To the extent that these population characteristics are correlated with household preferences for education, it seems likely that greater fractionalization in these characteristics would have some effect on local support for education. The question of how these trends in heterogeneity and income inequality have actually affected local school districts, however, is largely unexplored.

Economists and other social scientists have begun to examine the impact of heterogeneity on public finance in general, and much of the evidence thus far suggests that heterogeneity within jurisdictions tends to reduce the support for public programs. Explanations for this finding have varied. One model, for example, assumes that a voter's willingness to support public spending depends on how close their preferred use of funds is to their expected use. In this model, as preferences become more polarized, the typical voter becomes more likely to disagree with the median voter over the ultimate use of funds and chooses to reduce their support for public spending

---

[44] See http://www.census.gov/hhes/income/histinc/ie6.html (access date August 9, 2003) for Census data on nationwide household income inequality. As we show in Section 3.3, rising income inequality has not been strictly a between-district phenomenon.

altogether.[45] Other authors have suggested a special case of this model, where voters' support for public programs depends on the share of beneficiaries who come from the same socioeconomic or ethnic group.[46]

The link between income inequality and school finance has also received some attention. Epple and Romano (1996) and Fernandez and Rogerson (1995) have developed theoretical models that show that rising income inequality may be harmful to school spending. Epple and Romano describe the problem as "the ends against the middle," where high-income households oppose spending on public schools because of their high demand for private schools, and poorer families prefer lower taxes and expenditure in general.[47]

In this chapter, we use the demographic characteristics from a national panel of 8,700 unified school districts in 1970, 1980, 1990, and 2000 to explore how changing within-district population heterogeneity has affected local support for public education. We measure support for education in two ways—local tax dollars and student participation—and examine the impact of heterogeneity on both local education revenues per student and the fraction of students within district boundaries who enroll in private schools. We utilize a within-group estimator that holds constant the permanent unobserved characteristics of school districts and examines whether

---

[45] Alesina, Baqir and Easterly (1999).

[46] Luttmer (2001).

[47] Goldin and Katz (1997) find some evidence that income inequality slowed the development of secondary schooling in some communities during the first half of the 20th century. They attribute this in part to the opposition of both high- and low-income families to higher spending on education.

changing within-district support for schools over time is correlated with changes over time in population heterogeneity. Because school spending varies considerably across states and time, we also hold constant changes in spending that are common to all districts in a state for a given year.

We find that rising income inequality within a school district actually increases the level of real per-pupil expenditure at the local level, a finding consistent with a median voter model in which rising income inequality reduces the tax price to the voter with the median income. This result is robust to several measures of income inequality, and across numerous sample definitions. We find that increased heterogeneity in race and schooling reduce per-pupil expenditure somewhat; greater racial heterogeneity within a school district also tends to increase the proportion of children who enroll in private schools. Income inequality appears to have little impact on private schooling.

This chapter is organized as follows. In Section 3.2, we argue that despite opportunities to migrate between jurisdictions (a key assumption of the Tiebout hypothesis), communities in practice are in fact quite heterogeneous. Given that perfect stratification of households is rare, we explore some of the ways in which heterogeneity might affect support for schools by presenting a simple median voter model of school spending; we also briefly review other literature on this subject. Section 3.3 describes our empirical model and data. Section 3.4 presents our results, and Section 3.5 concludes.

3.2 Heterogeneity and Collective Choice

3.2.1 Some Evidence on Heterogeneity and Tiebout Sorting

Before we address why heterogeneity within a community might affect the demand for local public goods, a prerequisite question might be: why do local communities have heterogeneous populations? The benchmark local public finance theory of community formation—that of Tiebout (1956)—largely ignores heterogeneity within jurisdictions by assuming that households costlessly sort into homogeneous communities that offer their preferred level of public services.[48] Multiple jurisdictions, together with zoning regulations and a property tax, create an efficient "market" for public goods in which each household exactly satisfies their demand for public services, and (through the capitalization of taxes and benefits into property values) faces a "price" commensurate with their level of demand for those services.[49]

In practice, communities are often much more heterogeneous than the Tiebout model might predict.[50] Labor market decisions, moving costs, government interventions into the housing market, infrequent entry and exit of governments, the

---

48 We emphasize that perfect Tiebout sorting implies homogeneity in *preferences* for public spending, not necessarily in observed characteristics (see Hamilton (1976), Epple and Platt (1998)). In practice, however, demand for education spending appears to be strongly correlated across households with similar observed characteristics.

49 Hamilton (1975) introduced the idea of municipal zoning into the original Tiebout model. Zoning acts not only as a barrier to entry (resulting in an inelastic supply of housing in the locality) but also as a tool in which municipalities can indirectly set the tax price that households pay for local services.

50 In the words of Oates (1981), "the pure [Tiebout] model … involves a set of assumptions so patently unrealistic as to verge on the outrageous."

bundling of public services, and imperfect information about communities all act as barriers to perfect sorting across jurisdictions. Communities also seem to be willing to tolerate some heterogeneity in the population in return for economics of scale in the production of public goods. Alesina, Baqir, and Hoxby (2000) find evidence for such a tradeoff in U.S. school districts, municipalities, and special districts over the 1960-1990 period.

Yet even as Tiebout "market imperfections" have mitigated over time, communities remain significantly less homogeneous than one would expect.[51] In a recent paper, Rhode and Strumpf (forthcoming) document a persistent decline in stratification by race, income, education, and age (as well as other proxies for public spending preferences) across U.S. municipalities and counties over the past century, despite plummeting mobility costs that in theory should have facilitated greater Tiebout sorting. Perhaps the most striking of their findings relates to income—they show that within-county income inequality has remained roughly constant since 1949 (declining over the 1949-1979 period, and then rising thereafter), while between-county inequality declined over the 1949-1979 period (but rose modestly during the 1980s). This result parallels a similar finding for counties within the Boston

[51] Rhode and Strumpf (forthcoming) argue that improvements in transportation and communications technology have contributed to large reductions in moving costs over time; Fischel (2001) claims that the suburbanization of America has minimized the importance of labor market decisions and commuting costs in household location decisions, and increased households' ability to sort among jurisdictions.

metropolitan area over the same years. The authors interpret these results as evidence of little, if any, increase in sorting by income across U.S. counties over time.[52]

Other authors have pointed out patterns or trends in demographics that appear on the surface to be inconsistent with perfect Tiebout sorting. Pack and Pack (1978) in a frequently cited paper find sufficiently large variation in income and house prices within communities in Pennsylvania MSAs to rule out perfect Tiebout sorting. Cutler, Glaeser and Vigdor (1999)—while not intended as a test of the Tiebout hypothesis per se—find evidence of a persistent increase in racial integration within census tracts in metropolitan areas, and particularly in suburban tracts, since 1970.[53]

The question of how much heterogeneity is sufficient to refute the Tiebout hypothesis persists, and none of these findings should be taken as prima facie evidence against Tiebout sorting. Indeed, a score of capitalization studies, beginning with Oates (1969), have demonstrated that homebuyers almost certainly take local policies into account when making their housing investments. They do, however, suggest that forces other than Tiebout sorting have been important enough to prevent perfect

[52] The sharp decline in the number of local jurisdictions over this period of falling mobility costs would also appear to support this conclusion.

[53] It should be pointed out that race and ethnicity continue to be important correlates with preferences over local policies, independent of income. A recent article in *The Washington Post* about opinions in Prince George's County, Maryland—a county described in the article as "a symbol of black affluence"—is illustrative: "A stark divide exists in the way blacks and whites view a broad spectrum of institutions and conditions in Prince George's County, from the police and public schools … Blacks were twice as likely to be satisfied with the public schools and to view the police force as being overly aggressive … Half of blacks said African Americans do not hold enough political power, [and] half of the whites said it would be better if more whites moved to Prince George's." ["Prince George's Split Along Race Lines, Poll Shows," August 25, 2002].

stratification of households across communities along demographic characteristics—a condition that is key to our analysis that follows.

3.2.2 A Median Voter Model of School Expenditure

Absent the ability (or desire) to sort into perfectly homogeneous jurisdictions, conflicting household demands for public goods must instead be resolved through the political process.[54] When public decisions are made through non-market processes, it becomes necessary to consider the mechanism through which household preferences are aggregated into a collective choice. In the public finance literature, the median voter model remains a simple yet powerful tool for modeling the demand for public goods in heterogeneous communities. Under certain conditions (among other things, preferences that are single-peaked), the median voter model predicts that voting by majority rule over a single-dimensional public good will result in the provision of the median voter's desired level of public services.

To better understand how increased heterogeneity or income inequality might affect school expenditure (and to help motivate our empirical analysis in the next section), we present a stylized median voter model of a community voting over the size of a local school budget. Household-voters are allowed to be heterogeneous first in income, and later in preferences over the nature (but not the size) of the school budget. This simple model is intended only to be illustrative of some of the ways in which population heterogeneity might affect school spending—for convenience, we

[54] Drazen (2000) points out that even when preferences for public goods are identical within a community, voters often have the incentive to disagree over the incidence of taxation (he refers to this as *ex-post* heterogeneity in the electorate).

have ignored many important features of real communities, including mobility between jurisdictions, and private schooling alternatives. We discuss the implications of these extensions to the model at the end of this section.

Model

Suppose there are H households in a school district, each containing one school aged student, and one voter. Household income y_h is exogenous and distributed according to $f(y_h)$, a non-symmetric income distribution. Preferences over public schooling and private consumption are identical across households, and are given by the following utility function for household h:

(1) $u(g, c_h)$

where $g = G/H$ is per pupil expenditure (the same for all households), G is the total school budget, H is the number of households and c_h is private consumption (the price of both g and c_h is assumed to be one).

School district residents determine by majority vote the size of the district budget G, as well as a constant proportional tax rate on income t.[55] The individual household and government budget constraints are given by (2) and (3), respectively:

(2) $y_h = c + ty_h$

---

[55] Most school districts in the United States finance local school expenditures via property taxes, not income taxes (the average unified school district in 2000-01 received 65 percent of its local revenues from property taxes, and less than 1 percent from income taxes (see U.S. Bureau of the Census (2001))). Our interest in this paper, however, is the effect of income inequality on school spending. While income and housing values are highly correlated, we acknowledge the limitations of this model.

$$(3) \qquad G = tH\int_h y_h f(y_h) = tH\mu_y$$

It is illustrative to combine (2) and (3), noting that $t\mu_y$=g (that is, that per-pupil expenditure *G/H* must equal the expected income tax per household), and re-write the household's budget constraint as:

$$(4) \qquad y_h = c + \left(\frac{y_h}{\mu_y}\right)g$$

In this expression, the term in brackets represents the "tax price" of an additional unit of per-pupil expenditure to household *h*. Solving the household's optimization problem is straightforward, and results in the following first order condition which characterizes the preferred level of per-pupil spending for household *h*:

$$(5) \qquad \frac{u_g}{u_c} = \left(\frac{y_h}{\mu_y}\right)$$

If *u(·)* is quasi-concave, preferences will be single-peaked and the median voter theorem will apply. If demand for *g* is monotonic in income (see panel (a) of Figure 3.1 for an illustration)—the majority voting outcome will be that level of *g* demanded by the household with the median income (y_m). In this simple case, the level of per-pupil spending observed in each school district will be a function of both the median income in the district and the tax price to the median household (which differs from one whenever the distribution of income is skewed, and in particular will be less than one whenever $y_h < \mu_y$). As the median voter's tax share falls—as when, for example, the mean income in the district grows relative to the median—the price of an additional unit of school spending to the median voter falls, and the median voter will optimally choose a higher level of per-pupil expenditure.[56] In this scenario, an increase in income inequality that raises the mean district income relative to the median income should be expected to *increase* per-pupil spending.[57]

Demand for g may not, however, be monotonic in income.[58] If both low- and high-income households prefer lower levels of per-pupil expenditure (shown in panel (b) of Figure 3.1), per-pupil expenditure may have no clear systematic relationship with median income (the household with the median income will no longer be the decisive voter), and any increase in income inequality that results in higher concentrations of households in the tails of the income distribution is likely to result in *lower* levels of per-pupil spending.

From the baseline model presented above, greater income inequality will tend to increase expenditure when demand is monotonic in income (when greater income inequality represents a lower tax price for the median income household); it tends to lower expenditure when demand is not monotonic in income (a case of "the ends against the middle"). However, even when demand for school spending is monotonic in income, rising heterogeneity in the population may lead to reductions in total spending for other reasons, as we consider in the following extension.

Following Alesina, Baqir, and Easterly (1999) we extend our simple model by assuming that households have identical preferences over the *levels* of per-pupil

[56] Whether or not the median voter outcome is socially efficient is another question (it will not be, in general)—see Rubinfeld (1987).

[57] Extending this model to allow for the disincentive effects of taxation do not qualitatively change this result—see Meltzer and Richard (1981).

[58] See Stiglitz (1974).

spending and private consumption, but are allowed to disagree over the *nature* of school spending (a disagreement over an appropriate school curriculum is one example). Let the parameter φ_h represent the distance along the ideological line between household *h*'s ideal use of the school budget and the actual use of the budget, and modify the household utility function from (1) to read:[59]

(6) $u(g(1-\varphi_h), c_h)$, $0 < \varphi_h < 1$

In this modified utility function, the marginal utility from an additional dollar of per-pupil expenditure is reduced to the extent that household *h* disagrees with the *way* in which the school budget is spent (that is, the higher is φ_h). Put another way, to household *h* the *effective* level of per-pupil expenditure is $g^*=g(1-\varphi_h)$. In a homogeneous district with no disagreement over the nature of school expenditures, φ_h will be zero for all households.

Now suppose that the size and the nature of the school budget are determined in a two-stage process. First, households vote over the size of the school budget (G, t). They then vote over an appropriate curriculum (more generally, the allocation of the budget). Using backward induction, voters know that in the second stage, the median voter (i.e. the voter with the median ideology) will prevail in the choice of a curriculum. Forward-looking voters will be able to anticipate this outcome—and thus know their $(1 - \varphi_h)$—and vote accordingly for their optimal (G, t) in the first stage. The household optimization problem will be identical to that above, but with the following first order condition:

[59] The ideological spectrum is normalized to a length of one.

(7) $$\frac{u_g}{u_c} = \left(\frac{y_h}{\mu_y}\right)\left(\frac{1}{1-\varphi_h}\right)$$

As in (5), *h*'s optimal choice of *g* is decreasing in the household's tax share, but now support for school spending will be further diminished to the extent that household *h* expects to disagree with the median voter over the use of school funds. In households where φ_h=0, this first order condition reduces to (5). As φ_h goes up, however, the price of an effective unit of *g* is increased—if households do not receive the full benefit of an additional unit of *g,* they will (according to this result) be less willing to substitute away from private consumption.

In the above extension, even when demand for *g* is monotone in income, the majority-voting outcome may not correspond to that level of *g* preferred by the voter with the median income; instead, the effect of heterogeneity in preferences over the allocation of *g* may dominate and determine the identity of the decisive voter. Hence, we may observe lower levels of per-pupil spending in districts where heterogeneity exists along dimensions other than income that are correlated with preferences over the allocation of public funds—race, ethnicity, age, religion, native language and educational backgrounds are but a few possible examples.

The introduction of private schooling alternatives considerably complicates this theoretical analysis, as other authors, including Stiglitz (1974), Epple and Romano (1996), and Glomm and Ravikumar (1998) have shown. Both Glomm and Ravikumar (1998) and Epple and Romano (1996) prove the existence of a majority-voting equilibrium in the presence of private alternatives—in the former, the focus is on an equilibrium in which the voter with the median income is decisive, in the latter several possible equilibria are shown to exist. Epple and Romano argue that for public goods like education, the likely majority voting equilibrium will be one in which there are two opposing coalitions of voters—one comprised of high- and low-income households who prefer a low level of expenditure on public education (with high-income households opting for private schooling alternatives), and another made up of middle-income households who prefer a high level of public expenditure on education. Thus, where private schooling alternatives exist, greater income inequality may increase the likelihood of an "ends against the middle" outcome, with lower spending on education and higher rates of private schooling.

While the model presented in this section provides few clear predictions about the impact of income inequality and population heterogeneity on support for public education, it does illustrate some of the ways in which diversity in the local population might affect school financing and overall participation in the public schools. It remains an empirical question as to how these forms of heterogeneity actually do affect public support for education in practice. Before introducing our own empirical model in Section 3.3, we summarize in the following section a growing empirical literature that has begun to explore the relationship between population heterogeneity and public finance.

3.2.3 Related Literature

A small but growing empirical literature has begun to study the impact of heterogeneity on public expenditure, and much of this literature appears to suggest that population heterogeneity reduces the support for public programs. Alesina, Baqir, and

Easterly (1999), for example, find that ethnic fragmentation within cities, counties, and metropolitan areas is negatively related to public spending on education and infrastructure. Similarly, Poterba (1997), Murray and Ladd (2000), and Harris, Evans, and Schwab (2001) show that per-pupil education spending within school districts declines with the elderly share, an observation consistent with an elderly preference to vote down public programs that they themselves do not benefit from.[60]

In a study of the growth of secondary schooling the United States during the early part of the 20th century, Goldin and Katz (1997) find that localities that supported the expansion of secondary schools were more likely to have relatively equal income distributions, as well as populations that were homogeneous in religious or ethnic background. Hoxby (1998) points out that the time pattern of spending inequality across school districts closely follows that of national income inequality—i.e. "the two decades in which spending inequality rose the most (the 1930's and the 1970's) were also decades of rising income inequality. The decade in which spending inequality fell the most (the 1940's) was a decade of falling income inequality" (p. 310).

In another strand of this empirical literature, several authors have found that interpersonal preferences—that is, preferences for social programs that depend on the types of individuals who benefit from them—may also play a role in the support for public spending. Luttmer (2001), for example, finds using individual-level data from the General Social Survey that support for redistribution declines in the number of

[60] Both Harris et al and Murray and Ladd find that this effect is less pronounced at the local level, where elderly homeowners are likely to benefit indirectly from education spending via property values.

local welfare recipients, but rises with the share of recipients who come from the same racial group. Alesina, Glaeser and Sacerdote (2001) argue that differences in U.S. and European support for welfare spending can be largely explained by interpersonal preferences—because welfare recipients in the United States are disproportionately black, the majority of (non-black) voters seem to be unwilling to support a generous system of redistribution.

A smaller literature has looked at the impact of local population heterogeneity on private school enrollment. Betts and Fairlie (2003) use household-level Census data from 1980 and 1990 to test whether a rising foreign-born population is associated with higher enrollment of (native-born) white students in private schools. While they find no relationship between immigration and enrollment in private elementary schools, they do find a statistically significant impact of immigration on enrollment in private secondary schools (they attribute this effect primarily to differences in the native languages of school-aged children). Clotfelter (1997) finds that private school enrollment is positively affected by income inequality and the presence of nonwhites in the public schools. Hoxby (2000) includes measures of income inequality and racial and ethnic heterogeneity in district- and MSA-level regressions in a paper assessing the impact of school district competition on private school enrollment, school expenditure, and student achievement, but (because her paper focuses on the effects of competition between school districts) does not discuss her empirical findings on these measures.

Finally, in a separate context some authors have treated income inequality or other measures of within-district heterogeneity as endogenous variables; many of these papers might be characterized as tests of the Tiebout sorting hypothesis within urban areas. Urquiola (2000), for example, examines how competition between school districts within metropolitan areas affects the level of ethnic heterogeneity and income inequality within school districts. He finds that greater competition within MSAs tends to result in greater homogeneity in income and race within school districts located in those MSAs. Aaronson (1999) finds that the school finance reforms of the 1970s and 1980s (which effectively reduced the level of between-district competition in affected states) reduced sorting by income across school districts (which he measures using within-district measures of income inequality).[61]

Through our empirical model introduced in the next section, we hope to contribute to this literature by examining the relationship between changes in within-school district heterogeneity over time (specifically, fractionalization in race and schooling, and income inequality) and support for public education, as measured through per-pupil spending, and private school enrollment.

3.3 Empirical Model and Data

3.3.1 Econometric Strategy

The goal of this chapter is to examine how rising within-district heterogeneity has affected the support for public education. The stylized model and subsequent discussion in Section 3.2 suggested several ways that increased heterogeneity might affect this support. First, when demand is monotone in income, increased income

[61] See also Eberts and Gronberg (1981), Grubb (1982), and Schmidt (1992), for other empirical studies that treat within-district income inequality as an endogenous variable.

inequality may represent a reduction in the tax price to the voter with the median income, yielding an increase in the overall level of spending. Where demand is not monotonic in income, increased income inequality may decrease total spending on education. Second, even if demand is monotone with income, heterogeneity in preferences over the nature of school spending may result in reduced support for public education if households find themselves unable to agree over the allocation of school funds. Interpersonal preferences that reflect households' willingness to support members of other racial or socioeconomic groups may be important in this context. Third, where private schooling alternatives exist, rising income inequality may result in a coalition of "the ends against the middle," where low- and high-income households reduce their support for public education (and the wealthy enroll in private schools). Finally, to the extent that mobility is an option, some households may respond to greater within-district heterogeneity through exit.

In this section, we present an empirical model designed to explore how changes in within-district heterogeneity over time are associated with changes in local support for public education. We estimate this model with a balanced panel of roughly 8,700 school districts in the United States over a 30-year period.[62] The unique structure of our data allows us to control for certain unobserved characteristics correlated with tastes for education spending and private schooling that a cross-sectional analysis would not, a feature we describe in further detail below.

[62] This data is described in more detail in Section 3.3.2 (following), and in Appendix B.

Our basic empirical model is as follows:

(8) $$y_{ijt} = X_{ijt}\beta + H_{ijt}\gamma + \delta_{ij} + S_{jt} + \varepsilon_{ijt}.$$

Our endogenous variable y_{ijt} is, alternately, local per-pupil education revenues in district i in state j in year t, and the fraction of students within district i's boundaries enrolled in private school in year t.[63] X_{ijt} is a vector of exogenous demographic and socioeconomic characteristics in district i in year t thought to be correlated with tastes for school expenditure and private schooling, while H_{ijt} is a vector of within-district population heterogeneity measures (explained further below). Specifically, we include in X_{ijt} measures of the racial composition of the district, median family income, the fraction of households in poverty, the share of housing units in the district that are owner-occupied, and the educational attainment of district residents.[64] Our vector of heterogeneity measures H_{ijt} includes indices of within-district fractionalization in race and years of schooling, and varying measures of income inequality.[65] The error term ε_{ijt} is assumed to have a zero mean and constant variance.

[63] Per-pupil revenues and per-pupil expenditures are nearly identical measures (revenue data excludes capital expenditures), so we will use these terms interchangeably.

[64] The inclusion of these variables follows a long tradition of empirical estimation of demand functions for local public goods. See Bergstrom and Goodman (1973) for the seminal empirical work on the demand for local public goods. Lovell (1978), Brown and Saks (1985), Brazer and McCarty (1987), Rubinfeld and Shapiro (1989), Santerre (1989), Magna and Lee (1990), and Sass (1991) all provide examples of empirical estimates of demand functions for school spending. Most models typically include a "tax price" as a right-hand-side variable. We do not, however as discussed in Section 3.2, our measure of income inequality may proxy for the tax price to the median voter.

[65] Our data includes other interesting measures of population heterogeneity, including the foreign born share of the population, the distribution of English language ability

Because much of the variation in local revenues and private school enrollment shares is between districts—not within districts over time—we include district fixed effects (δ_{ij}) in each model. While much of the between-district variation in spending and private school enrollment can be explained through observed district characteristics, one might be concerned that there are unobserved characteristics (e.g., a stronger preference for education) that are correlated with our right-hand-side variables of interest (e.g., income). Our district fixed-effects will capture any of these unobserved characteristics that are permanent within districts over time.

Likewise, some of the variation in school revenues and private schooling shares may be due to state-specific trends in or shocks to these variables that are also correlated with our right-hand side variables. For example, a statewide economic downturn may affect both local revenues for education and income inequality. We include year-specific state effects S_{jt} in each of our models to capture any such trends.[66]

Taken together, our empirical model estimates how local per-pupil revenues and private school enrollment shares change as population demographics and within-district heterogeneity change over time, absent any trends in these variables within a state. Our use of district fixed-effects implies that we are using a within-group

---

within a district, and the distribution of housing values within a district. We hope to incorporate these measures into the empirical model in a later version of this paper.

[66] We will also be estimating this model on a sample of school districts within metropolitan areas. In those specifications, we include year-specific MSA effects, rather than state-year effects.

estimator, which means that we will use within-district variation in population heterogeneity and spending over time to identify our parameter of interest (γ).

One might be concerned that some of the changes over time in within-district population heterogeneity are in response to the policies or performance of local school districts (in particular, to the spending policies of the district). Work by Urquiola (2000), Aaronson (1999), and others discussed in the previous section suggested that income inequality within school districts might in part be a function of the competitiveness of the local market for education. Where households have a greater ability to Tiebout-sort into neighboring districts, dissatisfaction with local school spending may result in out-migration and a subsequent change in the local income distribution. To the extent this sorting occurs, our heterogeneity measures H_{ijt} will be endogenous and ordinary least squares estimates of γ will be biased and inconsistent.

We deal with the potential endogeneity of H_{ijt} in several ways. As a first pass, we interact H_{ijt} with several different measures of local school district competition, to see whether the impact of local population heterogeneity on local spending varies across districts with different sorting opportunities. If sorting is important, we might expect that any impact that local population heterogeneity might have (positive or negative) on district revenues would be lessened in districts with greater sorting opportunities. Next, we estimate (8) using subsamples of low-mobility districts (as measured by the percent of residents who lived in a different county five years earlier) and districts located outside of metropolitan areas, where Tiebout sorting would seem less likely to be an important force. Finally, we estimate an instrumental variables model, where we instrument for within-district income inequality using a measure of income inequality from a nearby district outside a specified distance. The idea here is to identify the effect of local income inequality only with that portion of within-district income inequality that is attributable to regional labor market conditions, not to inter-district sorting.

3.3.2 Data Sources

The empirical strategy outlined above requires a dataset that contains school-district level demographic and financial data, together with detailed measures of within-district population heterogeneity and income inequality. There is no readily available data source that meets all of these criteria—however, we constructed such a data set by merging eight national school district data sources: the *1970 Census of Population and Housing Special Fifth-Count Tallies,* the *1980 Census of Population and Housing Summary Tape File 3F,* the *1990 Census School District Special Tabulation,* the *2000 Census of Population and Housing School District Tabulation (STP2),* the *1972, 1982, and 1992 Census of Governments: School Districts,* and the *2000-01 F-33 School System Finance File.*[67] The first four of these data sets provide detailed information about the demographic characteristics of the population living within each U.S. school district, while the latter represent the primary historical source of school finance data in the United States. By merging these eight data sources, we were able to construct a national panel of the socioeconomic, demographic, and financial characteristics of unified public school districts for 1970, 1980, 1990, and

[67] Caroline Hoxby (1996 and 2001) was the first to merge these datasets (at least through 1992) to create a national panel of public school districts.

2000. Detailed information about the construction and contents of this panel can be found in Appendix B.[68]

School districts in the United States are typically organized as unified (K-12), elementary-only, or secondary-only districts. Because the organization and cost structure varies across these different types of school districts, we chose to restrict our analysis to unified districts. This decision was not overly restrictive—despite having only 62 percent of all U.S. districts in our panel (as measured in 2000-01), the included districts represent over 80 percent of all publicly enrolled K-12 students in the nation.[69]

In our analysis in Section 3.4, we make use of three different samples of school districts. The first is our full sample of unified public school districts, a balanced panel of 8,699 school districts in 1970, 1980, 1990, and 2000 (a total of 34,796 observations). The other two are subsets of the full panel—one a sample of districts located in metropolitan areas (MSAs), with 3,292 districts (13,168 observations), and the other the remaining sample of districts located outside of MSAs, with 5,407

[68] An earlier version of this panel was used in Harris, Evans, and Schwab (2001). For information about the construction of the original panel, see Harris (1999), at http://www.bsos.umd.edu/econ/evans/wkpap.htm (access date August 9, 2003).

[69] Total public Pre-K-12 enrollment in 2000-2001 was 47 million, according to the 2001 Digest of Education Statistics (see http://nces.ed.gov/pubs2001/digest/dt003.asp, access date August 10, 2003). Our panel represents a total of 37.7 million K-12 students in 2000. About a third of all unified districts in 2000 are missing from our panel—this is due primarily to the aggregation of small districts in the 1970 Census data, or the consolidation of districts over the sample period. See Appendix B for details.

districts (21,628 observations).[70] Within-district variation in population heterogeneity and income inequality tends to be greater and more prevalent in urban areas; private schooling also tends to be more important in urban areas where private schools are more plentiful than in rural or sparsely populated districts. Thus, we present most results for both the full sample of districts and our subset of districts in MSAs. On the other hand, our estimates from the MSA subsample are most likely to suffer from the endogeneity bias discussed in the previous section. Therefore, we supplement our results with estimates from the non-MSA subsample, where appropriate.

Our two endogenous variables are local school district revenues per pupil, and the fraction of school age children enrolled in private school. Data on local revenues per pupil comes from the Census of Governments data files. This variable represents all revenues raised locally for education, via property taxes, income taxes, sales taxes, special fees, or any other local revenue source. Of course, more than half of the typical school district's budget comes from outside the district, from state or federal sources. However, we focus on local revenues because we feel this measure best reflects local commitment to public education. Private school enrollment shares (at least in 1970, 1980, and 1990) are calculated using enrollment data from the special school district tabulations of the decennial Census. Unfortunately, the Census did not provide private school enrollment counts in 2000. To obtain these counts, we used

---

[70] Appendix B provides the counts of school districts in each state. Districts were included in the MSA subsample if they reside in a county that was considered part of a metropolitan area according to the U.S. Census 1973 standards (see http://www.census.gov/population/estimates/metro-city/73mfips.txt (access date August 9, 2003) for a complete list of 1973 metropolitan areas).

GIS shapefiles to match census tracts to school districts and aggregate private school enrollment counts to the school district level. See Appendix B for details.

No measures of within-district heterogeneity or income inequality were available in any of our source datasets. We were, however, able to generate our measures using data from the Census demographic files. Our measures of within-district heterogeneity in race and schooling are simple fractionalization indices—one minus the sum of the squared population shares across four different race or educational attainment categories.[71] These fractionalization indices range from zero to one, and can be interpreted as the probability that two randomly selected school district residents come from different race or educational backgrounds. A value of zero would indicate a perfectly homogeneous district; a value closer to one would represent greater heterogeneity.[72]

Generating a measure of within-district income inequality was a bit more difficult. While the Census frequently reports Gini coefficients of income inequality at the state and national level, they do not report these coefficients for smaller geographic areas like counties or school districts. Fortunately, our Census data does contain some information about the distribution of income in each school district, in the form of counts of families who fall into certain income ranges in each year. As we

---

[71] For race fractionalization, our categories in all years are: white, black, Asian and Pacific Islander (API), and "other." For fractionalization in schooling, our categories are the fraction of the population 25 and older who are high school dropouts, high school graduates only, recipients of some college education, or college graduates (and higher). See Appendix B.

[72] With four categories, the largest value that our fractionalization index can take is: $(1 - (.25^2 + .25^2 + .25^2 + .25^2)) = 0.75$. See Vigdor (2001) for a theoretical interpretation of these indices.

explain in Appendix B, if one assumes a flexible functional form for the CDF of income within a district, the counts of families in each income group can be used in a maximum likelihood procedure to estimate the parameters of this distribution. Given these parameters, the Gini coefficient can then (for some distributions) be directly calculated. We assume that family income in each school district follows a three-parameter Dagum (1977, 1980) distribution—a re-parameterization of the Burr Type III distribution—and compute Gini coefficients of income inequality for each district in our panel. In addition to these school district Ginis, we use the Dagum parameters to compute three other measures of within-district income inequality—the log of the ratio of the 95th centile of income to the median income (the log(95/50) ratio), the log of the ratio of the median income to the 5th centile of income (the log(50/5) ratio), and the log of the ratio of the 95th centile of income to the 5th centilie of income (the log(95/5) ratio).

3.3.3 Overview of Data

Summary statistics for our full sample and MSA subsample are provided in Tables 3.1 and 3.2. In these tables—as with all of our regression models—school district observations are weighted by the number of publicly enrolled students in the district, and all monetary variables are in constant (1992) dollars. Through weighting, we are estimating the impact of population heterogeneity or income inequality on a randomly selected public school student from the population.[73]

[73] In our regressions where the private school enrollment share is our left hand side variable, we weight using the total school age population in each district, rather than

The rise in within-district heterogeneity and income inequality alluded to in our introduction is observable in both our full sample and MSA subsample—heterogeneity in race and schooling have risen monotonically over the 1970-2000 period in both samples, as has income inequality, as measured by both the Gini coefficient and the log(95/50) ratio. In the full sample, we find that the average Gini coefficient has risen almost 15 percent, from 0.34 to 0.39; our indices of schooling and race fractionalization rose 14 percent and 90 percent, respectively.[74] With the exception of schooling fractionalization, the growth in within-district heterogeneity was even more striking in MSAs, with income inequality in the average district rising 19 percent, from 0.33 to 0.39; race fractionalization more than doubled, from 0.18 to 0.36.[75] Inequality in the lower half of the income distribution (as measured by the log(50/5) ratio) also rose over the 1970-1990 period in both samples, but fell slightly from 1990 to 2000.

---

public school enrollment. We have experimented with several other weights, including an average of public school enrollment over the 1970-2000 period. Our results are generally insensitive to our choice of weights.

[74] By comparison, household income inequality reported by the Census rose 16.2 percent nationwide, from .401 to .466 (see http://www.census.gov/hhes/income/histinc/ie6.html, access date August 9, 2003). It is interesting to point out that average schooling fractionalization in 2000 is near its four-category maximum of 0.75, as the fraction of school district residents with less than a high school degree, a high school degree only, some college, and college or more have converged to roughly 25 percent each in the full sample.

[75] The increase in income inequality, race and schooling fractionalization was less dramatic outside of urban areas. In our non-MSA sample, income inequality rose 5.1 percent on average from 1970 to 2000 (from a Gini coefficient of 0.362 in 1970 to 0.380 in 2000), compared to a 14.6 percent average increase in our MSA sample. The increase in our race and schooling fractionalization indices were 24.8 percent and 53.8 percent in our non-MSA sample (compared to 10.3 percent and 100 percent in the MSA sample).

Tables 3.1 and 3.2 also show a steady rise in real local per-pupil revenues over our sample period. While the average district in our full sample raised $1,861 in revenues per-pupil in 1970 (in 1992 dollars), by 2000 this sum had risen to $2,736. At the same time, local school districts were contributing a smaller and smaller share of their total per-pupil budget—we see that the local share of total per-pupil revenues has steadily fallen, from over half in 1970 to just above 40 percent in 2000. Most of this decline is attributable to large increases in state support for education over this period—on average, state contributions to district per-pupil spending rose more than 127 percent from 1970 to 2000.[76]

There was no discernible trend in private school enrollment shares over this time period in the full sample—mean district enrollment in private schools in our sample remained roughly constant at 10 percent. Not surprisingly, the private school enrollment share was higher in metropolitan areas—on average, about 12 percent of school age children were enrolled in private schools in our MSA sample; only 6 percent of K-12 children outside of MSAs were enrolled in private schools over our sample period (although there is a distinct upward trend in private school enrollment in non-MSA districts—the share in private schools was about 4.9 percent in 1970, and 7.0 percent in 2000).[77]

[76] See Corcoran, et al (2003) for a recent survey.

[77] We also noticed a downward trend in private school enrollment shares in some of the largest U.S. urban school districts over this period, particularly between 1990 and 2000. Our finding that private school enrollment shares rose significantly outside of MSAs may be due to measurement error in our 2000 private school enrollment share variable, which was constructed by matching census tracts to school districts (rural districts were most likely to suffer from measurement error under this method—see Appendix B).

One might be concerned that these observed increases in mean within-district heterogeneity and income inequality are driven by our use of enrollment weights—in other words, that income inequality or race fractionalization was only important in a few select heavily populated districts. Table 3.3 shows this not to be the case. In this table, we provide greater detail about the distribution (both weighted and unweighted) of within-district changes in income inequality, race, and schooling fractionalization in our sample. Just under 75 percent of the school districts in our full sample experienced a rise in income inequality between 1970 and 2000; nearly 90 percent of these districts became more fractionalized in race and schooling.[78] On the average, the likelihood that two randomly selected individuals from a school district come from a different ethnic or educational background rose by nine and eleven percentage points in the full sample, respectively.

3.4 Results

3.4.1 Local Revenues per Pupil

We begin by estimating our empirical model in equation (8), using local revenues per pupil as our endogenous variable. Again, we include district fixed effects and year-specific state effects in all specifications, and all observations are weighted by total district public school enrollment. The results are shown in Table 3.4.

---

[78] The shape of this distribution was virtually identical for districts outside of MSAs. Roughly 75 percent of districts saw an increase in income inequality, and 90 percent saw an increase in race fractionalization. Almost 99 percent saw an increase in schooling fractionalization.

Our baseline specification in column (1) includes a detailed set of demographic and economic controls, as well as a within-district Gini coefficient of income inequality. In columns (2) – (4) we add our two other heterogeneity measures (race and schooling fractionalization), and in column (5) two additional controls (the percent of district households with children, and the fraction of the district population residing in an urbanized area).

Our estimated coefficient on the Gini coefficient is positive, sizable, and statistically significant. Across these five columns, our point estimate of this coefficient ranges from \$3,028 to \$3,567, with a standard error of about \$225 in all cases. The inclusion of our additional heterogeneity measures and controls have little effect on our estimate of this coefficient—in fact, if we were to construct a 95% confidence interval around these five estimates, all pairwise comparisons of these intervals would overlap. Given that the mean Gini coefficient in our sample is about 0.36 (with a standard deviation of 0.06), this result suggests that a two standard deviation increase in income inequality within a district in the full sample is associated with a \$363 to \$428 increase in per-pupil spending. This estimate is quite large—roughly equivalent in magnitude (by our estimates) to a \$8,500 increase in real median district family income (or, about a 21 percent increase over the average median income across districts).

The estimated coefficients on our demographic and economic characteristics are generally of the expected sign, and qualitatively similar to those found in other empirical estimates of local demand functions for education. Real local per-pupil revenues are positively related to median family income, with a \$1,000 increase in real median family income associated with a \$43 to \$53 increase in per-pupil spending (implying an approximate income elasticity of demand for per pupil spending of 0.7 to 0.9 at the mean, quite in line with other estimates, see Rubinfeld (1987)). Per-pupil revenues tend to be lower, all else equal, in districts with higher poverty rates and in districts where a greater proportion of residents are nonwhite; per-pupil revenues are higher in districts with a greater fraction of college graduates and a greater share of residents who are 65 and older.[79] We also find the typical "renter effect" found in other empirical estimates of local demand functions for public goods—a one standard deviation increase in the fraction of homeowners is associated with roughly a \$170 to \$300 decrease in per-pupil spending.[80]

Race and schooling fractionalization appear to have a negative impact on real per-pupil revenues in the full sample (statistically significant at conventional levels), although our point estimate for the fractionalization in schooling coefficient is far larger than that for fractionalization in race. Given a mean race fractionalization index of 0.25 in the full sample (and a standard deviation of 0.20), we estimate that a one standard deviation increase in race fractionalization is associated with a \$40 decrease in local per-pupil spending. By contrast, the effect of a one standard deviation rise in schooling heterogeneity is about a \$178 to \$195 decrease in per-pupil spending.

---

[79] Our estimated coefficient on the elderly share would seem to contradict the conclusion of Harris, Evans and Schwab (2001) that a larger elderly population share reduces public spending on education. However, in their paper they find that the impact of the elderly is much less pronounced at the local level, where older voters may feel that greater spending on education is capitalized into property values. Their model also uses the log of total revenues or local revenues as left hand side variables, which seems to have some effect on their result.

[80] See Oates (1998) for a survey of the "renter effect" literature.

Based on our results in the first five columns of Table 3.4, it would appear that increases in within-district heterogeneity in race and schooling are associated with reductions in the level of real per-pupil spending, while rising within-district income inequality seems to *increase* (by a sizable amount) real per-pupil expenditure. These results are consistent with the idea that heterogeneity in preferences for educational spending correlated with race or educational background (but not necessarily with income) tend to reduce support for public schooling. They also appear to be consistent with the median voter theory outlined in Section 3.2 where changes in the income distribution affect the tax price to the median voter, and the level of educational expenditure demanded. We explore this idea further in columns (6) – (10).

In columns (6) through (10) of Table 3.4, we replace our Gini coefficient measure of income inequality with two other measures—the logged ratio of the 95th centile of income to the median (log(95/50)), and the logged ratio of the median income to the 5th centile (log(50/5)). The former can be considered a measure of income inequality in the top half of the distribution, while the latter measures income inequality in the bottom. Given our stylized model in Section 3.2, increases in income inequality that increase the mean income relative to the median should lower the tax price to the voter with median income, and increase per-pupil spending; increases that reduce the mean income relative to the median should reduce per-pupil spending. We find exactly this pattern in columns (6) through (10)—as the 95th centile of income rises relative to the median within a district, we observe increases in per-pupil spending; as the 5th centile of income falls relative to the median, we observe decreases. Specifically, a one percentage point increase in the 95th centile of income relative to the median is associated with (all else equal) an \$8.20 increase in real per-pupil revenues; a one percentage point increase in the median income relative to the 5^{th} centile of income is associated with a \$1.81 decline in per-pupil revenues. Given that the standard deviations of these variables in our full sample are 0.18 and 0.32, respectively, a two standard deviation increase in the log(95/50) ratio would increase spending per-pupil by \$294; a two standard deviation increase in the log(50/5) ratio would decrease spending by \$116. Our estimates of the coefficients on race and schooling fractionalization in these columns are nearly identical to those in columns (1) through (5).

Clearly, a 0.01 increase in the log(95/50) ratio will not have an equivalent effect on the tax price to the median voter as a 0.01 fall in the log(50/5) ratio. Because households at the top of the income distribution comprise a disproportionate share of total income, an increase in the log(95/50) ratio will add more to mean income than an equivalent decline in the log(50/5). Therefore, to compare the magnitudes of our coefficient estimates on these inequality measures, we performed the following exercise: using the estimated parameters for the income distribution in each district, we computed the fraction of total district income coming from the bottom quartile ($\ell(0.25)$, where $\ell(p)$ is the Lorenz curve for an individual district) and the fraction of total district income comprised by the top quintile ($1-\ell(0.75)$). Averaged across all districts, the upper quartile makes up roughly 47 percent of total income; the bottom quartile makes up about 8 percent of total income. Thus, for the mean district, a hypothetical rise in income of 10 percent among the top quartile of the distribution (with an accompanying rise in the 95/50 ratio of 10 percent) will increase mean income by 4.7 percent. A rise in income of 10 percent among the bottom quartile of the distribution (with an accompanying fall in the 50/5 ratio of 10 percent) will increase mean income by 0.8 percent. In other words, a 10 percent rise in income at the top of the distribution (for the mean district) will have an effect on total income that is roughly 5.8 times that of 10 percent rise in income at the bottom of the distribution. This ratio is similar in magnitude to the ratio of our two regression coefficients on the log(95/50) and the log(50/5): the coefficient on the log(95/50) is 4.5 times that on the log(50/5) ratio, suggesting that our coefficient estimates—in relative terms—are quite reasonable.

Table 3.5 presents the results of the same ten specifications in Table 3.4, estimated on our subsample of districts located in MSAs.[81] Our coefficient estimates for the effect of income inequality on school spending are quite similar to those in Table 3.4. Based on this sample of districts in MSAs, we estimate that a two standard deviation increase in the Gini coefficient corresponds to a \$442 to \$564 increase in per-pupil spending, about the same size effect when considered as a percentage of mean local revenues per pupil (mean per-pupil revenues in our sample of MSA districts is higher than that for our full sample). Increases in the log(95/50), log(50/5), and schooling fractionalization index all have similar effects in the MSA subsample as in the full sample. The primary difference between our estimates using the full sample and our estimates from the MSA subsample is our coefficient estimate for race fractionalization—in contrast to the full sample results, our point estimate for this

[81] In these regressions we use time-specific MSA effects (which subsume state level trends), rather than state effects. The results are not significantly changed when state effects are used.

coefficient is positive, and statistically significant in all specifications except column (5).

In summary, our results so far find a consistent positive relationship between income inequality and local per-pupil revenues, whether income inequality is measured using a Gini coefficient, or through the log(95/50) and log(50/5) ratios. This result holds for our full sample of districts, as well as our subsamples of districts in MSAs. The same is true for heterogeneity in schooling—we find a sizeable negative relationship between fractionalization in educational attainment and per-pupil spending across all samples. We find some evidence that race fractionalization decreases per-pupil revenues in our full sample; the opposite result arises in our subsample of MSAs. In the next section, we consider another possibility—whether heterogeneity in population characteristics or income result in an increase in a change in the fraction of children enrolled in private schools.

3.4.2 Private School Enrollment Share

Table 3.6 presents our coefficient estimates from model (8), where our dependent variable is now the fraction of school age children within a district enrolled in private schools. The order of model specifications is exactly the same as that in Table 3.5, and we continue to include district fixed effects and year-specific state effects in all models. In these regressions, we weight all observations by total K-12 enrollment (i.e. both public and private) in the district.[82]

---

[82] This weight (public plus private enrollment) is more appropriate in this context than total public enrollment. Rather than weighting to estimate the impact of population

We find that in the full sample, the Gini coefficient of income inequality has a negative, statistically significant relationship with the fraction of students enrolled in private school. Our point estimates range from –0.041 to –0.051, with a standard error of 0.011 in all cases. Given that the standard deviation of income inequality in our sample across all years is 0.06, these estimates suggest that a two standard deviation increase in income inequality within a school district is associated with a 0.2 to 0.3 percentage point reduction in the share of school aged children enrolled in private schools. With a mean private school enrollment share of 10 percent, this effect represents about a two to three percent rise in the private school enrollment share over the mean, a small but not insignificant effect.[83]

Our estimated coefficients on our demographic and economic controls are mostly as expected. Private schooling shares tend to be lower in districts where a greater share of the population is nonwhite, in poverty, or are homeowners; private schooling shares rise with the fraction elderly and with real median family income (we find that a $1,000 increase in real median family income is associated with about a 0.2 percentage point increase in private schooling).

---

heterogeneity on the average student (appropriate in our school spending regressions), we weight here to account for heteroskedasticity in our error term (the fraction of students enrolled in private schools is an estimate from the Census 1-in-6 longform data, an estimate that is likely to be more precise in more populated school districts). We have experimented with other weights with these specifications (public enrollment, average total enrollment over the sample period, etc), and the choice of weights has little effect on our results.

[83] This finding is dependent on our use of weights. In the unweighted version of these regressions, we find a positive and statistically significant effect of income inequality on private schooling (in the unweighted version of column (4), we estimate this coefficient to be 0.019, with a standard error of 0.009).

Our results for race and schooling fractionalization are somewhat contradictory. While we find that increased racial fractionalization has a positive, statistically significant relationship with private schooling, we find the opposite is true with fractionalization in schooling. According to Table 3.6, a one standard deviation increase in race fractionalization within districts (0.20 in the full sample) is associated with a 0.36 to 0.38 percentage point increase in the fraction of students enrolled in private schools; a one standard deviation increase in schooling fractionalization (0.07), on the other hand, corresponds with a 0.27 to 0.37 percentage point reduction in the private schooling share.

Table 3.7 presents the same five specifications, estimated using the subsample of districts in MSAs. Here, we find generally the same pattern—increases in within-district income inequality are associated with reductions in the share of students enrolled in private schools, while increases in racial fractionalization increase the private school enrollment share. However, in this case, our estimate for the coefficient on income inequality is imprecise—in no column in Table 3.7 is the negative coefficient on income inequality statistically significant.

3.4.3 The Influence of School District Competition

In Sections 3.4.1 and 3.4.2, we found that rising income inequality within school districts over the 1970-2000 period has been associated with higher local per-pupil revenues and lower private school enrollment ratios (although the latter effect is quite small and is statistically insignificant in our subsample of districts in MSAs). Greater heterogeneity in race shows a consistent and statistically significant negative relationship with per-pupil spending, and appears to have a positive impact on private school enrollment rates. Heterogeneity in schooling seems to be associated with lower per-pupil spending (across all samples), but also appears to lower private school enrollment rates, a result that is not easily explained by our stylized model in Section 3.2.

As we explained in Section 3.3, where mobility between school districts is possible, households may respond to rising within-district heterogeneity in income, race or schooling through exit. To the extent that this sorting occurs, we might expect to see our results weakened in school districts located in regions with a higher degree of choice.

To test this idea, we estimate model (8) again on our subsample of districts in MSAs, but now interact our Gini coefficient of income inequality or fractionalization indices with various indicators of local school district competition.[84] As Appendix B explains in greater detail, we constructed eight different measures of area school district competition—the count of public school districts within a 10-, 25-, and 50-mile radius (*dists10, dists25,* and *dists50*), the fraction of enrollment within a 10-, 25- and 50-mile radius that are enrolled in district *i* (*frac10, frac25,* and *frac50*), the number of districts per student in the same MSA (*diststud*), and finally (following Hoxby (2000)) an enrollment fractionalization index for the MSA (or, one minus a Herfindahl index using the enrollment shares of all districts within an MSA). We then placed districts into quintiles based on their placement in the distribution of these eight

[84] School district competition is generally a meaningful concept only within metropolitan areas. Therefore, we present these results only for districts in MSAs.

measures (in 2000), and interacted our Gini coefficient or fractionalization index with quintile dummies (the lowest quintile dummy is excluded in all regressions). The results of this exercise are in Table 3.8.[85]

Our results from this experiment are mixed, and vary across our different measures of public school district competition. When the number of districts within 25 or 50 miles is used as our measure of area competition, we find that the coefficient on income inequality is the largest in districts that face the least competition (with point estimates of $7,700 to $7,948), and lower in districts in higher quintiles of competition, although the pattern is not monotonically decreasing in the level of competition (nor is this pattern observed when other measures of competition are used). Similarly, the negative coefficient on schooling fractionalization is highest in districts with the lowest levels of competition (as measured by *dists25* and *dists50*); this coefficient gets smaller in magnitude in districts with more competition (although the coefficients on the interaction variables are not always statistically significant).

When using district *i*'s share of total enrollment within 10-, 25-, and 50-miles as our measure of competition, we find that the negative coefficient on race fractionalization is largest in districts with the least competition;[86] this negative

[85] In panel (1) of Table 3.8, we estimate a specification like that of column (1) in Table 3.6, but add interactions between the Gini coefficient and competition quintile dummies. In panels (2) and (3), we estimate a regression model similar to that of column (4) in Table 3.6, but add interactions between our race and schooling fractionalization indices and competition quintile dummies. Only the coefficient estimates on the interaction variables are reported.

[86] Quintiles of the *frac10, frac25,* and *frac50* variables are read in the opposite direction—when a greater fraction of enrollment within *m* miles is enrolled in district *i,* we think of district *i* as facing less competition.

coefficient is almost entirely offset in districts facing higher levels of competition. No discernible pattern is apparent, however, when this measure of competition is used with income inequality or schooling fractionalization. Likewise, this trend in coefficients is not observed under other measures of school district competition.

Our mixed results in Table 3.8 may be due to imperfect measures of school district competition, or may arise because Tiebout-sorting across districts is not strong enough to have a clear impact on our results. In Section 3.4.4 we attempt to further deal with the potential problem of sorting across districts by repeating our analysis in Section 3.4.1 and 3.4.2 with a subsample of low-mobility districts and a sample of districts residing outside of urban areas.

3.4.4 A Robustness Check: Low Mobility Districts and Districts Outside of MSAs

In this section, we repeat our analysis from Section 3.4.1 and 3.4.2, but with two additional subsamples—the first, a sample of "low mobility" districts, with a relatively low proportion of new residents, and the second a sample of districts that lie outside of metropolitan areas. To construct our "low mobility" sample, we selected all of those districts that were in the lowest three quintiles of household migration—that is, with the lowest proportion of households that lived in a different county five years prior to the 1980, 1990 and 2000 censuses.[87] 3,955 districts met our criteria for

[87] Specifically, we placed each district into a quintile of migration (the fraction of households who lived in a different county five years before) for 1980, 1990, and 2000 (migration information was not available in the 1970 census tabulation). If a district was in the lowest three quintiles of this distribution *in all three census years*, we included them in our subsample.

inclusion in this subsample.[88] These low mobility districts were not strictly rural, unpopulated districts—in fact, over 40 percent of the districts in our low mobility sample were located in MSAs.

Our findings are summarized in Table 3.9. Coefficient estimates on income inequality, race, and schooling fractionalization are qualitatively similar to those found for the full sample and MSA sample in Section 3.4.1 and 3.4.2. A two standard deviation increase in income inequality is associated with a $308 increase in per-pupil spending in our low mobility sample (about 13.8 percent of the sample mean in low mobility districts, a slightly smaller effect in percentage terms than the full sample result), and a $180 increase in per-pupil spending in our non-MSA sample (about 9.3 percent of the non-MSA sample mean). As in our MSA sample, there is no statistically significant impact of income inequality on the private schooling share in either our low mobility sample or our non-MSA sample. Schooling fractionalization continues to have a negative relationship with the private school enrollment share, while fractionalization in race continues to have a positive relationship with private schooling (although this coefficient is statistically insignificant in our non-MSA subsample).

It appears that our general findings from Sections 3.4.1 and 3.4.2 are robust to varying definitions of our underlying school district sample—with few exceptions, we find the same pattern of results across these subsamples.

[88] The mean fraction of households who lived in a different county, by quintile in 2000 were 9 percent, 12.7 percent, 15.7 percent, 19 percent, and 26.9 percent.

3.4.5 Instrumental Variables Results

Finally, in this section, we estimate our per-pupil spending regression in (8) using an instrumental variables model in which we instrument for within-district income inequality with a measure of inequality from another nearby school district. As we discussed in Section 3.3, our OLS estimates of the impact of income inequality on local per-pupil revenues may be biased and inconsistent if households sort into districts based on the level of per-pupil revenues—it may be that changes in income inequality over time within a school district are due in part to the in- and out-migration of families who are responding to changes in local school finance policies.

We address this possibility by estimating three instrumental variables models, where our instruments for within-district income inequality are the Gini coefficient for the closest school district, the Gini coefficient for the closest district in another county, and the Gini coefficient for the closest district in another state.[89] To the extent that changes in income inequality within a school district over time are due to (exogenous) regional labor market conditions, we would expect that the level of income inequality in a district and that of its neighboring districts to be highly correlated. It is this portion of variation in within-district income inequality that we would like to use to identify the impact of income inequality on local per-pupil revenues—not the variation in inequality that is a mechanical response to inter-district sorting. Of course, the problem is identifying districts that are close enough to be subject to the same economic conditions, but far enough away to preclude any sorting effects.

[89] More information about the construction of these instrumental variables can be found in Appendix B.

In Figure 3.2, we illustrate the spatial distribution of within-district income inequality across school districts in the United States.[90] Each block within this figure is a unified, elementary or secondary school district; shading within each block indicates the school district's quartile in the nationwide distribution of within-district income inequality, where darker shading represents a higher degree of income inequality. This figure suggests that the level of within-district income inequality is in many parts of the United States a regional phenomenon, with the level of income inequality highly correlated across neighboring districts. Much of the South, Appalachia, the Northeast corridor, Texas and California is characterized by smooth patterns of regional within-district inequality—not the checkerboard pattern that one might see if there were high degrees of sorting occurring between districts (the Upper Midwest is an exception to this rule—while some of this pattern in the Upper Midwest can be explained by missing data, the degree of spatial correlation in these districts is certainly lower).

The results of our first-stage regressions shown in column (1) of Table 3.10 confirm this pattern. The estimated coefficients on our three instruments are precisely estimated and have the correct sign—within-district income inequality is highly correlated with income inequality in the nearest district, nearest district in another county, and nearest district in another state, and the first-stage coefficient falls monotonically with distance.[91] Column (2) reports our instrumental variables estimate

[90] All school districts (unified, elementary, and secondary) and included in this Figure. Some districts (but not all) without shading are missing income inequality data.

[91] Note *t*-statistics are reported in this table rather than standard errors. Our first- and second-stage regressions both incorporate observation weights.

of the coefficient on income inequality from specification (4) in Table 3.4. With our first two instruments, our IV estimates of the coefficient on income inequality ($2,980 and $2,837) are only slightly less than our original point estimate from column (4) in Table 3.4 ($3,512), and well within a 95 percent confidence interval around our original estimate. Our third instrument (income inequality within the nearest district in another state) produces an IV estimate that is less than half that produced by the first two instruments ($1,167).

3.5 Summary and Conclusion

In this chapter, we have used a panel of unified school districts to estimate the effects of local population heterogeneity and income inequality on the support for public schools. We measured support for schools using two variables—local per-pupil education revenues and the fraction of district students enrolled in private schools. Our results suggest that rising income inequality within a district may actually increase per-pupil revenues, a result consistent with a median voter model in which rising income inequality lowers the tax price to the voter with the median income (which occurs when the mean income rises relative to the median income). The estimated impact of income inequality on per-pupil revenues is sizable—about a $200 (or 8 percent) per-pupil rise in local expenditure for a one standard deviation increase in income inequality—and robust across a number of specifications. We also find that, all else equal, increased fractionalization in race tends to reduce local per-pupil expenditure on education, and increase the fraction of school aged children in private school. Fractionalization in schooling also tends to reduce educational spending; it also tends to be associated with lower private school enrollment rates—a result that clearly deserves further study.

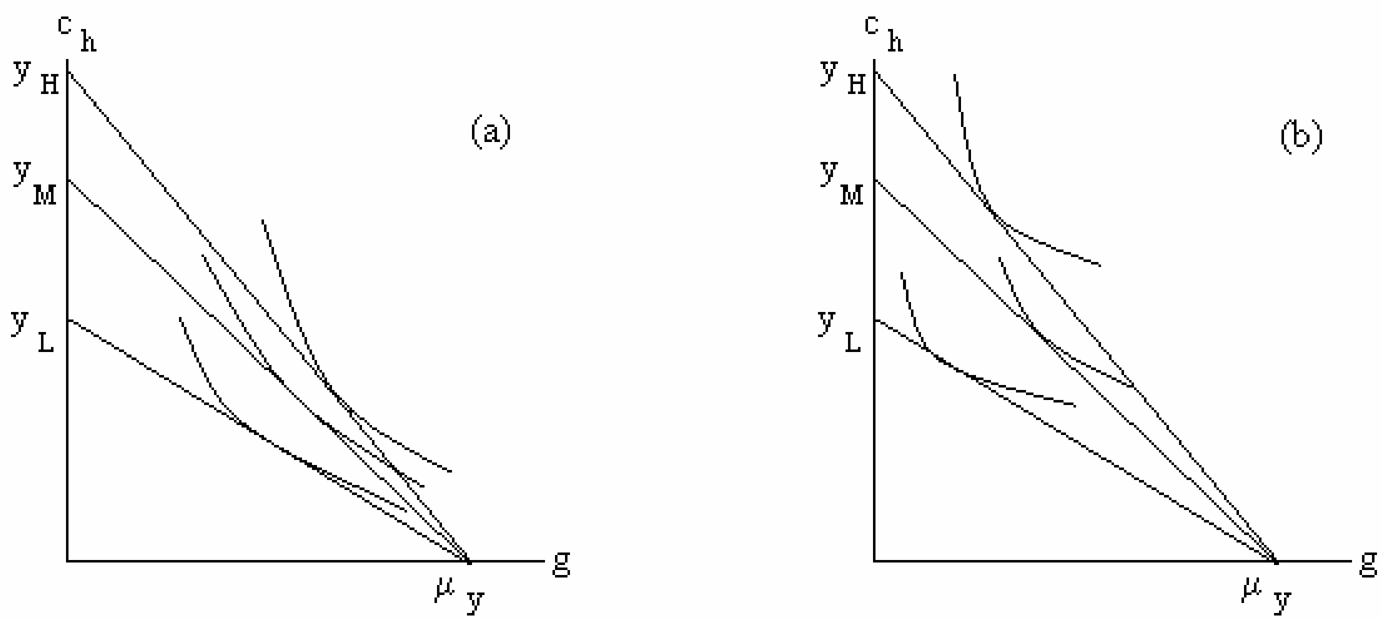
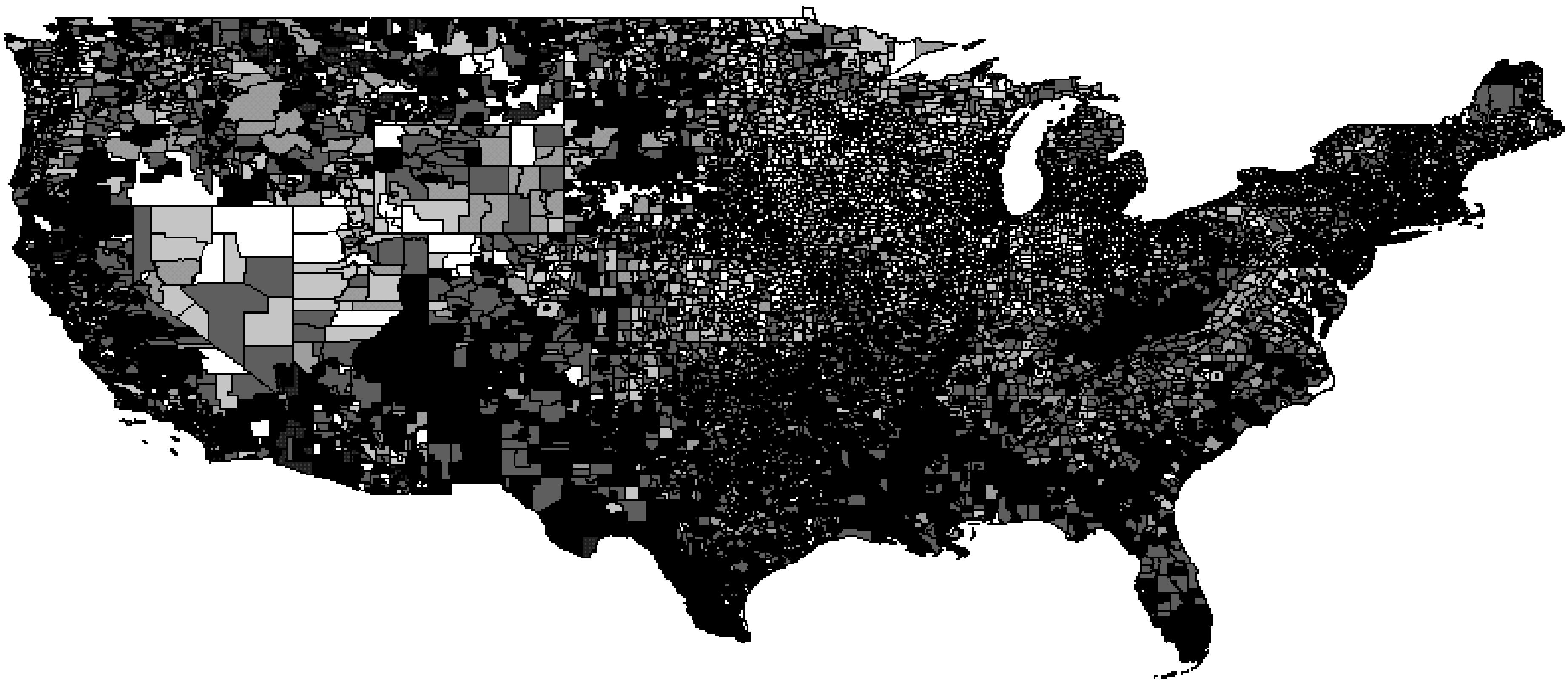


Figure 3.1: Income and the Demand for Per-Pupil Expenditure

Figure 3.2: Spatial Distribution of 2000 Income Inequality—Elementary, Secondary and Unified School Districts



Source: 2000 Census of Population and Housing School District Tabulation (STP2)

Table 3.1: Descriptive Statistics for Full Sample

	All Years		1970		1980		1990		2000	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Gini coefficient	0.36	0.06	0.34	0.05	0.35	0.05	0.38	0.05	0.39	0.06
Schooling fractionalization	0.68	0.07	0.62	0.08	0.68	0.05	0.71	0.03	0.71	0.04
Race fractionalization	0.25	0.20	0.17	0.17	0.22	0.19	0.26	0.20	0.33	0.20
Log(95/50) ratio	0.97	0.18	0.88	0.16	0.92	0.15	1.00	0.17	1.06	0.18
Log(50/5) ratio	1.59	0.32	1.51	0.28	1.57	0.28	1.66	0.37	1.61	0.31
Real local revenues per pupil	2,238.88	1,436.64	1,861.13	1,029.60	1,847.05	1,116.59	2,520.25	1,691.41	2,736.23	1,592.98
Real state revenues per pupil	2,305.86	1,140.92	1,377.46	577.81	2,004.72	743.27	2,736.82	1,088.42	3,134.63	1,128.50
Real federal revenues per pupil	143.58	242.41	79.41	185.37	22.78	66.58	23.41	84.38	421.61	264.83
Real total revenues per pupil	4,688.32	1,757.46	3,318.00	1,091.00	3,874.54	1,049.20	5,280.49	1,527.98	6,292.46	1,453.48
Local share of total revenues	0.47	0.19	0.54	0.18	0.46	0.19	0.45	0.20	0.42	0.18
Fraction enrolled in private school	0.10	0.07	0.10	0.08	0.10	0.07	0.10	0.07	0.10	0.06
Real median family income (thousands)	39.38	11.60	36.47	9.70	38.52	9.52	39.62	12.16	42.83	13.54
Percent of households in poverty	0.13	0.08	0.14	0.09	0.13	0.07	0.13	0.08	0.12	0.07
Percent high school dropouts	0.32	0.16	0.48	0.13	0.34	0.13	0.26	0.11	0.20	0.10
Percent high school grads only	0.31	0.08	0.31	0.07	0.35	0.07	0.31	0.08	0.29	0.08
Percent with some college	0.19	0.09	0.10	0.04	0.15	0.05	0.24	0.06	0.27	0.06
Percent college grads or higher	0.17	0.11	0.10	0.07	0.15	0.08	0.19	0.10	0.23	0.12
Percent nonwhite	0.23	0.22	0.16	0.17	0.20	0.20	0.24	0.23	0.30	0.25
Percent black	0.12	0.15	0.11	0.14	0.12	0.15	0.12	0.16	0.12	0.15
Percent hispanic	0.08	0.14	0.05	0.10	0.06	0.12	0.09	0.15	0.13	0.18
Percent of housing units owner occupied	0.67	0.15	0.66	0.15	0.67	0.15	0.66	0.14	0.68	0.14
Share of district population aged 65+	0.11	0.04	0.10	0.04	0.11	0.04	0.12	0.04	0.12	0.04
Share of district population aged 0-19	0.33	0.06	0.39	0.05	0.33	0.04	0.29	0.04	0.29	0.04
Percent urban	0.72	0.35	0.70	0.36	0.70	0.36	0.71	0.35	0.77	0.31
Share of households with children	0.41	0.10	0.43	0.10	0.48	0.08	0.39	0.07	0.35	0.07

The full sample consists of 8,699 districts for a total of 34,796 observations. All monetary values are in real 1992 dollars. Observations are weighted by district public enrollment.

Table 3.2: Descriptive Statistics for MSA Subsample

	All Years		1970		1980		1990		2000	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Gini coefficient	0.36	0.06	0.33	0.05	0.35	0.05	0.37	0.06	0.39	0.06
Schooling fractionalization	0.69	0.05	0.65	0.06	0.70	0.04	0.72	0.03	0.71	0.04
Race fractionalization	0.27	0.20	0.18	0.17	0.24	0.19	0.30	0.20	0.36	0.20
Log(95/50) ratio	0.96	0.19	0.86	0.14	0.90	0.15	1.00	0.18	1.08	0.20
Log(50/5) ratio	1.57	0.32	1.44	0.25	1.57	0.30	1.66	0.39	1.62	0.32
Real local revenues per pupil	2,477.32	1,511.99	2,086.06	1,045.29	2,012.04	1,142.70	2,817.00	1,809.03	3,001.09	1,677.53
Real state revenues per pupil	2,279.66	1,170.99	1,333.62	556.88	2,012.39	781.64	2,709.84	1,144.69	3,094.32	1,179.50
Real federal revenues per pupil	143.70	237.47	90.32	203.06	21.92	46.99	20.67	61.04	407.96	255.51
Real total revenues per pupil	4,900.68	1,805.41	3,510.00	1,134.36	4,046.35	1,045.86	5,547.52	1,588.72	6,503.36	1,513.36
Local share of total revenues	0.50	0.19	0.58	0.16	0.48	0.18	0.48	0.21	0.45	0.18
Fraction enrolled in private school	0.12	0.07	0.13	0.08	0.12	0.07	0.12	0.07	0.12	0.05
Real median family income (thousands)	42.56	11.74	39.85	8.90	41.37	9.49	43.14	12.46	45.82	14.31
Percent of households in poverty	0.11	0.07	0.11	0.07	0.11	0.07	0.12	0.08	0.11	0.07
Percent high school dropouts	0.30	0.15	0.45	0.12	0.31	0.11	0.23	0.10	0.19	0.10
Percent high school grads only	0.31	0.07	0.32	0.06	0.35	0.07	0.29	0.07	0.27	0.08
Percent with some college	0.20	0.08	0.11	0.04	0.17	0.04	0.26	0.06	0.28	0.05
Percent college grads or higher	0.19	0.11	0.12	0.07	0.17	0.09	0.22	0.10	0.26	0.12
Percent nonwhite	0.25	0.23	0.17	0.17	0.22	0.21	0.27	0.23	0.34	0.25
Percent black	0.12	0.15	0.11	0.13	0.13	0.15	0.13	0.16	0.13	0.16
Percent hispanic	0.09	0.15	0.05	0.10	0.07	0.13	0.11	0.16	0.15	0.18
Percent of housing units owner occupied	0.64	0.16	0.64	0.17	0.64	0.16	0.63	0.15	0.65	0.15
Share of district population aged 65+	0.10	0.04	0.09	0.04	0.10	0.04	0.11	0.04	0.11	0.04
Share of district population aged 0-19	0.32	0.06	0.39	0.05	0.32	0.04	0.29	0.04	0.29	0.04
Percent urban	0.86	0.25	0.84	0.27	0.84	0.26	0.86	0.25	0.89	0.20
Share of households with children	0.41	0.10	0.44	0.11	0.48	0.09	0.38	0.08	0.35	0.07

The MSA subsample consists of 3,292 districts for a total of 13,168 observations. Districts are included in this sample if they resided in a county that was located in a metropolitan area, according to 1973 standards. All dollar values are in real 1992 dollars. Observations are weighted by district public enrollment.

Table 3.3: 1970-2000 Changes in Income Inequality and Fractionalization (Full Sample)

	Percent Change in Gini Coefficient 1970 - 2000	Absolute Change in Race Fractionalization 1970 - 2000	Absolute Change in Schooling Fractionalization 1970 - 2000
Unweighted			
Mean	0.070	0.090	0.110
5th centile	-0.161	-0.030	-0.023
10th centile	-0.115	0.003	0.018
25th centile	-0.032	0.023	0.058
Median	0.065	0.053	0.098
75th centile	0.166	0.126	0.160
90th centile	0.259	0.253	0.242
95th centile	0.314	0.340	0.286
Weighted			
Mean	0.152	0.165	0.086
5th centile	-0.075	-0.033	-0.068
10th centile	-0.027	0.010	-0.015
25th centile	0.059	0.047	0.036
Median	0.153	0.133	0.079
75th centile	0.245	0.271	0.132
90th centile	0.327	0.383	0.207
95th centile	0.352	0.442	0.262

The full sample consists of 8,699 districts for a total of 34,796 observations. In the bottom half of the table, observations have been weighted by district public enrollment.

Table 3.4: OLS Results, Real Local Revenues per Pupil (Full Sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
gini coefficient	3027.949	3567.246	2972.546	3512.481	3166.224					
	(224.955)	(226.650)	(225.959)	(227.762)	(226.512)					
schooling fractionalization index		-2567.613		-2536.565	-2794.346				-2318.540	-2584.821
		(166.517)		(167.268)	(166.751)				(169.193)	(168.729)
race fractionalization index			-214.596	-181.801	-221.404				-130.935	-172.976
			(81.701)	(81.369)	(80.726)				(81.702)	(81.043)
log(95/50) ratio						817.549		817.551	901.052	821.097
						(60.366)		(60.329)	(60.933)	(60.559)
log(50/5) ratio							-181.953	-181.955	-96.071	-101.325
							(31.585)	(31.474)	(32.103)	(31.862)
%in poverty	-2482.635	-3264.822	-2531.773	-3299.914	-2988.629	-2130.684	-935.398	-1567.377	-2504.377	-2227.480
	(182.108)	(188.247)	(183.476)	(189.559)	(188.611)	(173.191)	(193.752)	(198.626)	(209.100)	(207.833)
real median family income (in thousands	53.794	46.158	53.663	46.112	43.327	54.428	49.171	54.182	47.087	44.366
	(1.485)	(1.559)	(1.490)	(1.565)	(1.587)	(1.495)	(1.454)	(1.495)	(1.577)	(1.600)
%25+ with hs degree	-3033.690	-1223.948	-3095.914	-1299.337	-1055.264	-3014.752	-3067.578	-2945.000	-1390.950	-1139.183
	(134.275)	(177.881)	(135.458)	(179.507)	(178.676)	(134.354)	(134.979)	(134.811)	(179.273)	(178.441)
%25+ with some college	-3036.286	-1578.145	-3078.951	-1625.641	-1594.207	-3074.307	-3444.511	-3068.799	-1802.402	-1746.642
	(149.977)	(176.726)	(150.774)	(178.093)	(177.702)	(149.385)	(147.214)	(149.295)	(177.177)	(176.787)
%25+ with college degree or more	541.574	1295.409	647.875	1397.811	1543.837	541.495	1170.399	650.531	1413.223	1560.989
	(188.450)	(193.862)	(189.597)	(195.130)	(193.803)	(188.395)	(185.941)	(189.220)	(195.737)	(194.360)
%nonwhite	-675.257	-417.098	-564.734	-326.214	-208.826	-627.699	-485.005	-537.525	-277.585	-162.267
	(80.052)	(81.429)	(98.048)	(98.874)	(98.356)	(79.797)	(81.451)	(81.258)	(99.363)	(98.806)
%owner-occupied housing	-2022.990	-1739.002	-2120.288	-1835.229	-1135.103	-2041.469	-1975.310	-2004.006	-1848.083	-1145.728
	(112.099)	(113.100)	(114.751)	(115.780)	(121.393)	(112.111)	(112.603)	(112.229)	(115.847)	(121.453)
%population 65+	3945.198	3890.683	3906.324	3848.956	988.935	4020.055	4228.813	4060.944	3950.442	1072.111
	(217.168)	(216.213)	(220.613)	(219.672)	(262.010)	(216.763)	(217.150)	(216.744)	(219.452)	(262.020)
%hh with kids					-2131.128					-2136.685
					(109.398)					(109.377)
%population in urban area					-315.219					-332.743
					(41.576)					(41.599)
Constant	1446.526	2226.333	1574.761	2331.974	3512.677	1676.509	2594.959	1802.047	2618.816	3778.438
	(147.113)	(154.932)	(149.273)	(156.778)	(165.600)	(138.901)	(128.189)	(140.502)	(151.397)	(160.015)
Observations	34,795	34,795	34,620	34,620	34,620	34,795	34,795	34,795	34,620	34,620
Adjusted R^2	0.8616	0.8629	0.8976	0.8637	0.8659	0.8616	0.8962	0.8618	0.8637	0.8659

The sample consists of a balanced panel of 8,699 unified public school districts located in 48 U.S. states, for the years 1970, 1980, 1990, and 2000. All specifications include district fixed effects and year-specific state effects. Observations have been weighted by district public enrollment. Standard errors are in parentheses.

Table 3.5: OLS Results, Real Local Revenues per Pupil (MSA Sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
gini coefficient	3685.229	4531.783	3889.188	4706.562	4231.799					
	(435.097)	(438.988)	438.210	441.922	441.342					
schooling fractionalization index		-3754.109		-3695.838	-3803.477				-3469.368	-3568.159
		(340.090)		341.410	340.602				(343.427)	(342.617)
race fractionalization index			315.7557	292.3407	181.181				386.9414	271.3389
			139.292	138.424	137.916				(139.282)	(138.745)
log(95/50) ratio						984.854		1027.915	1304.26	1192.615
						(119.381)		(119.648)	(121.812)	(121.415)
log(50/5) ratio							-271.2465	-321.1577	-253.4194	-274.3108
							(72.053)	(72.001)	(72.721)	(72.377)
%in poverty	-5140.159	-5730.376	-5089.314	-5672.242	-4843.792	-4556.552	-2436.957	-3206.573	-3829.93	-3000.027
	(430.656)	(431.153)	(433.281)	(433.882)	(438.358)	(407.534)	(501.308)	(507.287)	(515.469)	(517.527)
real median family income (in thousands	26.347	14.54663	26.25154	14.6048	14.29247	27.301	22.96902	27.62077	16.84638	16.80965
	(2.579)	(2.776)	(2.586)	(2.786)	(2.859)	(2.602)	(2.554)	(2.601)	(2.808)	(2.884)
%25+ with hs degree	-3724.542	-1131.218	-3642.823	-1096.829	-851.493	-3702.848	-3588.624	-3536.498	-1085.003	-841.3435
	(278.334)	(362.834)	(279.757)	(364.125)	(364.302)	(278.407)	(281.662)	(280.608)	(363.976)	(363.964)
%25+ with some college	-3888.257	-1052.375	-3782.982	-997.0155	-1129.456	-3941.223	-4436.519	-3872.899	-1166.55	-1257.938
	(283.336)	(381.088)	(284.524)	(382.311)	(383.321)	(282.307)	(275.802)	(282.430)	(380.191)	(381.271)
%25+ with college degree or more	363.332	1511.418	404.1174	1534.511	1839.34	348.407	965.3603	510.0054	1548.377	1862.301
	(341.018)	(354.382)	(343.708)	(357.130)	(356.013)	(341.600)	(340.399)	(343.164)	(358.226)	(357.019)
%nonwhite	-1520.788	-1304.978	-1768.74	-1535.767	-1319.796	-1493.100	-1447.24	-1402.302	-1511.869	-1294.81
	(151.794)	(152.058)	(183.088)	(183.193)	(183.961)	(151.867)	(153.590)	(153.069)	(183.590)	(184.279)
%owner-occupied housing	-1582.617	-1246.164	-1604.787	-1274.594	-353.7409	-1639.867	-1653.105	-1636.604	-1348.155	-428.2555
	(206.170)	(207.070)	(212.249)	(213.094)	(236.213)	(206.041)	(206.642)	(205.829)	(212.650)	(235.924)
%population 65+	5290.021	5321.697	5192.584	5245.644	2527.209	5405.570	5588.27	5478.383	5388.6	2650.369
	(384.262)	(381.747)	(397.484)	(394.988)	(476.903)	(383.619)	(384.886)	(383.569)	(394.365)	(476.632)
%hh with kids					-1991.623					-1987.403
					(209.181)					(209.016)
%population in urban area					-429.4353					-469.2901
					(79.781)					(79.763)
Constant	2733.165	3853.622	2625.426	3736.48	4636.67	3026.195	4122.569	3182.752	4125.287	5006.741
	(271.269)	(287.968)	(275.063)	(291.950)	(301.776)	(255.865)	(234.592)	(257.999)	(281.183)	(290.215)
Observations	13,171	13,171	13,120	13,120	13,120	13,171	13,171	13,171	13,120	13,120
Adjusted R^2	0.8848	0.8863	0.8852	0.9218	0.8881	0.8847	0.884	0.8849	0.8867	0.8882

The sample consists of a balanced panel of 3,292 unified public school districts located in MSAs in 1973, for the years 1970, 1980, 1990, and 2000. All specifications include district fixed effects and year-specific MSA effects. Observations have been weighted by district public enrollment. Standard errors are in parentheses.

Table 3.6: OLS Results, Private School Enrollment Share (Full Sample)

	(1)	(2)	(3)	(4)	(5)
gini coefficient	-0.051	-0.042	-0.049	-0.041	-0.050
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
schooling fractionalization index		-0.039		-0.040	-0.053
		(0.008)		(0.008)	(0.008)
race fractionalization index			0.018	0.019	0.018
			(0.004)	(0.004)	(0.004)
%in poverty	-0.005	-0.017	-0.005	-0.017	-0.010
	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)
real median family income (in thousands)	0.002	0.002	0.002	0.002	0.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
%25+ with hs degree	-0.079	-0.051	-0.077	-0.048	-0.036
	(0.007)	(0.009)	(0.007)	(0.009)	(0.009)
%25+ with some college	-0.018	0.005	-0.016	0.008	-0.004
	(0.007)	(0.009)	(0.007)	(0.009)	(0.009)
%25+ with college degree or more	-0.073	-0.061	-0.074	-0.062	-0.063
	(0.009)	(0.010)	(0.009)	(0.010)	(0.010)
%nonwhite	-0.053	-0.049	-0.064	-0.061	-0.054
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
%owner-occupied housing	-0.038	-0.033	-0.034	-0.030	0.001
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
%population 65+	0.126	0.125	0.121	0.120	0.020
	(0.011)	(0.011)	(0.011)	(0.011)	(0.013)
%hh with kids					-0.077
					(0.005)
%population in urban area					0.017
					(0.002)
Constant	0.105	0.116	0.101	0.112	0.140
	(0.007)	(0.008)	(0.007)	(0.008)	(0.008)
Observations	34,793	34,793	34,618	34,618	34,618
Adjusted R^2	0.8763	0.8764	0.8762	0.8763	0.8776

The sample consists of a balanced panel of 8,699 unified public school districts located in 48 U.S. states, for the years 1970, 1980, 1990, and 2000. All specifications include district fixed effects and year-specific state effects. Observations have been weighted by total (public + private) enrollment in district. Standard errors are in parentheses.

Table 3.7: OLS Results, Private School Enrollment Share (MSA Sample)

	(1)	(2)	(3)	(4)	(5)
gini coefficient	-0.005	-0.002	-0.004	-0.001	-0.010
	(0.020)	(0.020)	(0.020)	(0.021)	(0.021)
schooling fractionalization index		-0.013		-0.012	-0.028
				(0.016)	(0.016)
race fractionalization index			0.011	0.011	0.008
			(0.006)	(0.006)	(0.006)
%in poverty	-0.031	-0.033	-0.029	-0.031	0.002
	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)
real median family income (in thousands)	0.002	0.002	0.002	0.002	0.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
%25+ with hs degree	-0.133	-0.123	-0.132	-0.123	-0.100
	(0.013)	(0.017)	(0.013)	(0.017)	(0.017)
%25+ with some college	0.021	0.032	0.024	0.033	0.007
	(0.013)	(0.018)	(0.013)	(0.018)	(0.018)
%25+ with college degree or more	-0.079	-0.075	-0.082	-0.078	-0.067
	(0.016)	(0.017)	(0.016)	(0.017)	(0.017)
%nonwhite	-0.031	-0.031	-0.039	-0.038	-0.024
	(0.007)	(0.007)	(0.008)	(0.009)	(0.009)
%owner-occupied housing	0.000	0.001	0.004	0.005	0.057
	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)
%population 65+	0.206	0.206	0.203	0.203	0.086
	(0.018)	(0.018)	(0.018)	(0.018)	(0.022)
%hh with kids					-0.098
					(0.010)
%population in urban area					0.020
					(0.004)
Constant	0.088	0.092	0.085	0.088	0.106
	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)
Observations	13,171	13,171	13,120	13,120	13,120
Adjusted R^2	0.8963	0.8963	0.8963	0.8963	0.8977

The sample consists of a balanced panel of 3,292 unified public school districts located in MSAs in 1973, for the years 1970, 1980, 1990, and 2000. All specifications include district fixed effects and year-specific MSA effects. Observations have been weighted by total (public + private) enrollment in district. Standard errors are in parentheses.

Table 3.8: OLS Results--Real Local Revenues per Pupil: Interactions of Gini Coefficient and Fractionalization with Quintiles of District Competition (MSA Sample)

	competition measure							
(1)	dists10	dists25	dists50	frac10	frac25	frac50	enrlfrac	diststud
gini coefficient	3,448.782	7,948.584	7,699.180	4,709.360	5,154.200	6,161.835	1,153.928	5,068.279
	(797.721)	(718.294)	(733.192)	(607.692)	(639.985)	(737.329)	(819.161)	(570.521)
gini * 2nd quintile of competition measure	3,266.309	-4,829.156	-4,459.694	-1,535.712	-1,429.748	-2,134.445	-4,701.060	-5,019.806
	(851.355)	(1003.919)	(976.305)	(623.310)	(705.917)	(854.192)	(433.502)	(784.707)
gini * 3rd quintile of competition	262.025	-7,699.066	-2,794.989	747.541	-852.726	-3,066.431	26.742	-1,592.150
	(865.526)	(893.629)	(851.329)	(647.136)	(740.223)	(841.058)	(2.580)	(955.116)
gini * 4th quintile of competition	-2,146.291	-4,397.862	-2,540.098	-2,129.882	-2,966.919	-2,848.194	-3,655.120	-277.420
	(817.784)	(800.979)	(836.437)	(644.202)	(714.086)	(826.730)	(279.482)	(940.348)
gini * 5th quintile of competition	57.459	-5,233.352	-8,769.054	-2,640.082	-1,753.668	-3,178.816	-3,654.620	-23.987
	(794.163)	(808.620)	(931.431)	(601.074)	(645.583)	(767.766)	(284.692)	(1172.456)

	competition measure							
(2)	dists10	dists25	dists50	frac10	frac25	frac50	enrlfrac	diststud
schooling fractionalization	-4186.589	-4211.917	-4789.185	-2575.528	-2266.518	-2187.080	-6607.061	-4440.995
	(450.066)	(467.083)	(454.013)	(439.854)	(475.753)	(509.463)	(521.260)	(382.833)
schooling fractionalization * 2nd quintile of competition measure	1774.865	742.898	2118.742	551.524	-99.941	745.831	2844.812	995.857
	(484.869)	(586.074)	(578.104)	(453.943)	(502.304)	(548.127)	(567.785)	(470.913)
schooling fractionalization * 3rd quintile of competition measure	2111.798	1813.151	3163.463	13.029	-626.847	-687.801	1872.055	336.311
	(503.314)	(577.602)	(558.717)	(485.973)	(505.815)	(533.379)	(566.018)	(480.557)
schooling fractionalization * 4th quintile of competition measure	1466.830	521.629	2443.922	-1007.782	-1329.263	-960.685	5481.199	3575.544
	(481.998)	(506.847)	(529.146)	(439.316)	(485.027)	(501.094)	(597.724)	(522.354)
schooling fractionalization * 5th quintile of competition measure	35.305	638.055	120.720	-2917.969	-2363.871	-2482.182	4828.460	801.004
	(458.475)	(491.535)	(499.946)	(430.251)	(438.652)	(470.896)	(581.722)	(720.810)

	competition measure							
(3)	dists10	dists25	dists50	frac10	frac25	frac50	enrlfrac	diststud
race fractionalization	161.871	324.472	172.014	623.658	1278.585	2357.504	-1461.411	235.740
	(229.793)	(245.801)	(253.978)	(192.596)	(220.317)	(293.509)	(318.244)	(177.847)
race fractionalization * 2nd quintile of competition measure	515.583	138.046	349.862	-391.352	-979.328	-1704.179	1997.004	-417.220
	(271.202)	(349.492)	(350.120)	(197.396)	(220.983)	(317.949)	(349.003)	(251.774)
race fractionalization * 3rd quintile of competition measure	93.391	-803.671	-3.785	-187.345	-849.721	-2056.477	1348.209	-474.098
	(236.039)	(289.424)	(289.494)	(212.403)	(240.219)	(292.880)	(345.515)	(298.805)
race fractionalization * 4th quintile of competition measure	-511.008	-239.177	724.441	-689.580	-1267.762	-2138.053	1826.615	706.042
	(232.951)	(272.324)	(281.514)	(193.466)	(223.693)	(294.288)	(374.814)	(362.165)
race fractionalization * 5th quintile of competition measure	366.098	45.640	-363.160	-951.671	-1328.140	-2417.723	2886.449	1902.613
	(237.441)	(267.854)	(316.530)	(194.441)	(208.485)	(280.618)	(357.700)	(364.604)

Standard errors in parentheses. Panel (1) displays the results of specification (1) in Table 3.6, allowing the coefficient on the within-district Gini coefficient to vary across quintiles of regional public school district competition. We use six different measures of district competition: dists10, dists25, and dists50 (indicating the number of districts within 10, 25, and 50 miles respectively), and frac10, frac25, and frac50 (indicating the fraction of public school students within 10, 25, and 50 miles that are enrolled in that particular district). In columns (1)-(3), higher quintiles of dists10, dists25, and dists50 represent districts in areas with greater competition; in column (4)-(6), higher quintiles of frac10, frac25, and frac50 represent districts in areas with less competition. Panel (2) and (3) display the results of specification (4) in Table 3.6, allowing the coefficient on within-district schooling and race fractionalization to both vary across quintiles of district competition (the gini coefficient enters linearly into these regressions).

Table 3.9: OLS Results--Low Mobility Districts and Districts Outisde of MSAs

Dependent variable	(1) Local Revenues per Pupil	(2) Local Revenues per Pupil	(3) Private School Enrollment Share	(4) Private School Enrollment Share
Sample	Low Mobility	non-MSA	Low Mobility	non-MSA
gini coefficient	2,969.142	1,952.237	-0.007	0.007
	(340.651)	(211.992)	(0.019)	(0.013)
schooling fractionalization index	-2,391.493	-1,126.774	-0.081	-0.024
	(256.286)	(172.169)	(0.014)	(0.011)
race fractionalization index	-1,113.340	-100.286	0.013	0.005
	(105.508)	(105.862)	(0.006)	(0.007)
%in poverty	-2,889.482	-301.055	-0.021	-0.062
	(272.831)	(164.387)	(0.015)	(0.010)
real median family income (in thousands)	50.059	46.562	0.002	0.001
	(2.463)	(2.143)	(0.000)	(0.000)
%25+ with hs degree	-2,152.233	-92.508	-0.031	-0.018
	(267.297)	(176.275)	(0.015)	(0.011)
%25+ with some college	1,630.735	-125.728	0.051	-0.011
	(309.263)	(238.982)	(0.017)	(0.011)
%25+ with college degree or more	2,302.358	2,485.727	-0.106	0.061
	(344.765)	(238.982)	(0.019)	(0.015)
%nonwhite	-458.831	84.124	-0.066	-0.057
	(131.883)	(111.346)	(0.007)	(0.007)
%owner-occupied housing	-1,538.289	-228.106	-0.016	0.005
	(202.771)	(131.891)	(0.011)	(0.008)
%population 65+	5,705.483	2,536.925	0.072	-0.095
	(337.716)	(250.076)	(0.019)	(0.015)
Constant	2,236.261	-278.943	0.140	0.079
	(244.492)	(177.653)	(0.014)	(0.011)
Observations	15,750	21,506	15,750	21,504
Adjusted R^2	0.8730	0.8377	0.921	0.841

Standard errors are in parentheses. Our low mobility sample consists of 3,955 districts in 1970, 1980, 1990, and 2000, who in 1970, 1980, and 1990 were in the lowest three quintiles of household migration. The non-MSA sample consists of districts located outside of MSAs (as defined in 1973). District fixed effects are included in all specifications, as are year-specific state effects. In columns (1) and (2), observations are weighted by total public K-12 enrollment. In columns (3) and (4), observations are weighted by total (public + private) enrollment.

Table 3.10: Instrumental Variables Estimates of Local Revenues per Pupil Model (Full Sample)

	Instrument:	First stage: Gini coefficient (1)	Second stage: Local revenues per pupil (2)
(a)	gini coefficient of closest district	0.110	2,980.16
		(27.500)	(2.289)
(b)	gini coefficient of closest district in another county	0.095	2,837.87
		(23.750)	(1.806)
(c)	gini coefficient of closest district in another state	0.054	1,167.61
		(13.500)	(0.427)

t -statistics are in parentheses. Coefficient estimates in column (1) are the results of three separate regressions, where our Gini coefficient of income inequality is regressed on each of our three instrumental variables, along with district fixed-effects, year-specific state effects, and all other exogenous variables included in column (4) of Table 3.4.

CHAPTER FOUR

CONCLUSION

The unifying theme of this dissertation has been the examination of several important economic trends of the past four decades and their effects on the quantity and quality of inputs available for public education in the United States. In chapter two, we showed that the first of these trends—the dramatic gender desegregation of the labor market since 1960—was accompanied by a decline in the propensity for women with the highest math and verbal skills (relative to their peers) to enter the teaching profession. While it has long been presumed that gender desegregation of the professions has adversely affected the quality of teachers, there has been surprisingly little evidence to date measuring the extent to which this is true. This essay makes a significant contribution to the existing literature by being the first to combine data from several large longitudinal surveys spanning the 43-year period 1957 to 2000 to provide empirical evidence on this subject. Our results demonstrated that the math and verbal ranking of the average new female teacher has fallen only slightly over this period, but the likelihood that a female in the top decile of her high school class enters the teaching profession has fallen dramatically (a similar decline is observed in the *bottom* decile of the skill distribution).

In chapter three we examined the relationship between two other significant trends—rising population diversity and income inequality—and local support for public education. Nationwide, household income inequality (as measured by the Gini coefficient) has increased over 16 percent since 1970; within-school district heterogeneity in race and educational attainment have risen over 90 percent and 14 percent, respectively (as measured by simple fractionalization indices). To the extent that income and demographic characteristics like race and schooling are correlated with the demand for public education, one would expect that these long-run compositional changes within U.S. school districts have had an impact on the quantity of local resources devoted to education. While some theoretical work has addressed the potential effects of income inequality on school finance, and related empirical work has studied the effects of population heterogeneity on other public programs, there has been little evidence on how these forces have affected support for public schools in practice. Bringing together four decades of demographic and financial data for 8,700 unified public school districts, we demonstrated that rising within-school district income inequality has been associated with *greater* local per-pupil spending on education. Increased heterogeneity in race appears to be associated with lower per-pupil spending, although the effects are modest.

Our results on the relationship between income inequality and spending appear to run contrary to the theoretical prediction that rising income inequality shrinks the support for public schools, as a coalition of the "ends against the middle" votes down public spending on education and wealthier families enroll in private schools. Rather, our findings are more consistent with a median voter model in which rising income inequality (particularly at the top of the income distribution) represents a decline in the tax price to the median voter, encouraging greater spending on education. We also examined the effect of rising heterogeneity and income inequality on enrollment in private schools, and found no evidence that greater income inequality increases private school enrollment. To the contrary, in our full sample of school districts we found that increased income inequality within school districts tends to *reduce* private school enrollment. It may be that the increased levels of per-pupil spending observed in districts with rising income inequality raised the attractiveness of public schooling relative to private schooling, or at least offset any positive effect on private schooling that income inequality may have had.

APPENDIX A

DETAILS ON THE LONGITUDINAL SURVEY DATASETS USED IN CHAPTER TWO

A.1 Overview

In chapter two, I combined data from five longitudinal surveys of U.S. high school students spanning four decades (1957-2000) to see how the propensity for high test-scoring women to enter the teaching profession has changed over time. In this appendix, I provide a brief description of each of the datasets used in that chapter (following), as well as detailed information about the occupational codes used to identify teachers in each dataset (Section A.2), the aptitude tests administered by each survey (Section A.3) and the sampling weights used in our analysis, where applicable (Section A.4). Section A.5 presents the chapter two regression results again, without the use of sampling weights.

Wisconsin Longitudinal Study

The Wisconsin Longitudinal Study (WLS) is perhaps the earliest major longitudinal survey of high school graduates. It is a random sample of 10,317 (virtually all non-Hispanic white) men and women who graduated from Wisconsin high schools in 1957. Follow-up surveys collected in 1964, 1975 and 1992 contain, among other things, information about each respondent's schooling, work history, and labor market experiences.[92] Also included are test scores from the Henmon-Nelson Test of Mental

[92] The National Institute on Aging is supporting a new wave of interviews during 2003-2004.

Ability (see Section A.3). A module containing very thorough information about female work histories was added in 1993. In chapter two, we focus primarily on the original 1957 survey and the 1964 follow up, which was conducted when most respondents were 25 years of age. Contrary to the other follow-up surveys, the 1964 survey was completed by the graduates' parents. More information and complete documentation for the WLS can be found on-line at http://dpls.dacc.wisc.edu/wls/index.html (access date August 9, 2003).

Project Talent

Project Talent is another early longitudinal survey of high school students, originally administered to 400,000 students in grades 9-12 in 1960. This survey was intended to be a representative sampling of American youth. Follow-up surveys were conducted 1, 5, 11, and 20 years following the expected year of graduation from high school. Designed to allow investigation into the relationship between students' cognitive abilities, family and school environment, and post-high school education, work experience, and family development, the study includes measures of numerous aptitudes, and work history. This paper makes use of the Project Talent Public Use File, a 4,000-student subsample of the original participants. The subsample includes 1,000 students from each graduating class 1960-1963, and was chosen such that every included student was a participant in the 11-year follow up in 1971-74. We focus primarily on this 11-year follow-up, when most respondents were 26-29 years of age. For more information about Project Talent, refer to Flanagan, et. al. (1981). The public use data and documentation is available through the Inter-university Consortium for Political and

Social Research (ICPSR), at http://www.icpsr.umich.edu:8080/ICPSR-STUDY/07823.xml (access date August 9, 2003).

National Longitudinal Study of the High School Class of 1972

The National Longitudinal Study of the High School Class of 1972 (NLS-72) was the first major nationwide longitudinal study of high school students conducted by the National Center for Education Statistics of the U.S. Department of Education. 16,683 high school seniors were surveyed in 1972, and then contacted again in 1973-74, 1974-75, 1976-77, 1979-80 and 1986. The study includes information on each respondent's family background, community, education, family development, and labor market experience. Also included are several measures of academic ability—SAT and ACT scores (where available), and raw scores on a battery of tests administered by the NLS to most seniors in 1972. The fifth follow up in 1986 included a supplemental survey of all of those in the original 1972 sample who had obtained teaching certificates and/or who had teaching experience. We focus primarily on the fourth NLS-72 follow-up in 1979, when most respondents were 25 years of age. More information on the NLS-72 can be found online at http://nces.ed.gov/surveys/nls72/ (access date August 9, 2003).

High School and Beyond

High School and Beyond (HSB) is another major longitudinal study of high school students conducted by the National Center for Education Statistics. It was designed to be comparable to and improve upon the NLS-72. Two cohorts of students—sophomores and seniors—were surveyed and tested in 1980, and then re-surveyed in

1982, 1984, 1986 and 1992. Like the other surveys mentioned above, HSB and its follow-ups include detailed information on each student's family background, school experiences, higher education and training, and career. More detailed academic data (such as GPA, class rank, attendance, courses taken and grades received, etc.) is available for a subset of students in the HSB Transcripts Survey. SAT and ACT scores are available for many students, as are raw scores on the HSB battery of tests, administered to both sophomores and seniors in 1980 (sophomores were re-tested in 1982—see Section A.3). We use only the *sophomore* cohort 1982 (senior year) survey, and the 1992 follow-up, when most respondents were 28 years of age. Although there is a public use version of this data available, we used the restricted use version of the data in this chapter (through a confidentiality agreement with the National Center for Education Statistics). More information on HSB can be found at http://nces.ed.gov/surveys/hsb/ (access date August 9, 2003).

National Education Longitudinal Study of 1988

The National Education Longitudinal Study of 1988 (NELS)—also designed and administered by the National Center for Education Statistics—is a nationally representative survey of eighth graders begun in the spring of 1988. A subsample of these students was re-surveyed in 1990, 1992, 1994, and 2000. As with the other NCES surveys described above, NELS and its follow-ups include detailed information about students' school, work, and home experiences, higher education and training, and career. Students were administered a battery of tests in mathematics, reading, science, and social studies during each wave of the survey that they were in high school (1988, 1990 and

1992). We focus primarily on the 1992 follow-up survey when the NELS participants were seniors in high school (we take this as our base sample—not the original sample of eighth graders), and the 2000 follow-up, when most respondents were approximately 26 years of age. We use the public use version of this dataset. More information on the NELS can be found at: http://nces.ed.gov/surveys/nels88/ (access date August 9, 2003).

Current Population Survey

The Current Population Survey (CPS) is a monthly survey of about 50,000 households, conducted by the Bureau of Labor Statistics and intended to be the primary source of information about the labor force characteristics of the United States. In this paper, we use public use microdata from the CPS March Supplement, or Annual Demographic Survey, 1964-1996. For more information, see the Bureau of Labor Statistics' CPS website at: http://www.bls.census.gov/cps/cpsmain.htm (access date August 9, 2003).

IPUMS

IPUMS is the "integrated public use micro data series," a collection of individual-level samples of the U.S. population from the decennial U.S. Census of Population and Housing. In chapter two, we make use of the 1980 and 1990 5% State IPUMS samples, a 1-in-20 sample of the U.S. population, and the 1970 1% State PUMS, a 1-in-100 sample. See the IPUMS website at http://www.ipums.umn.edu/usa/index.html (access date August 9, 2003) for more information.

A.2 Occupational Coding

One of the preliminary steps in preparing our data for analysis in chapter two was distinguishing teachers from non-teachers in each of the longitudinal datasets described above. While the method of classifying and reporting occupations varied widely from one dataset to the next, teachers (fortunately) were identified quite consistently across datasets. There were, however, some exceptions and possible miscodes. Below, we briefly discuss the occupational coding system used in each of the five longitudinal surveys, and describe how we used these variables to identify elementary and secondary school teachers.[93]

Wisconsin Longitudinal Survey

In the WLS, respondents' occupations in 1964 were recorded using a numeric code from the 1950 Census Occupational Classification system. Using the variables *ocx64* (respondent's detailed occupation code as reported in the 1964 follow-up survey) and *ocms64* (male respondent's detailed occupation as reported in 1963 or 1964 tax records), we identified 447 women and 211 men who were classified as "Teachers—NEC" (code #51).[94]

The 1950 Census Occupational Classification system was a considerably less detailed set of codes than that used in later census years. Upon inspection of other

[93] Unfortunately, we were unable to distinguish elementary teachers from secondary teachers (nor public school teachers from private school teachers) with much confidence. Therefore, in chapter two we count all K-12 teachers simply as "teachers."

[94] It is important to note that many women who reported themselves to be out of the labor force also reported an occupation—therefore, for females this occupation code should be interpreted as respondents' current or most recent job.

variables available in the WLS data, including educational attainment in 1964 (*edat64*) and reported occupation in 1975 (*ocxcuru*), we determined that the broad category "Teachers—NEC" included a number of individuals who were most likely not elementary or secondary teachers in 1964. For example, there was no unique category in the 1950 coding system for school principals or librarians. It appeared that many respondents who were school principals or librarians in 1964 may have been assigned the "teacher" occupation code in error because it was the closest available code. It also appeared that PhDs (or graduate students) who in 1964 were likely to be working as instructors or teaching assistants may have also been classified as teachers. To avoid improper inclusion of these "false positives," we excluded a number of individuals from our initial selection of teachers. Table A.1 lists the counts of excluded teachers, along with our criteria for making these exclusions. After these exclusions, our sample includes 368 female teachers and 168 male teachers.

Project Talent

Project Talent did not use a standard Census classification system in reporting occupations. Instead, the survey administrators developed their own unique 3- and 4-digit occupational coding system to identify the occupation of those individuals who responded to the follow up surveys. Only those respondents currently in the labor force were allowed to report an occupation. Our initial definition of teachers included 106 women and 69 men who were assigned the 3-digit occupation codes #400 – 433, categories listed in Table A.2 (along with the counts of individuals in each category). Note that these occupation categories exclude college or university teachers, teaching assistants, school administrators (principals, superintendents, etc), and teachers of the handicapped, librarians, speech therapists, and the like.

Seven females and one male with codes #400 – 433 had an insufficient amount of education in 1971-74 to plausibly be considered teachers (i.e. a high school education or less). These eight individuals were excluded from our identified teachers, leaving a total of 99 female and 68 male teachers in our Project Talent sample.

National Longitudinal Study of the High School Class of 1972

Participants in the 1979 NLS-72 follow-up survey were asked to "describe the job that you held the first week of October of 1979." The NCES matched the respondents' reply to a numeric code from the 1970 Census classification of occupations, and reported this occupation code as the variable *ft12ad*. We identified 468 females and 143 males whose occupation code in 1979 was #142, 143, 144 or 145 ("elementary teachers," "pre-kindergarten and kindergarten teachers," "secondary teachers," and "teachers except college and university, other"). The counts of these individuals (by sex and occupation code) are reported in Table A.3. There were 37 females and 8 males with these codes who had too little education in 1979 to be a teacher (34 and 6 females and males, respectively) or were a school principal during other waves of the survey (3 and 2). We excluded these 45 respondents from our teacher definition, leaving a total of 431 female and 135 male teachers in our sample.

High School and Beyond Sophomore Cohort

Several variables in the HSB dataset were used to determine respondents' occupation in 1992. The variables *y4303fa* ("occupation code in 1992") and *y4303ca* ("industry code in 1992") were assigned by the NCES based on the verbal or handwritten responses of participants to the survey; the original (verbatim) responses were also included in the HSB dataset as the variables *y4303d07* and *y4303e07*. Concerned about the accuracy of the occupation codes (the NCES occupational coding scheme included only 29 general categories, one of which being the broad "School Teacher," which was likely to include individuals that were not elementary or secondary teachers, such as daycare workers, adult education instructors, etc.) we hand-coded a new occupation variable by matching the verbal responses for each individual to an occupation in the 1970 Census Occupational Classification list.

In general, we used the HSB coded occupation variable (*y4303fa* equal to code 26, "School Teacher") to identify teachers. Where values were missing for this variable, we used the occupation code for an earlier year (*y4303f9* for 1991, *y4303f8* for 1990, etc).[95] We cross-checked this variable against our hand-coded occupation variable, and in most cases, they matched. Where occupations codes did not match, we relied on our hand-coded variable as the definitive occupation. In total, we identified 219 female teachers and 62 male teachers in our sample.

[95] Thus, this variable should be considered the respondents current or most recent job in 1992.

National Education Longitudinal Study of 1988

As with HSB, the NCES devised its own simple categorization of occupations for use in the NELS. However, unlike HSB, elementary and secondary teachers are more readily identified under the specific category "Educators: K-12 Teachers" (code 24). There were 302 females and 87 males who were identified as teachers in 2000 using the variable *f4bxoccd*. No respondents were excluded from this category for any other reason.[96]

A.3 Test Scores

In chapter two, respondents to the five longitudinal surveys were assigned a centile rank and standardized score based on their raw score on a test (or battery of tests) administered to their high school cohort. While the content of the exams differed somewhat across the five surveys, they were all (with the exception of WLS) tests of verbal and mathematical aptitude. In this section, we provide a brief summary of the content of the exams administered to each high school cohort, and describe how we calculated our test score centiles and standard scores.

[96] This is due in part to insufficient time available to closely scrutinize this data. We incorporated the NELS data into this chapter upon its release in Spring 2003, and have not held this data up to the same standards of accuracy as the other four datasets. Hence, appropriate caution should be used when interpreting results from this data.

Wisconsin Longitudinal Study

The WLS reports for every student a raw test score from the Henmon-Nelson Test of Mental Ability, as well as an IQ score computed from this raw score (this IQ score uses a formula that adjusts for the respondents' age when the exam was administered). National and state centile rankings for each student's raw score and IQ are also provided. The Henmon-Nelson test was administered to Wisconsin high school students during their freshman and junior years—in most cases, the WLS reports both scores.

In chapter two, we used the raw Henmon-Nelson score to compute centile ranks and standardized scores. Specifically, we use the WLS reported variable *ghnrs_bm*, the "best measure" of the respondent's raw Henmon-Nelson score, and computed a centile rank (or standard score) based on each students placement in the distribution of all high school graduates of the same sex in our sample.[97] This "best measure" variable contains the junior year score, or a score from another source (usually the freshman year score) when the junior year score is missing. In our sample of respondents to the 1964 follow-up, most of the raw scores were junior year scores, as Table A.4 shows. Summary statistics for the Henmon-Nelson raw score for our 1964 sample of WLS high school graduates are displayed in Table A.5.

Project Talent

The Project Talent test battery included tests of a wide variety of cognitive abilities including language and mathematical skills, memory, reasoning ability, and

[97] Centile ranks were based upon the distribution of all high school graduates in our dataset who took the exam, not just those who responded to the 1964 follow-up survey. The same method was used with all five cohorts.

creativity.[98] We based our centile ranks and standardized scores on a "general academic aptitude composite" test score provided by Project Talent (again, basing these centile ranks on each students' placement in the distribution of all high school graduates of the same gender who took the test in our sample). This composite test score includes scores from nine different test components: Math Information, Vocabulary (I and II), English, Reading Comprehension, Creativity, Abstract Reasoning, and Mathematics (I and II). Summary statistics for this composite score for the high school graduates in our sample are provided in Table A.6.

National Longitudinal Study of the High School Class of 1972

All NLS-72 participants were asked to complete a battery of six tests, which included Vocabulary (15 questions), Picture Number (30 questions, an associative memory test), Reading (20 questions), Letter Groups (25 questions, an abstract reasoning test), Mathematics (25 questions), and Mosaic Comparisons (116 questions, used to test the ability to detect patterns). To generate centile ranks and standard scores, we first summed four "formula scores," which NLS-72 calculated from the raw scores on these tests: the reading formula score (variable *forsc_rd*), mathematics formula score (*forsc_mt*), letter groups formula score (*forsc_lt*) and vocabulary formula score (*vocabsc*). The NCES calculated each of these four formula scores as:

(1) $FS = R-[W/(C-1)]$

where R is the number of right responses, W the number of wrong responses, and C is the

[98] See Flanagan, et. al. (1981) for a detailed description of all of the available Project Talent test scores.

number of item response alternatives. Unweighted summary statistics for this composite score for our sample of high school graduates are provided in Table A.7.

Based on these composite scores, we computed centile ranks and standardized scores for all high school graduates in our sample, based on each student's placement in the distribution of all high school graduates of the same gender in our sample. In order to appropriately place these students into the distribution of the population of high school students in their cohort, we incorporated the sampling weights provided in the NLS-72 data. These weights are described in greater detail in Section A.4.

High School and Beyond Sophomore Cohort

Sophomores participating in the HSB base year survey were asked to complete a battery of tests nearly identical to that given to the NLS-72 base year participants (about 86 percent of the test items were the same). Both the sophomore and senior cohort test battery included achievement tests in verbal ability (vocabulary and reading, a total of 41 questions) and mathematics (38 questions). In addition, sophomores were asked to complete tests on science (20 questions), writing (17 questions), and civics (16 questions). The sophomore cohort completed this test battery twice—one during their sophomore year (the base year, 1980), and one during their senior year (the first follow-up, 1982).

To calculate our HSB centile ranks and standard scores, we summed the four formula scores for reading (variable *fyreadfs*), vocabulary (*fyvocbfs*), and mathematics part one and two (*fymth1fs* and *fymth1fs*) from the test battery taken during the *senior* year (formula scores were calculated in the same manner as in the NLS-72—see above).

Unweighted summary statistics for these composite scores for our sample of high school graduates from the sophomore cohort are provided in Table A.8. We then based each student's centile rank and standardized score on the distribution of test scores among all high school graduates of the same gender in our sample. In order to appropriately place these students into the distribution of the larger population of high school students in their cohort, we incorporated the sampling weights provided in the HSB data. These weights are described in greater detail in Section A.4.

National Education Longitudinal Study of 1988

Student participants in the NELS were asked to complete a series of achievement tests during every wave of the survey that they were still in school (1988, 1990, 1992). Each cognitive test battery consisted of multiple-choice questions in four subject areas: reading comprehension (21 questions), mathematics (40 questions), science (25 questions), and social studies (30 questions), for a total of 116 questions. Students in each year were not necessarily given the same test—after the base year, NELS developed multiple tests to administer to students of varying levels of ability; they then used "item response theory" (IRT) scoring to make the students' raw scores comparable to each other (see Appendix H in Curtin, et. al (2002)). We calculated a math and verbal composite score by summing the two IRT "estimated number right" scores for reading comprehension and mathematics (source variables *f22xrirr* and *f22xmirr*). Unweighted summary statistics for this composite score for the high school graduates in our NELS cohort are provided in Table A.9.

We used this composite score to generate centile ranks and standardized scores, based on each student's placement in the distribution of all high school graduates of the same gender in our sample. In order to appropriately place these students into the larger distribution of high school graduates, we incorporated the sampling weights provided in the NELS data. These weights are described in greater detail in Section A.4.

A.4 Sampling Weights

Three of our longitudinal datasets (NLS-72, HSB, and NELS) provided sampling weights to adjust for unequal selection probabilities or nonresponse to the follow-up surveys. Because we would like our samples to reflect the population of each cohort of high school graduates as closely as possible, we make use of these sampling weights in our summary statistics and regressions wherever possible and appropriate. The following is a description of each of the original weighting variables used in our analysis.

National Longitudinal Study of the High School Class of 1972

w1: this weighting variable projects the sample of base year respondents to the population of high school seniors in 1972. We used this weight when computing centile ranks and standard scores, so that respondents' test score centiles better reflect their position in the distribution of all high school seniors. To the extent that NLS-72 over-sampled any particular group of students, this weight will adjust for this over-sampling.

w21: this weighting variable adjusts the sample of fourth follow-up (1979) participants for survey nonresponse. Because we are observing NLS-72 high school graduates in

1979, the sample may be biased toward those individuals who were willing and able to respond to this follow-up survey seven years after high school. The use of this weight in our summary statistics and regression coefficient estimates will reduce the bias that might arise from survey nonresponse.

High School and Beyond Sophomore Cohort

fu1wt: this weighting variable projects the sample of first follow-up (1982) participants to the population of 1980 high school sophomores.[99] This weight adjusts both for unequal selection probabilities and survey nonresponse (since this is the senior year follow-up, not the base year of the survey). We chose to use the weight for the first follow-up—which was administered during this cohort's senior year—for consistency with the other longitudinal surveys whose base year was the respondents' senior year. We used this weight in our calculation of centile ranks and standard scores, so that respondents' test score centiles will better reflect their position in the distribution of all high school sophomores, and not merely all participants in the first-year follow up survey.[100] To the extent that HSB over-sampled any particular group of students, or to the extent that some individuals were more likely to be included in the first follow-up survey, the use of these weights will reduce the bias associated with these factors.

---

[99] Note, the weighting variable *bywt* was used in an earlier version of chapter two—see Corcoran, Evans, and Schwab (2001).

[100] This weight is not ideal, because it makes the sample representative of all sophomores in 1980, not all high school graduates in 1982. Unfortunately, a superior weighting variable was not available. See Section A.5 for unweighted versions of our results.

fu4wt: this weighting variable in the HSB data adjusts the sample of fourth follow-up (1992) participants for survey nonresponse. Because we wish to make inferences about the population of high school students who graduated in 1982 (and were observed again in 1992), unequal response to the 1992 follow-up survey may bias our results toward those individuals who were willing and able to complete the 1992 follow-up survey. Therefore, the use of these weights in our summary statistics and regression coefficient estimates will reduce the bias associated with these factors.

National Education Longitudinal Study of 1988

f4qwt92g: unlike the NLS-72 and HSB weights which only approximately project to our population of interest, our NELS weighting variable *f4qwt92g* projects exactly to the population of 1992 high school graduates—our population of interest for the NELS cohort. It was determined that this weight would be appropriate both for computing centile ranks in 1992, and summary statistics and regression coefficient estimates for the 2000 follow-up. See Curtin, et. al. (2002) for more information on this and other weights provided in the NELS data.

A.5 Unweighted Regression Results

Tables A.10, A.11 and A.12 correspond to Tables 2.7a, 2.7b and 2.8 (the third through fifth columns, i.e. logit regressions estimating the likelihood of becoming a teacher, conditional on test score ranking for the NLS-72, HSB, and NELS cohorts). These specifications, however, exclude the sampling weights used in chapter two. Table

A.13 corresponds to Table 2.9 in the main text (predicted probabilities of becoming a teacher by decile), but again sampling weights have been omitted.

As these tables show, there is little difference in our coefficient estimates when weights are used compared to when they are not used. The sample most affected by the use of weights in Tables 2.7a and 2.7b appears to be HSB—elasticities calculated from the weighted regressions appear to be generally larger in magnitude than those calculated from the unweighted regressions, in Tables A.10 and A.11. The unweighted results only strengthen our finding that the propensity for high-ability women to teach was weaker in later years; on the other hand, without the use of weights, we observe a large fall in the elasticity in 1992 (the HSB sample) followed by a large rise in the elasticity from 1992 to 2000 (NELS). The less erratic pattern of the elasticities calculated from the weighted regressions would seem to suggest that the use of weights with these samples was probably appropriate.

The predicted probabilities in Table A.13 (the unweighted version) are virtually identical to those in Table 2.9, with the exception of NELS. For example, the predicted probability of becoming a teacher for female high school graduates in the top decile is 0.089 from the unweighted regressions, and 0.079 from the weighted regressions. The pattern is the opposite for the 9th decile. While in this case, this weakens our argument that the decline in the likelihood of becoming a teacher was largest in the top decile, the difference is fairly small (1 percentage point).

Table A.1: "False Positive" Teachers Excluded from WLS Teacher Sample

Reason:	Females	Males
Educational attainment was too low in 1964 or 1975	45	6
School principal in 1975	6	23
Librarian in 1975	6	1
Adult educator in 1975	5	0
College instructor in 1975, with 18+ years of education, and not a coach, theater, music, or art teacher.	3	5
Other individuals with 18+ years of education in 1975 (chemists, research workers, seminary instructors, etc.)	8	6
Other reasons	5	2
Total exclusions	78	43

Table A2: Teaching Occupation Codes in Project Talent

Code	Occupation	Females	Males
400	Teaching (NEC)	12	6
410	Teaching young children	1	
411	Teaching pre-school children	6	
412	Teaching elementary school	44	8
420	Teaching high school (NEC)	8	7
421	Teaching high school math	2	6
422	Teaching high school science		8
423	Teaching high school social studies	3	9
424	Teaching high school English	6	4
425	Teaching high school foreign languages		1
426	Teaching high school commercial ed	2	3
427	Teaching high school home economics	4	
428	Teaching high school trade & industrial ed		10
429	Teaching high school physical education	5	3
431	Teaching art	3	3
432	Teaching music	10	1
433	Teaching speech in high school		
	Total	106	69

Table A.3: Teaching Occupation Codes in NLS-72

Occupation Code	Females	Males
142 – Elementary teachers	102	13
143 – Pre-kindergarten and kindergarten teachers	57	3
144 – Secondary teachers	66	34
145 – Teachers except college and university, other	243	93
Total	468	143

Table A.4: Source for WLS “Best Measure” of Henmon-Nelson Raw Score

Source for *ghnrs_bm:*	Males	Females
Junior year test score	4,000	4,285
Freshman year test score	268	259
Other source	111	65
Total	4,379	4,609

Table A.5: Descriptive Statistics for WLS Henmon-Nelson Raw Score

	N	Mean	Standard deviation	Min	Median	Max
Males	4,379	56.65	11.89	10	57	90
Females	4,609	56.49	11.31	8	57	90

Table A.6: Descriptive Statistics for Project Talent General Aptitude Composite Scores

	N	Mean	Standard deviation	Min	Median	Max
Males	1,527	499.14	121.90	174	499	796
Females	1,634	503.64	112.93	66	506	779

Table A.7: Descriptive Statistics for NLS-72 Composite Scores

	N	Mean	Standard deviation	Min	Median	Max
Males	6,444	44.85	19.25	-9.3	46.6	85
Females	6,751	44.54	18.73	-5.6	46.2	85

Table A.8: Descriptive Statistics for HSB Sophomore and Senior Composite Scores

	N	Mean	Standard deviation	Min	Median	Max
Males (sophomore score)	4,245	32.26	17.75	-9.8	31.8	76.8
Males (senior score)	4,984	37.66	19.52	-6.9	38.2	78.0
Females (sophomore score)	4,735	29.47	17.00	-10.7	28.7	76.8
Females (senior score)	5,389	34.23	18.71	-9.5	34.5	78.0

Table A.9: Descriptive Statistics for NELS Composite Scores

	N	Mean	Standard deviation	Min	Median	Max
Males	3,855	84.91	22.86	28.6	86.8	128.6
Females	4,284	84.82	21.80	28.6	86.4	128.0

Table A.10: Unweighted Regression Results, Logit Regresions of Teacher Entrance Models for Females with at least a High School Degree

Dependent variable	NLS-72 Teacher in 1979	HSB Teacher in 1992	NELS Teacher in 2000
Sample mean	0.068	0.041	0.070
Sample size	6,751	5,389	4,284
	(3)	(4)	(5)
Intercept	0.460	3.942	-0.046
	(3.080)	(4.537)	(4.112)
Centile score	0.021	0.013	0.015
	(0.002)	(0.003)	(0.002)
Age	-0.166	-0.283	-0.129
	(0.118)	(0.163)	(0.157)
Marginal effect of centile score at:			
Mean centile	0.0011	0.0005	0.0009
	(0.0001)	(0.0000)	(0.0001)
85^{th} centile	0.0020	0.0007	0.0014
	(0.0003)	(0.0002)	(0.0003)
Elasticity of centile score at:			
Mean centile	0.0196	0.0124	0.0141
	(0.0018)	(0.0026)	(0.0021)
85^{th} centile	0.0185	0.0121	0.0135
	(0.0016)	(0.0025)	(0.0019)
Log-likelihood	-1530.4	-902.2	-1059.3
Pseudo r-squared	0.045	0.018	0.030

Standard errors in parentheses.

Table A.11: Unweighted Regression Results, Logistic Regressions of Teacher Entrance Models for Females with at least a High School Degree

Dependent variable	NLS-72 Teacher in 1979	HSB Teacher in 1992	NELS Teacher in 2000
Sample mean	0.068	0.041	0.070
Sample size	6,751	5,389	4,284
	(3)	(4)	(5)
Intercept	1.135	4.774	0.284
	(3.080)	(4.491)	(4.107)
Standardized score	0.622	0.354	0.462
	(0.058)	(0.073)	(0.067)
Age	-0.152	-0.290	-0.113
	(0.118)	(0.162)	(0.157)
Marginal effect of standardized score at:			
Mean score	0.0318	0.0129	0.0270
	(0.0026)	(0.0026)	(0.0037)
One standard deviation above mean	0.0556	0.0173	0.0396
	(0.0070)	(0.0044)	(0.0074)
Elasticity of standardized score at:			
Mean score	0.5886	0.3407	0.4327
	(0.0556)	(0.0706)	(0.0636)
One standard deviation above mean	0.5605	0.3359	0.4178
	(0.0503)	(0.0686)	(0.0592)
Log-likelihood	-1529.7	-902.6	-1058.0
Pseudo r-squared	0.046	0.018	0.031

Standard errors in parentheses.

Table A.12: Unweighted Regression Results, Who Enters Teaching, Females with at least a High School Degree

Dependent variable	NLS-72 Teacher in 1979	HSB Teacher in 1992	NELS Teacher in 2000
Sample mean	0.068	0.041	0.070
Sample size	6,751	5,389	4,284
	(3)	(4)	(5)
Intercept	-1.641	3.588	-2.121
	(3.255)	(4.629)	(4.251)
10th decile	2.398	1.121	1.961
	(0.407)	(0.404)	(0.482)
9th decile	2.553	1.079	2.335
	(0.404)	(0.402)	(0.478)
8th decile	2.633	0.918	1.795
	(0.403)	(0.411)	(0.489)
7th decile	2.342	1.206	2.305
	(0.406)	(0.398)	(0.478)
6th decile	2.209	0.800	2.072
	(0.409)	(0.419)	(0.481)
5th decile	2.042	0.942	1.657
	(0.412)	(0.414)	(0.496)
4th decile	1.604	0.273	1.629
	(0.427)	(0.463)	(0.496)
3rd decile	1.146	0.123	1.273
	(0.444)	(0.480)	(0.506)
2nd decile	0.699	0.204	0.585
	(0.472)	(0.479)	(0.553)
Age	-0.116	-0.272	-1.978
	(0.123)	(0.165)	(1.008)
Log-likelihood	-1513.1	-899.4	-1043.2
Pseudo r-squared	0.056	0.021	0.045

Standard errors in parentheses.

Table A.13: Predicted Probabilities of Entering Teaching as an Occupation, Females with at least a High School Degree, Based on Unweighted Regression Results

	NLS-72	HSB	NELS	NLS-72	HSB	NELS
Sample mean	0.068	0.041	0.070	0.068	0.041	0.070
Decile of test score	(3)	(4)	(5)	(3b)	(4b)	(5b)
10th	0.096	0.057	0.089	1.41	1.39	1.27
9th	0.109	0.054	0.123	1.60	1.32	1.76
8th	0.117	0.046	0.076	1.72	1.12	1.09
7th	0.089	0.062	0.120	1.31	1.51	1.71
6th	0.079	0.041	0.097	1.16	1.00	1.39
5th	0.068	0.047	0.066	1.00	1.15	0.94
4th	0.048	0.024	0.065	0.71	0.59	0.93
3rd	0.029	0.021	0.046	0.43	0.51	0.66
2nd	0.018	0.022	0.024	0.27	0.54	0.34
1st	0.001	0.017	0.013	0.02	0.42	0.19

Values in columns (3) – (5) are the average predicted probability of entering the teaching profession, by decile, for a female with at least a high school degree. These columns correspond to the unweighted regression results in Table A.12. Values in columns (3b) – (5b) are the predicted probabilities in columns (3) – (5), normalized by the sample mean for each cohort.

APPENDIX B

DETAILS OF SCHOOL DISTRICT PANEL CONSTRUCTION

B.1 Overview

In chapter three, I used a national panel of unified public school districts to study the relationship between rising within-district population heterogeneity and support for public education, as measured by the local contribution to per-pupil district revenues, and the share of K-12 students enrolled in private school. Following Hoxby (1996 and 2001), Evans, Murray, and Schwab (1997) and Harris, Evans, and Schwab (2001), we combined population and housing data from the 1970, 1980, 1990, and 2000 decennial census with school finance data from the 1972, 1982, 1992, and 2001 Census of Governments, to create a balanced panel of 8,699 unified school districts. As stated in chapter three, this panel represents over 62 percent of all school districts and over 80 percent of all publicly enrolled K-12 students in 2000-2001.

In the following sections, I describe in greater detail the contents of our panel, how this panel was compiled, how certain variables were calculated or constructed, and various problems that were encountered during its construction. Section B.2 discusses the panel's demographic, income and school enrollment data, B.3 our measures of population heterogeneity and income inequality, B.4 the financial data, B.5 our instrumental and district competition variables, and B.6 the final assembly of the full panel, and MSA subsample.[101]

---

[101] This appendix relies heavily on Amy Rehder Harris (1999), whose national public school district panel was the source for virtually all of our first three decades of data (this paper can be found at http://econ.bsos.umd.edu/econ/evans/wrkpap.htm (access date August 9, 2003); see also Harris, Evans, and Schwab (2001)). Without Amy's painstaking work on this panel dataset, this paper would not have been possible.

B.2 Demographic, Enrollment and Family Income Data

Demographic information about the population and housing in each school district comes from the *1970 Census of Population and Housing Special Fifth-Count Tallies,* the *1980 Census of Population and Housing Summary Tape File 3F,* the *1990 Census School District Special Tabulation,* and the *2000 Census of Population and Housing School District Tabulation (STP2).* Because grade-specific public and private enrollment data were excluded from the 2000 Census tabulation, we supplemented our panel with data from the National Center for Education Statistics' (NCES) *Common Core of Data: Local Education Agency Universe Survey Data 2000-2001,* the *Common Core of Data: Public Elementary/Secondary School Universe Survey Data 2000-2001,* and tract-level enrollment data from the *2000 Census of Population and Housing Summary File 3*.

Table B.1 summarizes the number of school districts found in each of the original Census files in each year (note these files do not necessarily contain the universe of school districts). For a number of reasons, not all of the school districts found in the original census files were included in the construction of this panel. First, not all school districts are unified (K-12) districts. Those districts classified as "elementary only" or "secondary only" were excluded for the reasons mentioned in chapter three.[102] Second, in 1970, school districts with fewer than 300 enrolled students were aggregated into one quasi-district in each of thirty-nine affected states (these are included under "other or

[102] In 1970, unified districts were identified where the grade span variable was equal to any of the following values: (-C, 0C, OD, OE, 1C, 1E). In other years, unified, elementary, and secondary districts were coded with a "U," "E," or "S," respectively. The 2000 census data did not include grade span information for school districts—this information was obtained from the 2000-01 *Common Core of Data: Local Education Agency Universe Survey Data 2000-2001.*

invalid," above), accounting for the loss of approximately 1,500 individual unified districts. Third, we excluded observations that covered areas that lie outside of any organized school district jurisdiction, or lacked any school district level code (many of these districts are non-operating). Finally, in all years we excluded community areas within the New York City Public School District, districts in Alaska, as well as state-managed districts located in Hawaii, the District of Columbia and Puerto Rico.[103] These criteria resulted in an upper bound of 9,508 unified districts eligible for inclusion in the panel. Additional districts that could not be matched to financial data, or were involved in a merger or consolidation, further limited the number of districts we could ultimately include in the panel to 8,699 (see section B.6 for details).[104]

Table B.2 shows the distribution of unified and all school districts across states, by year. The table also indicates—for each state—the count of unified districts found in our panel, and the percent of that state's unified and total K-12 districts that the districts in our panel comprise. As the table shows, our panel represents roughly 62 percent of all elementary, secondary and unified districts, and 81 percent of all unified districts in the United States.[105] It is clear from Table B.2 that our panel's representation varies

---

[103] The "other or invalid" row in Table B.1 includes the following: in 1970, D.C. (1), Hawaii (1), 39 aggregated areas, 107 "other" districts and 14 out of area; in 1980, D.C. (1), 234 education areas in Hawaii, 42 NYC community areas, 4 "other" and 15 out of area; in 1990, D.C. (1), 232 Hawaii areas, 36 NYC areas, and 29 out of area; in 2000, D.C. (1), Puerto Rico (1), 51 Hawaii districts, 34 NYC areas, 199 "other" districts, and 97 out of area (includes 46 unorganized rural areas, 16 are military bases, and 4 Indian reservations).

[104] In future work, I hope to compile a new 1980-2000 panel of school districts, which will allow the recovery of almost 3,000 (admittedly small) districts lost in 1970.

[105] As measured in 2000. Hawaii, Puerto Rico and District of Columbia are excluded from this calculation.

significantly across states. For example, our panel includes fewer than one third of all K-12 districts in states that make widespread use of elementary- or secondary-only districts; these states are found largely in the west (Arizona, California, and Montana, for example) and New England (Maine, New Hampshire and Vermont). Our coverage is significantly better in states where districts are mostly unified. States in which our representation of unified districts is low are those that were most likely to be affected by the 1970 small district aggregation in the census file, or had very few large districts in 1970, relative to 2000 (Arizona, Maine, North Dakota, and Nebraska, for example). Overall, we feel that in most states unified districts are quite well represented here; despite the lack of many districts in a few states, our data contains more than 90 percent of all unified districts in 16 states, and more than 75 percent of all unified districts in 31 states.

What follows is a description of all demographic, enrollment, or income variables extracted from the Census files and used in chapter three, categorized by subject:

Demographic Characteristics of the Population

Note: unless otherwise specified, these variables were reported consistently across the four decennial censuses.

hholds: the number of households in the district.

pop: the total population in the district.

black: the percent of district population that whose reported race was black. Only the 1990 and 2000 census tables report separate counts for Hispanic and non-Hispanic blacks. In those years, we include both Hispanic and non-Hispanic blacks in "black." In 2000, respondents were allowed to report more than one race. Those reporting more than one race were counted in an "other race" category.

hispanic: the percent of the district population whose reported ethnicity was Hispanic. The 1970 and 1980 census files do not categorize Hispanics in the same manner as the 1990 and 2000 census files do. Certain assumptions had to be made about the categorization of Hispanic individuals in the 1970 and 1980 census to compute this variable (see Harris (1999), pp. 6-7 for details).

white: the percent of the district population whose reported race was white. In 2000, only Hispanic and non-Hispanic whites reporting one race are counted as "white."

nonwhite: calculated as 1 – percent non-Hispanic white. Here again, certain assumptions had to be made about the classification of Hispanic individuals in the 1970 and 1980 census.

dropouts, hsgrad, somecol, colgrad: the percent of the district population 25 and older that dropped out of high school, completed a high school diploma only, completed some college, or received a college degree or higher (respectively). These variables were reported consistently across census years, except in 1990, where the educational attainment variables apply to adults aged 20 and older.

hhwkids: the percent of households in the district with children. In 1970, this variable was calculated as: (count of families with children under 18)/(count of families + count of unrelated individuals).[106] In 1980, this variable was calculated as: (count of households with 3+ children)/(total households).[107] The 1990 and 2000 calculation is the true fraction of households with children under 18.

sdp65_: the share of the total district population that is aged 65 or older.

sdp0_19: the share of the total district population that is aged 0—19.

urban: the percent of the district population living within an urban area.

pforborn: the percent of the district population that is foreign born.[108]

owner: the percent of housing units in the district that are owner-occupied.

pmobile: the percent of the district population aged five and up whose county of residence in 1965/75/85/95 differed from that in 1970/80/90/00.

---

[106] "Households" may contain families, or unrelated individuals. "Families" are defined in the 1970 census as households containing two or more related individuals. Thus, the total number of households in 1970 is simply the sum of families and unrelated individuals.

[107] It was not possible to determine how many households contain children in 1980, so it was assumed that all households with three or more persons contained children.

[108] This variable will be utilized in a future version of this chapter.

Enrollment

Note: these enrollment counts are not used to calculate per-pupil revenues (see Section B.4). They are, however, used as weights in weighted regressions and summary statistics.

enroll: the count of K-12 students in the district enrolled in public schools. For all 1970—1990 observations, this count comes directly from the census files. To obtain public enrollment counts in 2000, we used the *Common Core of Data: Public Elementary/Secondary Universe Survey File 2000-2001,* aggregating all K-12 students across schools to the district level.

enrkg_8: the count of K-8 students in the district enrolled in public schools. See *enroll* for the source of this variable.

enr9_12: the count of grade 9-12 students in the district enrolled in public schools. See *enroll* for the source of this variable.

pvtenrl: the count of K-12 students in the district enrolled in private schools. As with *enroll,* all 1970—1990 observations were taken directly from the census files. Private school enrollment counts were not, however, provided in the *2000 Census School District Tabulation.* Instead, we were able to approximate these counts by mapping census tracts to school districts and aggregating across tracts. See below for a more detailed explanation.

pvtkg_8: the count of K-8 students in the district enrolled in private schools. See *pvtenrl* for the source of this variable.

pvt9_12: the count of grade 9-12 students in the district enrolled in private schools. See *pvtenrl* for the source of this variable.

frpriv: the fraction of K-12 students in the district enrolled in private schools. For 1970—1990 observations, this variable is calculated as (*pvtenrl/*(*enroll* + *pvtenrl*)), using enrollment counts from the census files. However, as mentioned above, *pvtenrl* was not available in the 2000 census tables. Thus, for 2000 observations we calculated *frpriv* using this formula, but instead used the aggregated census tract enrollment counts for both *enroll* and *pvtenrl,* to avoid using enrollment counts from two different data sources in one variable calculation.

ttlenrl: the sum of public and private enrollment in the district (*enroll* + *pvtenrl*).

Family or Household Income

medfminc: median family income in the district in 1969/79/89/99. This variable was reported in all census years except 1970. In that year, we compute median family income using the distributional assumptions discussed in Section B.3.

rmfminc: real median family income in the district in 1969/79/89/99 (median family income expressed in 1992 dollars). This variable was computed as: *medfminc/cpi2*, where *cpi2* takes the values of 0.262, 0.517, 0.884, and 1.187 in 1969, 1979, 1989, and 1999 (respectively). See U.S. Bureau of Labor Statistics (2003).

inc_1—inc_N: counts of families whose income in 1969/79/89/99 fell into each of N income groups. As described in greater detail in Section B.3, these counts are used to estimate the parameters of an assumed income distribution in each district. N=15, 17, 25, and 16 in the 1970, 1980, 1990, and 2000 censuses, respectively. See Table B.3 for a list of the specific income categories in each census year.

inpov: the fraction of households with income in 1969/79/89/99 below the poverty line.

Computing private school enrollment counts for 2000 was one of the more challenging tasks of this dataset compilation. Because the *2000 Census of Population and Housing School District Tabulation (STP2)* failed to include a table reporting public and private school enrollments by grade, we instead used the following procedure to estimate private school enrollment for 2000.[109]

---

[109] The 2000 School District Tabulation does include one table (PCT23) that reports school enrollment for males and females age three and older, by various age categories: 3-4, 5-9, 10-14, 15-17, 18-19, etc., with no public-private distinction. Together with the total K-12 public enrollment counts taken from the *Common Core*, it would be possible to use the residual (census enrollment – CCD public enrollment) as an estimate of the number of children enrolled in private school. This would, however, require the use of two different data sources in the construction of one variable, as well as critical

First, we obtained public and private K-12 enrollment for every census tract in the United States, from the *2000 Census Summary File 3.*[110] These tracts were then mapped into school districts using census tract and unified school district cartographic boundary files provided by the Census Bureau.[111] Using ArcInfo GIS software, these boundary files were overlaid and merged, creating thousands of tract-district intersections (see Figure B.1 for an illustration using census tracts and school districts in the Lawrence, Kansas area).

While census tracts are almost always smaller than school districts, they are not necessarily contained entirely within the boundaries of one school district. In cases where census tracts crossed school district boundaries, we allocated total public and private school enrollment to school districts based on the fraction of the census tract land area residing in each district (in Figure B.1, for example, Census tract 0014 lies in at least five different school districts).

Of course, this method has some obvious problems. To provide a simple example, suppose that 10 percent of a large census tract lies in District A, while 90 percent of the same tract lies in District B. Allocating enrollment by land area would place 10 percent of the total enrollment in A and 90 percent in B. This method only works well when the population is uniformly distributed over the census tract. If, say, the 10 percent region were a densely populated urban area, and the 90 percent region were a

---

assumptions about the fraction of five-year olds in kindergarten and the fraction of 15-17 and 18-19 year olds in secondary school.

[110] Summary File 3 contains estimates of public and private school enrollment, based on responses to a 1-in-6 sample of households (the Census long-form questionnaire).

[111] These boundary files can be downloaded from http://www.census.gov/geo/www/cob/bdy_files.html (access date August 9, 2003).

rural or mountainous area, we would clearly be understating enrollment in A and overstating enrollment in B. Without the help of smaller geographic units—like census block group—this problem seems to be unavoidable.[112]

Fortunately, the most densely populated census tracts are located in urban areas, are quite small in land area, and typically are contained in one and only one district; tracts that traverse boundaries are more frequently located in large, sparsely populated rural areas. The measurement error in our private school variable, then, is likely to be highest in those districts with the smallest populations. The use of enrollment weights in most of our regressions in chapter three should thus ameliorate at least some of the effects of this measurement error, where it exists. We also benefit from the coterminous boundaries of school districts and counties in certain Mid-Atlantic and Southern states like Maryland, Virginia, and Georgia. In these cases, where school districts are typically operated by county governments, census tracts almost always lie within one and only one district.

B.3 Heterogeneity and Income Inequality Measures

Measures of population heterogeneity and income inequality within school districts are calculated using data from the 1970, 1980, 1990, and 2000 census files described in Section B.2. Our indices of race and schooling heterogeneity are simple fractionalization indices, i.e. one minus the sum of the squared shares of J different race or K different educational attainment categories (or, one minus a Herfindahl index based on these shares):

---

[112] Census data and cartographic boundary files are indeed available at the block group level. We have not yet attempted to map block groups to tracts—however, the use of tracts seems to work reasonably well for the vast majority of school districts.

(1) $$racefrac_{it} = 1 - \sum_{j=1}^{J} s_{jit}^{\;2}$$

(2) $$educfrac_{it} = 1 - \sum_{k=1}^{K} s_{kit}^{\;2}$$

where s_{jit} and s_{kit} are the shares of race group j or education category k in school district i in census year t. Both of these indices can take on values from zero to one, with values closer to one representing greater heterogeneity (a value of zero would indicate perfect homogeneity). The race and education categories we use are defined as follows:

race categories: white, black, Asian/Pacific Islander (API), and "other." We were limited to these four race categories because of the inconsistencies in race reporting across census years. In the 1970 Census, we include in the API category Japanese, Chinese, and Filipino. American Indians and individuals of any other race are counted in the "other" category. For the 1980 Census, we count Japanese, Chinese, Filipino, Korean, Asian Indian, Vietnamese, Hawaiian, Guamian, Samoan, and "Other Asian/Pacific Islanders" in the API category, and Eskimo, Aleutian, American Indian, "Race NEC Spanish," and "Race NEC not Spanish" in "other." In 1990, race was defined using the five categories white, black, API, other, and American Indian, who we combined with "other." The 2000 Census data uses the same categorization as 1990, except for the possibility that individuals may report more than one race. Those reporting two or more races in 2000 were counted as "other."

education categories: fraction of the district population 25 and older who are high school dropouts, high school graduates only, recipients of some college, or college graduates and higher. These categories were reported consistently across Census years, except in 1990, where the educational attainment variables apply to adults aged 20 and older.

There are no measures of income inequality reported by the Census at the school district level. As mentioned in Section B.2, the Census does report counts of families (or households) in various income categories in each district (inc_1 - inc_N above), as well as the median and total family (or household) income. Given aggregate income in each income *group*, one could compute an approximate Gini coefficient using simple procedures like those outlined in Gastwirth (1972), but these group aggregates are unfortunately not available. It is possible, however, to generate measures of within-district inequality by assuming a functional form for the distribution of income in each district, using the grouped income data to estimate the parameters of this distribution, and then using this distribution to compute common measures of income inequality. We use such a procedure—but it was necessary to first agree upon an appropriate income distribution.

Economists have experimented with literally dozens of distribution functions for income. The lognormal distribution, for example, is commonly thought to closely approximate the shape of the income distribution in the United States. In an article in *Econometrica,* McDonald (1984) examined a large number of one to four parameter distributions in order to assess their ability to fit the U.S. income distribution. Among these he included the popular lognormal, gamma, beta, and Pareto distributions—many of which were shown to be special cases of "generalized beta" distributions. As might be expected, the four-parameter generalized beta distribution provided a better fit than almost all other distributions, but its lack of a closed-form representation for its moments makes it quite difficult to estimate in practice. Among the three-parameter models, McDonald concluded that the Dagum (1977, 1980) distribution—a re-parameterization of the Burr Type III distribution—outperformed all other three-parameter models (and even some four-parameter models), at least for U.S. family income in 1970-1980. These results suggest that the Dagum distribution would be a good candidate for our estimation procedure, and we implement it as described below.

For a random variable *z*, the cumulative distribution function for the three-parameter Dagum distribution is as follows, for $z \geq 0$ and $(a, b, p) > 0$:

$$(3) \qquad F(z) = \left[1 + \left(\frac{b}{z}\right)^{a}\right]^{-p}$$

The r^{th} moments of this distribution are defined as:

$$(4) \qquad E[z^{r}] = pb^{r}\beta\left(1 - \left(\frac{r}{a}\right), p + \left(\frac{r}{a}\right)\right),$$

(where $\beta(*)$ is the complete beta function).

Given this distribution, $K+1$ income groups, and N_{K+1} families in each group, the probability that a family falls in group *k* is Pr(k; *a, b, p*). For the lowest income group this can be written (with y_1 as the upper limit on the first income category):

$$(5) \qquad \Pr(1; a, b, p) = \Pr(y \leq y_1) = \left[1 + \left(\frac{b}{y_1}\right)^{a}\right]^{-p},$$

and for the next $K-1$ income groups (where y_{k-1} and y_k are the lower and upper bounds of the *kth* income category ($k=2 \ldots K-1$)):

(6) $$\Pr(k;a,b,p) = \Pr(y_{k-1} \le y \le y_k) = \left[1+\left(\frac{b}{y_k}\right)^a\right]^{-p} - \left[1+\left(\frac{b}{y_{k-1}}\right)^a\right]^{-p},$$

and for the highest income group:

(7) $$\Pr(K+1;a,b,p) = \Pr(y > y_K) = 1-\left[1+\left(\frac{b}{y_K}\right)^a\right]^{-p}.$$

To estimate the parameters *a, b,* and *p*, we write the log-likelihood function as:

(8) $$\ln L(a, b, p) = \sum_{k=1}^{K+1} N_k \ln\left[\Pr(k;a,b,p)\right]$$

and estimate values for *a, b,* and *p* using maximum likelihood. According to McDonald (1984), "the estimators obtained by maximizing the multinomial likelihood function will be asymptotically efficient relative to other estimators based on grouped data; however, they will be less efficient than maximum likelihood estimators based on individual observations." With parameter estimates for *a, b,* and *p,* Dagum (1980) showed that the Gini coefficient can be calculated simply as:

(9) $$Gini = -1+\frac{\beta(p,p)}{\beta(p,p+(1/a))}.$$

We estimated the parameters *a, b,* and *p* for every school district in our panel, for 1970, 1980, 1990, and 2000. We then calculated gini coefficients for each district using (9). Various centiles of the income distribution were also estimated with these parameters, to generate several other measures of inequality in the income distribution (*log95_50, log50_5, log95_5,* and so on).

As one check of this procedure, we aggregated our counts of families in each school district income category to the state level for 1970, 1980, and 1990, estimated state-specific parameter values using our maximum likelihood procedure, and computed state-specific estimates of the Gini coefficient. We then compared our results to the actual Gini coefficients reported by the Census for each state (which they produce using individual family-level data from their one-in-six sample).[113] The results for 1970 and 1990 are plotted in Figure B.2 (panels a and b, respectively). The maximum likelihood model seems to do remarkably well at the state level—the correlation coefficient between our estimated state Ginis and the actual Ginis in 1970, 1980 and 1990 are 0.998, 0.996, and 0.980, respectively.

Because school districts are much smaller than states, we were also interested in seeing how this procedure would perform in smaller geographic areas. Using the same maximum likelihood procedure, we estimated county-specific parameters of the Dagum distribution for 1970, 1980 and 1990. While the Census does not report Gini coefficients at the county level, they do report several other aggregate measures of the income distribution. For example, in 1990 the Census reports the fraction of families in each county earning \$50,000 or more. We compared these fractions to the same fraction calculated with our estimated Dagum parameters (i.e. $1 - F(50{,}000; a,b,p)$)—again, the correlation between these values is quite high—0.996 for 1990. Analagously, we calculated average family income in each county using the moment generating function for the Dagum distribution, and compared these to the average family income reported by the Census—the correlation between the actual and predicted values in this case was 0.997. Overall, it appears our maximum likelihood procedure performs remarkably well.

[113] State Gini coefficients for 2000 had not been released as of this writing.

B.4 Financial Data

Data on per-pupil revenues were extracted from the electronic files of the *Census of Governments School System Finance (F33) Files* for 1972, 1982, 1992, and 2001.[114] These files included 15,780 observations in 1972, 16,457 in 1982, 16,236 in 1992 and 15,470 in 2001; however—like the Census demographic data—not all of these observations are operating unified K-12 districts.[115] Key variables in our panel that were extracted from these datasets are defined below:

r_fed: school district revenues obtained from the federal government (typically funds targeted for specific programs or students—bilingual or special education funds, for example).

r_state: school district revenues obtained from the district's state government (also includes funds targeted towards specific programs—special and gifted education, or school lunch and transportation programs, for example).

r_local: school district revenues raised at the local level, through property taxes, sales taxes, interest earnings, special fees and contributions, and the like.

---

[114] With the exception of the 2001 file, the Census of Government surveys were the closest to our years of demographic data. Beginning in 1992 the F-33 survey of school finances began to be administered on an annual basis. For consistency with earlier years' data, we used 2000-01 data instead of 1999-2000 data (the 2002 survey was not available as of this writing).

[115] Dependent school districts (those operated by a higher level of government) were not included in the 1972 file. For these districts, financial data was pulled from other sources (see Harris (1999) for details).

r_total: the sum of federal, state, and local district revenues. Long-term debt issued during the fiscal year is not counted in district revenues.

cogenrl: the count of K-12 students enrolled in the district in the fall of 1971, 1981, 1991, and 2000. This enrollment count will not necessarily be equal to that reported in the Census (where it was reported in all years except 2000, as discussed in Section B.3).

frevp: real federal revenues per pupil in 1972/82/92/01 (federal revenues per pupil expressed in 1992 dollars), calculated as: *(r_fed / cogenrl) / slpi*, where *slpi* is a state and local expenditure price index that takes the values of 0.279, 0.611, 1.0, and 1.323 in 1972, 1982, 1992, and 2001, respectively (see U.S. Bureau of Economic Analysis (2003)). We used the Census of Governments enrollment variable (*cogenrl*) in this calculation instead of the Census data enrollment (*enroll*) to avoid using two different data sources in the calculation of one variable.

srevp: real state revenues per pupil in 1972/82/92/01 (state revenues per pupil expressed in 1992 dollars), calculated analogously to *frevp.*

lrevp: real local revenues per pupil in 1972/82/92/01 (local revenues per pupil expressed in 1992 dollars), calculated analogously to *frevp.*

trevp: real total revenues per pupil in 1972/82/92/01 (total revenues per pupil expressed in 1992 dollars), calculated analogously to *frevp.*

sfrevp: real state + federal revenues per pupil in 1972/82/92/01 (the sum of *srevp* and *frevp*).

B.5 Instrumental and District Competition Variables

Concerned about the endogeneity of within-district income inequality changes over time, we used several different variables in chapter three to instrument for this inequality measure. These included the Gini coefficient of the closest school district (*ivdag_c1*), that of the closest district in another county (*ivdag_c2*), and finally the Gini coefficient of the closest district in another state (*ivdag_c3*).

To determine the identity of the school district closest to each district in our panel, we consulted the NCES *Common Core of Data: Local Education Area Universe Survey 2000-01* and obtained the zip code for every school district's headquarters in 2000. We then calculated the distance from each district's zip code centroid to the zip code centroid of every other U.S. school district headquarters, and captured the identity (*closest1, closest2* and *closest3*) and distance to (*milesto1, milesto2* and *milesto3*) the closest district, closest district in another county, and closest district in another state.[116] Due to

[116] Mappings from zip codes to centroid latitudes and longitudes are available from many sources. We used ZipList5, a commercially available dataset, downloadable at http://www.zipinfo.com/products/products.htm (access date August 9, 2003).

the limited availability of Gini coefficients for all U.S. districts, we restricted our search to the closest district *that was also in our panel.*

We also used a distance calculation procedure similar to that used above to generate measures of the degree of local school district choice, or, the availability of other public school alternatives within a designated region. As we did with our instrumental variables, we relied on the *Common Core of Data: Local Education Area Universe Survey 2000-01* to generate these measures, as we describe below:

dists10, dists25, dists50: the number of public school districts in 2000 within 10, 25, and 50 miles of the zip code of the district's headquarters. The higher are these measures—all else equal—the more public school district options the typical household has.

frac10, frac25, frac50: the fraction of total K-12 public enrollment within 10, 25, and 50 miles enrolled in this district. When a district is the only school district within 10, 25, or 50 miles, these variables will take a value of one. As a district becomes a less significant portion of regional enrollment (perhaps representing more competition, but possibly just representing greater concentration in one or more other districts) these variables will approach zero.

enrlfrac: a fractionalization index of school district competition based on enrollment shares (calculated for MSA districts only). One can interpret this index as the probability that a randomly selected student placed in a randomly selected district will end up in a district that is *different from her own.* This measure takes a value of zero if a region consists of only one district. With N_m students enrolled in J public school districts in metropolitan area m, and N_{jm} enrolled in district j, this index would be calculated as:

$$(10) \qquad enrlfrac_{im} = 1 - \sum_{j=1}^{J} \left(\frac{N_{jm}}{N_m} \right)^2$$

schlfrac: a fractionalization index of school district competition based on school shares (calculated for MSA districts only). This index is calculated like *enrlfrac,* except that N_m would represent the total number of schools in metropolitan area *m,* and N_{jm} the number of schools in district *j*.

diststud: the number of public school districts per student in a metropolitan area (calculated for MSA districts only).

For our regressions, we calculate each district's quintile of local school district competition based on each of these measures. In all MSA district regressions, these quintiles are based only upon those districts residing in metropolitan areas (most districts in MSAs by definition would likely be in the top quintiles of district competition if we were to base these quintiles on the entire distribution of districts).

comp1, comp2, comp3: the district's quintile in the panel-wide distribution of *dists10, dists25,* and *dists50.* (*comp1m – comp3m* are quintiles based on these distributions across MSA districts only). For example, if for one district *comp1*=1, that district is in the lowest quintile of district competition (based on *dists10*) among all districts.

comp4, comp5, comp6: the district's quintile in the panel-wide distribution of *frac10, frac25,* and *frac50.* (*comp4m – comp6m* are quintiles based on these distributions across MSA districts only). For example, if *comp5m*=2, that district is in the second quintile of district competition (based on *frac25*) among all districts residing in MSAs.

compenrl: the district's quintile in the distribution of *enrlfrac* across MSA districts only. For example, if for one district *compenrl=3*, that district is in the third quintile of district competition (based on *enrlfrac*) among all districts residing in MSAs.

compschl: the district's quintile in the distribution of *schlfrac* across MSA districts only. For example, if for one district *compschl=4*, that district is in the fourth quintile of district competition (based on *schlfrac*) among all districts residing in MSAs.

compd: the district's quintile in the distribution of *diststud* across MSA districts only. For example, if for one district *compd=2*, that district is in the second quintile of district competition (based on districts per student) among all districts residing in MSAs.

Table B.4 provides summary statistics for each of the closest district and district competition measures described above, for both the full sample and MSA subsample.

B.6 Panel Assembly and Conclusion

School districts in the demographic and finance files were matched in one of two ways—using a unique NCES ID code, or—if necessary—by hand.[117] The matching of 2000 private school enrollment data from the Census tract files, or public enrollment from the CCD was accomplished in a similar fashion. Except in cases of district mergers or splits (discussed below), the vast majority of districts were easily matched by NCES ID, at least in 1980/82, 1990/92, and 2000/01. In 1970/72, the demographic and finance files used different identifying codes for school districts. For most of these 1970/72 observations, it was possible to obtain a match by first matching each district to its 1982 counterpart (using a unique Census of Governments ID) and then capturing the 1980 NCES ID for use in 1970. Districts that were manually matched were done so using district names, grade spans, enrollment, county, or any other field that might aid in obtaining matches across datasets.

While none of the districts in our panel were involved in a consolidation before 1990, a sizeable number of districts in the panel (124) were part of a consolidation between 1990 and 2000.[118] Most of these were the result of a wave of rural school district consolidations in Minnesota (40) and Iowa (40); we consulted Minnesota House of Representatives (2003) and Iowa Department of Education (2003) for lists of specific consolidations in these states. Following Hoxby (1996) and Harris (1999), we

[117] A detailed description of the matching process and outcome for the 1970-1990 data is provided in Harris (1999). The NCES ID code is a seven-digit number, with the first two digits corresponding to the state FIPS code, and the last five digits corresponding to the NCES number for that district, in that state.

[118] In the original version of this panel dataset (which included elementary- and secondary-only districts, as well as unified), Harris (1999) identified 122 districts that were involved in mergers or splits between 1980 and 1990, as well as 39 that were part of a merger or split between 1970 and 1980. None of these districts appear in this version of the panel.

“preconsolidated” districts who took part in a school district consolidation by summing or taking weighted averages of all variables across these districts (where the use of sums or weighted averages is applied where appropriate). There was only one school district split between 1990 and 2000—the affected districts were dropped in this case.

In the end, our full sample panel contains 34,796 district/year observations, with 8,699 in each year. From this full sample, we created a subsample of 3,292 unified districts (13,168 total district/year observations) that reside in a county that was determined to be part of a metropolitan area in 1973.[119]

[119] As urban boundaries extend over time, new counties will meet Census Bureau criteria for inclusion in a metropolitan area. We wanted to ensure that all of the districts in our MSA subsample had been considered to be in a metro area for the entire sample period.

Figure B.1: Matching 2000 Census Tracts to School Districts—Example of Lawrence, Kansas Area School Districts

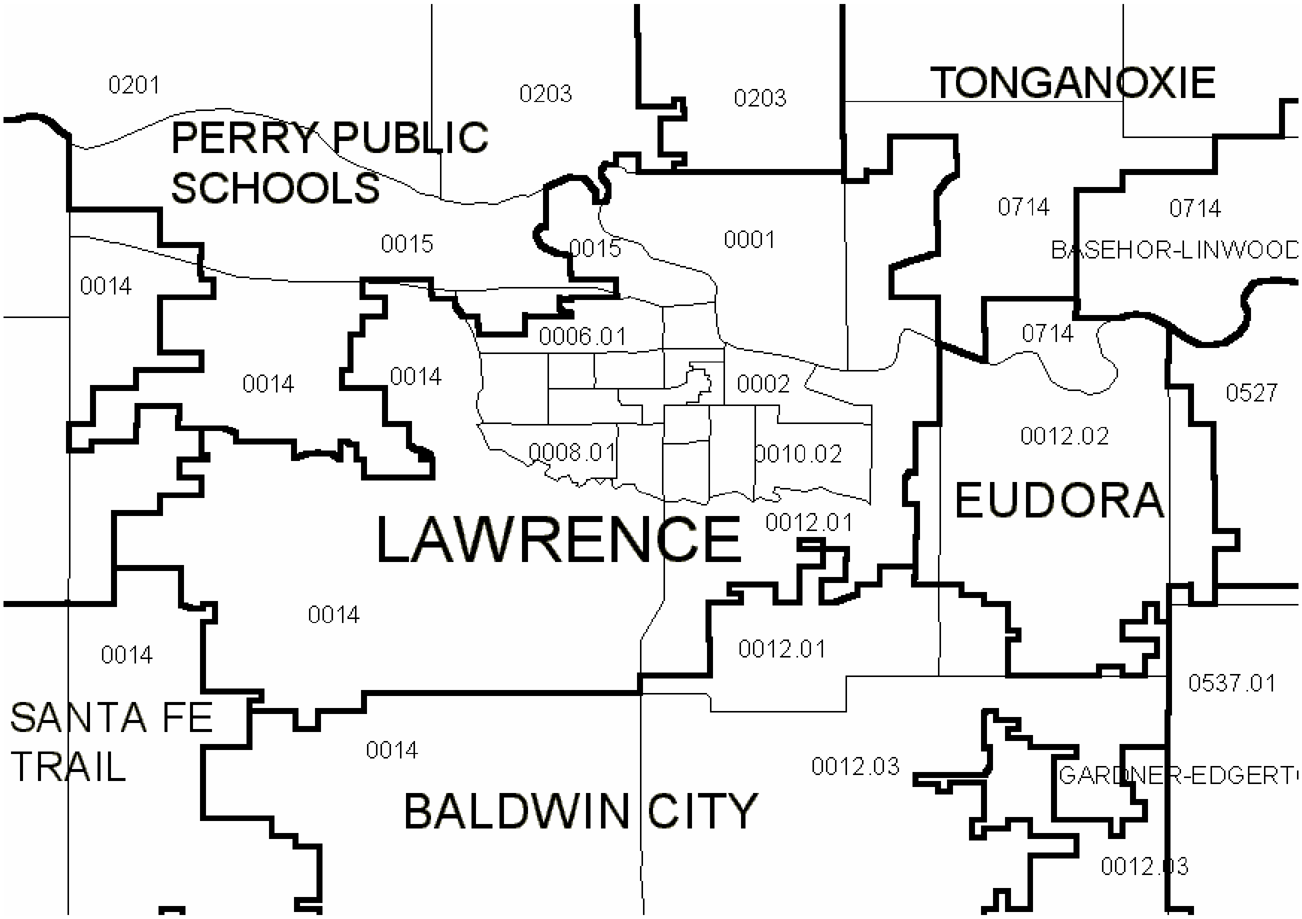


Figure B.2a: Estimated Gini Coefficient vs. Actual Census Gini Coefficient Based on State Family Income (1969)

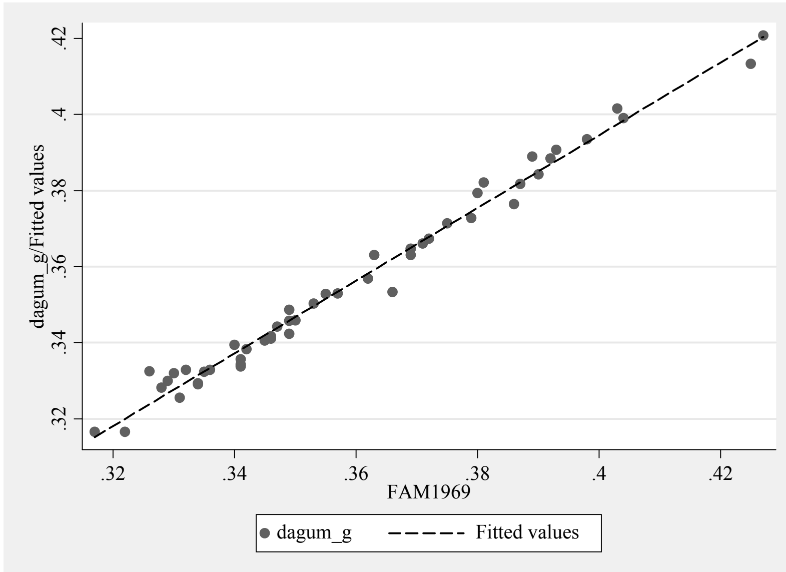


Figure B.2b: Estimated Gini Coefficient vs. Actual Census Gini Coefficient Based on State Family Income (1989)

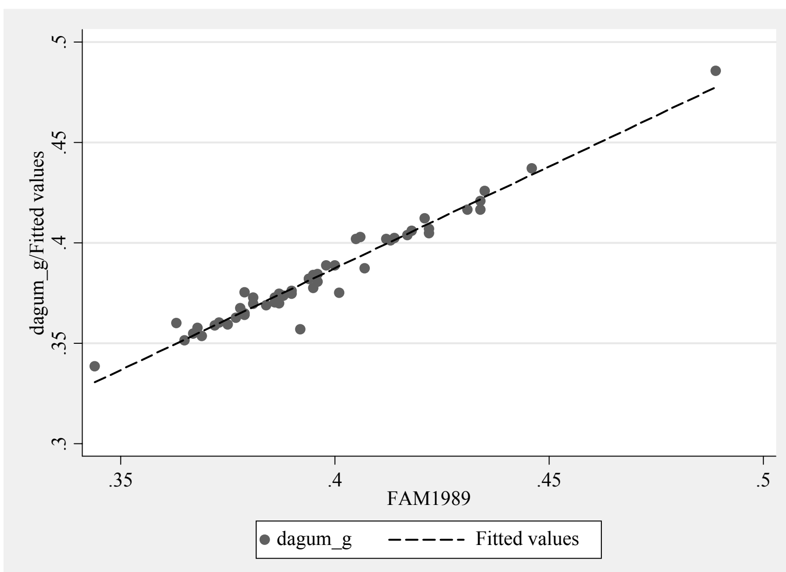


Table B.1: Number of School Districts in Original Census Files by Type and Year

	1970	1980	1990	2000
Unified	9,508	11,239	11,263	10,692
Elementary only	1,547	3,903	3,165	2,867
Secondary only	464	602	561	462
Other or invalid (includes DC, HI, NYC)	168	296	298	383
Total	11,687	16,040	15,287	14,404

Table B.2: Unified and Total Districts Across States in Census Data and Panel, by Year

	Unified Districts				Total Districts						
state	1970	1980	1990	2000	1970	1980	1990	2000	in panel, 2000	% of unified districts in panel, 2000	% of all districts in panel
AK	14	50	54	53	14	52	54	53		0.00	0.00
AL	115	125	131	128	118	126	132	128	111	0.87	0.87
AR	287	369	329	310	288	371	329	310	243	0.78	0.78
AZ	5	74	139	104	153	217	222	210	32	0.31	0.15
CA	248	278	278	349	700	1,026	821	984	207	0.59	0.21
CO	111	178	175	176	111	181	176	176	109	0.62	0.62
CT	109	110	113	166	152	165	166	166	105	0.63	0.63
DE	22	14	16	15	23	16	16	16	11	0.73	0.69
FL	67	67	67	67	67	67	67	67	67	1.00	1.00
GA	185	181	188	174	187	187	189	180	168	0.97	0.93
IA	426	443	430	374	426	443	430	374	360	0.96	0.96
ID	83	106	112	107	83	115	112	113	81	0.76	0.72
IL	430	448	427	406	937	1,012	958	890	351	0.86	0.39
IN	268	290	294	291	275	299	303	292	250	0.86	0.86
KS	281	306	304	303	281	307	304	304	271	0.89	0.89
KY	184	175	178	171	186	180	178	176	170	0.99	0.97
LA	66	66	67	66	66	66	67	66	64	0.97	0.97
MA	185	231	202	207	308	403	329	295	179	0.86	0.61
MD	24	24	24	24	24	24	24	24	24	1.00	1.00
ME	70	168	266	111	99	287	291	224	61	0.55	0.27
MI	508	529	527	524	510	575	561	554	493	0.94	0.89
MN	399	434	436	327	399	437	436	341	305	0.93	0.89
MO	394	454	452	449	402	547	541	522	378	0.84	0.72
MS	149	151	151	147	149	151	151	149	141	0.96	0.95
MT			1	55	103	553	536	447		0.00	0.00
NC	151	144	136	117	151	144	136	118	96	0.82	0.81
ND	108	265	224	170	108	330	282	226	97	0.57	0.43

NE	168	294	278	259	183	1,016	828	569	153	0.59	0.27
NH	42	84	83	64	65	169	170	159	35	0.55	0.22
NJ	188	229	232	213	472	582	576	550	155	0.73	0.28
NM	67	88	88	89	68	89	88	89	63	0.71	0.71
NV	14	16	17	17	14	17	17	17	14	0.82	0.82
NY	641	664	695	636	682	716	709	682	598	0.94	0.88
OH	610	612	612	610	611	615	612	611	596	0.98	0.98
OK	308	451	453	429	312	614	605	542	292	0.68	0.54
OR	118	157	181	176	169	312	300	195	114	0.65	0.58
PA	502	504	501	498	503	504	501	500	480	0.96	0.96
RI	28	31	31	31	38	40	37	36	27	0.87	0.75
SC	93	92	91	86	93	92	92	86	84	0.98	0.98
SD	145	191	184	172	147	196	184	176	133	0.77	0.76
TN	122	121	123	122	140	142	139	137	109	0.89	0.80
TX	728	1,069	976	974	735	1,070	1,054	1,038	699	0.72	0.67
UT	38	40	40	40	38	40	40	40	38	0.95	0.95
VA	128	139	136	135	129	139	138	135	124	0.92	0.92
VT	27	60	55	35	97	274	259	230	22	0.63	0.10
WA	203	244	290	246	212	299	296	295	195	0.79	0.66
WI	357	372	372	368	395	433	429	426	309	0.84	0.73
WV	55	55	55	55	55	55	55	55	54	0.98	0.98
WY	37	46	49	46	41	49	49	48	33	0.72	0.69
Total	9,508	11,239	11,263	10,692	11,519	15,744	14,989	14,021	8,701	0.81	0.62

Source: 1970 Census of Population and Housing Special Fifth-Count Tallies, the 1980 Census of Population and Housing Summary Tape File 3F, the 1990 Census School District Special Tabulation, and the 2000 Census of Population and Housing School District Tabulation (STP2). Excludes DC, HI, PR, and NYC education areas.

Table B.3: Income Ranges for Census Family Income Categories

1970 Census (15 categories):

under \$1000	\$8000 – 9000
\$1000 – 2000	\$9000 – 10000
\$2000 – 3000	\$10000 – 12000
\$3000 – 4000	\$12000 – 15000
\$4000 – 5000	\$15000 – 25000
\$5000 – 6000	\$25000 – 50000
\$6000 – 7000	over \$50000
\$7000 – 8000	

1980 Census (17 categories):

under \$2500	\$25000 – 27500
\$2500 – 5000	\$27500 – 30000
\$5000 – 7500	\$30000 – 35000
\$7500 – 10000	\$35000 – 40000
\$10000 – 12500	\$40000 – 50000
\$12500 – 15000	\$50000 – 75000
\$15000 – 17500	over \$75000
\$20000 – 22500	
\$22500 – 25000	

1990 Census (25 categories):

under \$5000	\$40000 – 42500
\$5000 – 10000	\$42500 – 45000
\$10000 – 12500	\$45000 – 47500
\$15000 – 17500	\$47500 – 50000
\$17500 – 20000	\$50000 – 55000
\$20000 – 22500	\$55000 – 60000
\$22500 – 25000	\$60000 – 75000
\$25000 – 27500	\$75000 – 100000
\$27500 – 30000	\$100000 – 125000
\$30000 – 32500	\$125000 – 150000
\$32500 – 35000	over \$150000
\$35000 – 37500	
\$37500 – 40000	

2000 Census (17 categories):

under \$10000	\$60000 – 75000
\$10000 – 15000	\$75000 – 100000
\$15000 – 20000	\$100000 – 125000
\$25000 – 30000	\$125000 – 150000
\$30000 – 35000	\$150000 – 200000
\$35000 – 40000	over \$200000
\$40000 – 45000	
\$45000 – 50000	
\$50000 – 60000	

Table B.4: Summary Statistics for Closest District and School Competition Variables

	Full Sample					MSA Subsample				
	Mean	Std Dev	Min	Median	Max	Mean	Std Dev	Min	Median	Max
Miles to nearest district	9.4	7.3	0.0	8.2	110.5	6.0	5.1	0.0	5.1	100.3
Miles to nearest district in another county	13.7	8.2	1.2	12.2	110.5	10.7	6.7	1.2	9.3	100.3
Miles to nearest district in another state	58.1	52.9	2.0	43.3	430.2	57.2	58.0	2.0	39.5	390.3
Number of districts within 10 miles	5.0	10.4	0	2	111	11.0	14.9	0	6	111
Number of districts within 25 miles	27.9	39.0	0	15	284	53.2	52.9	0	34	284
Number of districts within 50 miles	92.3	100.3	4	60	596	155.3	130.6	1	117	596
Fraction of enrollment within 10 miles in this district	0.52	0.38	0.01	0.44	1.0	0.28	0.30	0.00	0.15	1.0
Fraction of enrollment within 25 miles in this district	0.12	0.19	0.00	0.05	1.0	0.06	0.13	0.00	0.02	1.0
Fraction of enrollment within 50 miles in this district	0.03	0.08	0.00	0.01	1.0	0.03	0.06	0.00	0.01	1.0
Enrollment fractionalization index						0.86	0.15	0	0.91	0.99
School fractionalization index						0.87	0.13	0	0.92	0.99
Districts per student						0.0003	0.0002	0	0.0003	0.0015

REFERENCES

Aaronson, Daniel. 1999. "The Effect of School Finance Reform on Population Heterogeneity." *National Tax Journal.* 52(1), pp. 5—29.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Journal of Political Economy.* 114(4), pp. 1243—1284.

Alesina, Alberto, Reza Baqir, and Caroline Hoxby. 2000. "Political Jurisdictions in Heterogeneous Communities." NBER Working Paper #7859.

Alesina, Alberto, Edward Glaeser and Bruce Sacerdote. 2001. "Why Doesn't the United States Have a European-Style Welfare State?" *Brookings Papers on Economic Activity.* v. 2, pp. 187—254.

Bacolod, Marigee P. 2003. "Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Supply and Teacher Quality 1940-1990." University of California at Irvine Working Paper #02-03-02, revised.

Ballou, Dale and Michael Podgursky. 1997. *Teacher Pay and Teacher Quality.* Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research.

Beller, Andrea. 1992. "Occupational Segregation by Sex: Determinants and Changes." *Journal of Human Resources,* 17(3), pp. 371—392.

Benabou, Roland. 1996. "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance." *American Economic Review.* 86(3), pp. 584—609.

Bergstrom, Theodore C. and Robert P. Goodman. 1973. "Private Demands for Public Goods." *American Economic Review.* 63(3), pp. 280—296.

Betts, Julian R. and Robert W. Fairlie. 2003. "Does Immigration Induce 'Native Flight' from Public Schools into Private Schools?" *Journal of Public Economics.* 87(5), pp. 987—1012.

Blau, Francine D., Patricia Simpson, and Deborah Anderson. 1998. "Continuing Progress? Trends in Occupational Segregation Over the 1970's and 1980's." *Feminist Economics.* 4(3), pp.29—71.

Brazer, Harvey E. and Therese A. McCarty. 1987. "Interaction Between Demand for Education and for Municipal Services." *National Tax Journal.* 40(4), pp. 555—566.

Brown, Byron W. and Daniel H. Saks. 1985. "The Revealed Influence of Class, Race, and Ethnicity on Local Public-School Expenditures." *Sociology of Education.* v. 58, pp. 181—190.

Bureau of Labor Statistics. 2003. “Consumer Price Index, All Urban Consumers, U.S. All Items, 1982-84=100,” (http://www.bls.gov/cpi/home.htm, access date May 11, 2003).

Clotfelter, Charles T. 1997. “Diverging Incomes, School Desegregation and Private School Enrollment.” *Duke University Working Paper #97-22.*

Coleman, J., E. Campbell, C. Hobson, J. McPartland, A. Mood, F.D. Weinfeld and R. York. 1966. *Equality of Educational Opportunity*. Washington D.C.: Department of Health, Education and Welfare.

Corcoran, Sean P., William N. Evans and Robert M. Schwab. 2002. “Changing Labor Market Opportunities for Women and the Quality of Teachers, 1957-1992.” NBER Working Paper #9180.

Corcoran, Sean P., William N. Evans, Jennifer Godwin, Sheila E. Murray, and Robert M. Schwab. 2003. “The Changing Distribution of Education Finance: 1972 – 1997.” Russell Sage Foundation Working Paper.

Curtin, T.R., JJ. Ingels, S. Wu, and R. Heuer. 2002. *National Education Longitudinal Study of 1988: Base Year to Fourth Follow-up Data File User’s Manual (NCES 2002-323)*. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Cutler, David, Edward Glaeser and Jacob Vigdor. 1999. “The Rise and Decline of the American Ghetto.” *Journal of Political Economy.* 107(3), pp. 455—506.

Dagum, Camillo. 1977. “A New Model of Personal Income Distribution. Specification and Estimation.” *Economie Apligue.* 30(3), pp. 413—436.

Dagum, Camillo. 1980. “Inequality Measures Between Income Distributions with Applications.” *Econometrica.* 48(7), pp. 1791—1803.

Drazen, Allen. 2000. *Political Economy in Macroeconomics.* Princeton: Princeton University Press.

Eberts, Randall W. and Timothy J. Gronberg. 1981. “Jurisdictional Homogeneity and the Tiebout Hypothesis.” *Journal of Urban Economics.* 10(2), pp. 227—239.

Ehrenberg, Ronald G. and Dominic J. Brewer. 1994. “Do School and Teacher Characteristics Matter? Evidence from High School and Beyond.” *Economics of Education Review.* 13(1), pp. 1—17.

Ehrenberg, Ronald G. and Dominic J. Brewer. 1995. “Did Teachers’ Verbal Ability and Race Matter in the 1960s? *Coleman* Revisited.” *Economics of Education Review.* 14(1), pp. 1—21.

Epple, Dennis and Glenn J. Platt. 1998. "Equilibrium and Local Redistribution in an Urban Economy When Households Differ in Both Preferences and Incomes." *Journal of Urban Economics.* v. 43, pp. 23—51.

Epple, Dennis and Richard E. Romano. 1996. "Ends Against the Middle: Determining Public Service Provision when there are Private Alternatives." *Journal of Public Economics.* v. 62, pp. 297—325.

Evans, William N., Sheila E. Murray, and Robert M. Schwab. 1997. "Schoolhouses, Courthouses, and Statehouses after *Serrano*," *Journal of Policy Analysis and Management.* 16(1), pp. 10—31.

Ferguson, Ronald F. 1998. "Can Schools Narrow the Black-White Test Score Gap?" in *The Black-White Test Score Gap,* ed. Christopher Jencks and Meredith Phillips. Washington, D.C.: Brookings Institution.

Ferguson, Ronald F. and Helen F. Ladd. 1996. "How and Why Money Matters: An Analysis of Alabama Schools" in *Holding Schools Accountable: Performance-Based Reform in Education,* ed. Helen F. Ladd. Washington, D.C.: Brookings Institution.

Fischel, William A. 2001. *The Homevoter Hypothesis: How Home Values Influence Local Government Taxation, School Finance, and Land-Use Policies.* Cambridge, Massachusetts: Harvard University Press.

Flanagan, John C., David V. Tiedeman, William V. Clemans, Lauress L. Wise. 2001. *Project TALENT Public Use File, 1960-1976* [Computer File]. ICPSR version. Palo Alto, CA: American Institutes for Research [producer], 1979. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

Flyer, Frederick and Sherwin Rosen. 1997. "The New Economics of Teachers and Education." *Journal of Labor Economics.* 15(1), pp. S104—S139.

Gastwirth, Joseph L. 1972. "The Estimation of the Lorenz Curve and Gini Index." *Review of Economics and Statistics.* v. 54, pp. 306—316.

Gitomer, Drew H. and Andrew S. Latham. 1999. "The Academic Quality of Prospective Teachers: The Impact of Admissions and Licensure Data." *Educational Testing Service Research Report*.

Glomm, Gerhard, and B. Ravikumar. 1998. "Opting Out of Publicly Provided Services: A Majority Voting Result." *Journal of Political Economy.* 100(4), pp. 187—199.

Goldhaber, Dan. 2002. "The Mystery of Good Teaching: Surveying the Evidence on Student Achievement and Teachers' Characteristics." *Education Next.* 2(1), pp. 50—55.

Goldin, Claudia, and Lawrence F. Katz. 1997. "Why the United States led in Education: Lessons from Secondary School Expansion: 1910 to 1940." NBER Working Paper #6144.

Grubb, W. Norton. 1982. "The Dynamic Implications of the Tiebout Model: The Changing Composition of Boston Communities, 1960-1970." *Public Finance Quarterly.* 10(1), pp. 17—38.

Hamilton, Bruce. 1975. "Zoning and Property Taxation in a System of Local Governments." *Urban Studies.* 12(2), pp. 205—211.

Hamilton, Bruce. 1976. "Capitalization of Intrajurisdictional Differences in Local Tax Prices." *American Economic Review.* 66(5) pp. 743—753.

Hanushek, Eric A. 1970. "The Production of Education, Teacher Quality, and Efficiency." *Do Teachers Make a Difference? A Report on Recent Research on Pupil Achievement*. Washington, D.C.: U.S. Department of Health, Education, and Welfare.

Hanushek, Eric A. 1971. "Teacher Characteristics and Gains in Student Achievement." *American Economic Review.* 61(2), pp. 280-288.

Hanushek, Eric A. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature.* v. 14, pp. 1141-1177.

Hanushek, Eric A. 1996. "School Resources and Student Performance" in *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success,* ed. Gary B. Burtless. Washington, D.C.: Brookings Institution.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 1999. "Do Higher Salaries Buy Better Teachers?" NBER Working Paper #7082.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2000. "Teachers, Schools, and Academic Achievement." NBER Working Paper #6691 (revised).

Hanushek, Eric A. and Richard R. Pace. 1995. "Who Chooses to Teach (and Why)?" *Economics of Education Review.* 14(2), pp. 101-117.

Harris, Amy Rehder. 1999. "Data Chapter: The Construction of a National Public School District Panel," *mimeo,* University of Maryland, College Park.

Harris, Amy Rehder, William N. Evans, and Robert M. Schwab. 2001. "Public Education Financing in an Aging America," *Journal of Public Economics.* 81(3), pp. 449—472.

Hoxby, Caroline M. 1996. "How Teachers' Unions Affect Education Production." *Quarterly Journal of Economics.* 111(3), pp. 671—718.

Hoxby, Caroline M. 1998. "How Much Does School Spending Depend on Family Income? The Historical Origins of the Current School Finance Dilemma." *American Economic Review.* 88(2), pp. 309—314.

Hoxby, Caroline M. 2000. "Does Competition Among Public Schools Benefit Students and Taxpayers?" *American Economic Review.* 90(5), pp. 1209—1238.

Hoxby, Caroline M. 2001. "All School Finance Equalizations are not Created Equal," *Quarterly Journal of Economics.* 116(4), pp. 1189—1231.

Hoxby, Caroline M. 2003. "Our Favorite Method of Redistribution: School Spending Equality, Income Inequality, and Growth." Russell Sage Foundation Working Paper.

Ingels, Steven J. and others. *National Education Longitudinal Study of 1988: Second Follow-up: Student Component Data File User's Manual*. Washington, D.C.: U.S. Department of Education, 1994.

Iowa Department of Education, Bureau of Planning, Research and Evaluation. 2003. "District Name Changes," www.state.ia.us/educate/directory/doc/name_change.html [Access date: May 12, 2003].

Krueger, Alan. 2002. "Economic Considerations and Class Size." Princeton University, Industrial Relations Section Working Paper #447.

Lakdawalla, Darius. 2001. "The Declining Quality of Teachers." NBER Working Paper #8263.

Lankford, Hamilton, Susanna Loeb, and James Wyckoff. 2001. "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis." Stanford University Working Paper.

Lloyd, Cynthia B. and Beth T. Niemi. 1979. *The Economics of Sex Differentials.* New York: Columbia University Press.

Loeb, Susanna and Marianne E. Page. 2000. "Examining the Link Between Teacher Wages and Student Outcomes: The Importance of Alternative Labor Market Opportunities and Non-Pecuniary Variation." *Review of Economics and Statistics.* 82(3), pp. 393-408.

Lovell, Michael C. 1978. "Spending for Education: The Exercise of Public Choice." *Review of Economics and Statistics.* v. 60, pp. 487—495.

Luttmer, Erzo F. P. 2001. "Group Loyalty and the Taste for Redistribution." *Journal of Political Economy.* 109(3), pp. 500—528.

Manski, Charles F. 1985. "Academic Ability, Earnings, and the Decision to Become a Teacher: Evidence from the National Longitudinal Study of the High School Class of 1972." NBER Working Paper #1539.

McDonald, James B. 1984. "Some Generalized Functions for the Size Distribution of Income." *Econometrica,* 52(3), pp. 647—663.

Megna, Richard H. and Lee Tong Hun. 1990. "Estimation of the Demand for Local Public Education under a Kinked Budget Constraint." *Review of Economics and Statistics.* v. 72, pp. 596—602.

Meltzer, Allan H. and Scott F. Richard. 1981. "A Rational Theory of the Size of Government." *Journal of Political Economy.* 89(5), pp. 914—927.

Minnesota House of Representatives. 2003. "House Research Issues and Information: K-12 Education: School District Consolidations," http://www.house.leg.state.mn.us/hrd/issinfo/schdistcon.htm [Access date: May 12, 2003].

Murnane, R.J., J.D. Singer, J.B. Willett, J.J. Kemple and R.J. Olsen. 1991. *Who Will Teach?* Cambridge: Harvard University Press.

Murray, Sheila E. and Helen F. Ladd. 2001. "Intergenerational Conflict Reconsidered: County Demographic Structure and the Demand for Public Education." *Economics of Education Review.* 20(4), pp. 343—357.

National Center for Education Statistics, Department of Education. 2002. *Common Core of Data: Local Education Agency Universe Survey Data 2000-2001, machine-readable data file.* U.S. Department of Education, National Center for Education Statistics [producer and distributor], Washington, DC.

National Center for Education Statistics, Department of Education. 2002. *Common Core of Data: Public Elementary/Secondary School Universe Survey Data 2000-2001, machine-readable data file.* U.S. Department of Education, National Center for Education Statistics [producer and distributor], Washington, DC.

Oates, Wallace E. 1969. "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis." *Journal of Political Economy.* 77(6), pp. 957—971.

Oates, Wallace E. 1981. "On Local Finance and the Tiebout Model." *American Economic Review, Papers and Proceedings of the American Economic Association.* 71(2), pp. 93—98.

Oates, Wallace E. 1998. "Property Taxation and Local Public Spending: The Renter Effect." *Mimeo.*

Pack, Howard and Janet Rothenburg Pack. 1978. "Metropolitan Fragmentation and Local Public Expenditures." *National Tax Journal.* v. 31, pp. 349—362.

Pavalko, Ronald M. 1970. "Recruitment to Teaching: Patterns of Selection and Retention." *Sociology of Education.* 43(3), pp. 340—353.

Podgursky, Michael, Ryan Monroe and Donald Watson. 2002. "Teacher Mobility, Pay, and Academic Quality." University of Missouri—Columbia Working Paper.

Polachek, Solomon W. 1981. "Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure." *Review of Economics and Statistics.* 63(1), pp. 60—69.

Poterba, James M. 1997. "Demographic Structure and the Political Economy of Public Education." *Journal of Policy Analysis and Management*. 16(1), pp. 48—66.

Rhode, Paul W. and Koleman S. Strumpf. forthcoming. "Assessing the Importance of Tiebout Sorting: Local Heterogeneity from 1850 to 1990." *American Economic Review.*

Rice, Jennifer King. 2003. *Teacher Quality: Understanding the Effectiveness of Teacher Attributes.* Washington, D.C.: Economic Policy Institute.

Rubinfeld, Daniel. 1987. "The Economics of the Local Public Sector," in *Handbook of Public Economics, v. 2,* ed. Alan J. Auerbach and Martin Feldstein. New York: Elsevier Science Publishers.

Rubinfeld, Daniel and Perry Shapiro. 1987. "Microestimation of the Demand for Schooling: Evidence from Michigan and Massachusetts." *Regional Science and Urban Economics.* 19(3), pp. 381—398.

Santerre, Rexford E. 1989. "Representative vs. Direct Democracy: Are There any Expenditure Differences?" *Public Choice.* v. 60, pp. 145—154.

Sass, Tim R. 1991. "The Choice of Municipal Government Structure and Public Expenditures." *Public Choice.* v. 71, pp. 71—87.

Schmidt, Amy. 1992. "Private School Enrollment in Metropolitan Areas." *Public Finance Quarterly.* 20(3), pp. 298—320.

Schwab, Robert M. and Wallace E. Oates. 1991. "Community Composition and the Provision of Local Public Goods: A Normative Analysis." *Journal of Public Economics.* v. 44, pp. 217—237.

Siltanen, Janet, Jennifer Jarman and Robert M. Blackburn. 1995. *Gender Inequality in the Labour Market: Occupational Concentration and Segregation.* Geneva: International Labour Office.

Stiglitz, Joseph E. 1974. "The Demand for Education in Public and Private School Systems." *Journal of Public Economics.* v. 3, pp. 349—385.

Stoddard, Christiana. 2003. "Why Has the Number of Teachers Per Student Risen While Teacher Quality Has Declined?" *Journal of Urban Economics.* 53(3), pp. 458—481.

Temin, Peter. 2002. "Teacher Quality and the Future of America." *Eastern Economic Journal.* 28(3), pp. 285—300.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy.* v. 64, pp. 416—424.

Urquiola, Miguel. 2000. "Demand Matters: School District Concentration, Composition, and Educational Expenditure." *University of California, Berkeley, Center for Labor Economics Working Paper #14.*

U.S. Bureau of the Census, Department of Commerce. 1973. *Census of Population and Housing, 1970: Special Fifth Count Summary Tapes, machine-readable data file.* U.S. Department of Commerce, Bureau of the Census [producer], Washington, DC; U.S. Department of Education, Office of Educational Research and Improvement, National Center for Educational Statistics [distributor].

U.S. Bureau of the Census, Department of Commerce. 1982. *Census of Population and Housing, 1980: Summary Tape File 3F, school districts, machine-readable data file.* U.S. Department of Commerce, Bureau of the Census [producer and distributor], Washington, DC.

U.S. Bureau of the Census, Department of Commerce. 2001. *U.S. Census 2001 Annual Survey of Local Government Finances: School Systems, machine-readable data file.* U.S. Department of Commerce, Bureau of the Census [producer and distributor], Washington, DC.

U.S. Bureau of the Census, Department of Commerce. 2002. *Census of Population and Housing, 2000: School District Tabulation (STP2).* U.S. Department of Commerce, Bureau of the Census [producer], Washington, DC, and United States Department of Education, National Center for Education Statistics [distributor].

U.S. Bureau of the Census, Department of Commerce. 2003. *Census of Population and Housing, 2000: Summary Tape File 3, machine-readable data file.* U.S. Department of Commerce, Bureau of the Census [producer and distributor], Washington, DC.

U.S. Department of Commerce, Bureau of the Census, Census of Governments, Census of Government School System Finance (F33) File, 1972, 1982, 1992. Machine Readable Data files. United States Department of Commerce, Bureau of the Census [producer], Washington, DC, and United States Department of Education, National Center for Education Statistics [distributor].

U.S. Bureau of Economic Analysis. 2003. "Table 7.14: Chain-Type Quantity and Price Indexes for Gross Domestic Product by Sector," http://www.bea.doc.gov/bea/dn/nipaweb/index.asp. [Access date: May 11, 2003].

U.S. Department of Education. 1994. 1990 Census School District Special Tabulation, School District Data Book [CD-ROMs]. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics [producer and distributor], Washington, DC.

U.S. Department of Education, National Center for Education Statistics. 2003. *Digest of Education Statistics, 2002.* Washington, DC: ED Pubs.

Vance, Victor S. and Philip C. Schlechty. 1982. "The Distribution of Academic Ability in the Teaching Force: Policy Implications." *Phi Delta Kappan.* September, pp. 22—27.

Vegas, Emiliana, Richard J. Murnane, and John B. Willett. 2001. "From High School to Teaching: Many Steps, Who Makes It?" *Teachers College Record.* 103(3), pp. 427—449.

Vigdor, Jacob L. 2001. "Interpreting Ethnic Fragmentation Effects." *mimeo,* Duke University*,* May 3, 2001.

Wayne, Andrew J. and Peter Youngs. 2003. "Teacher Characteristics and Student Achievement Gains: A Review." *Review of Educational Research.* 73(1), pp. 89—122.

Weaver, W. Timothy. 1983. *America's Teacher Quality Problem: Alternatives for Reform.* New York: Praeger Publishers.

Wisconsin longitudinal study (WLS) [graduates and siblings]: 1957-1977. [machine-readable data file] / Hauser, Robert M. and Sewell, William H. [principal investigator(s)]. Madison, WI: University of Wisconsin-Madison, Data and Program Library Service. [distributor]

Zahs, Daniel, Steven Pedlow, Marjorie Morrissey, Patricia Marnell, and Bronwyn Nichols. 1995. *High School and Beyond Fourth Follow-up Methodology Report.* Prepared by National Opinion Research Center (Chicago, Ill) for the National Center of Education Statistics (Washington, DC).