WHAT SCHOOLS CONTRIBUTE TO EDUCATION:
A COMPARATIVE STUDY OF CATHOLIC AND PUBLIC
HIGH SCHOOL STUDENTS USING A VALUE-ADDED APPROACH

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

William P. Marquis, B.S., M.Div., S.T.L.

*****

The Ohio State University

1996

Dissertation Committee:
Lucia Dunn
Stephen Cossett
Gene Mumy

Approved by

[Signature]
Lucia Dunn
Adviser
Department of Economics
Copyright by
William Marquis
1996
ABSTRACT

Student achievement has generally been identified to be a function of socioeconomic status, teacher quality, school quality, and peer group characteristics. Empirical studies attempting to estimate educational production functions often failed to identify school specific effects of the educational process because the aggregate nature of the data disguised the school's true impact on student learning. The purpose of this study is to compare the academic achievement of public and Catholic high school students and identify which student, teacher, school, and family characteristics are significantly related to learning. Student-specific data on a sample of approximately 26,000 students, collected over an eight year period by the National Center for Education Statistics for the National Longitudinal Study Base Year 1988 is in this study. The data set includes information on individual standardized test scores, family, teacher, and school characteristics. A simultaneous equation model, consisting of a probit equation for Catholic or public high school choice, and a value-added educational production function corrected for selectivity bias was estimated for public and Catholic high school sectors. A Heckman correction technique was applied to the results of the probit equation and entered into the educational production functions to correct for selectivity bias. The dependent variable in the production functions was a standardized measure of the change in student test scores between the eighth and twelfth grades. The results of this study indicate that the student and other background
characteristics significantly related to gain scores in mathematics, reading, and social studies are more positively impacted in Catholic high schools.
Dedicated to my parents
ACKNOWLEDGMENTS

I express sincere appreciation to Dr. Lucia Dunn for her guidance and insight throughout the research. Thanks go to other members of my advisory committee, Drs. Stephen Cossett and Gene Mumy for their suggestions and comments. The Technical assistance of John Gibb is gratefully acknowledged. To the Dominican Fathers at St. Patrick Church, I offer sincere thanks to endure with me the vicissitudes of my endeavors.
VITA

September 20, 1954 ......................................................... Born - Nashua, New Hampshire

1977 ................................................................. B.S., University of New Hampshire,
Durham, New Hampshire

1977-1984 ............................................................. Seminarian, Dominican House of Studies,
Washington DC

1982 ................................................................. M.Div., Dominican House of Studies,
Washington DC

1984 ................................................................. S.T.L. Dominican House of Studies,
Washington DC

1984-1985 ............................................................ Associate Pastor, Saint Gertrude Catholic
Church, Cincinnati, Ohio

1985-1990 ............................................................ Instructo in Theology, Chaplain, Ohio
Dominican College, Columbus, Ohio

1990-1996 ............................................................ Graduate Teaching Assistant
Ohio State University

FIELDS OF STUDY

Major Field: Economics

Labor Economics: Lucia Dunn
Stephen Coslett
Gene Mumy
TABLE OF CONTENTS

ACKNOWLEDGEMENTS........................................................................................................iii

VITA......................................................................................................................................iv

LIST OF TABLES..................................................................................................................vi

LIST OF CHARTS..................................................................................................................vii

CHAPTER

I. INTRODUCTION................................................................................................................1

Background of the Study.................................................................................................1
Coleman Report..............................................................................................................4
Selectivity Bias................................................................................................................5
Efficiency in Education..................................................................................................8
Value-Added Approach..................................................................................................9
National Education Longitudinal Study, Base Year 1988.............................................11

II. LITERATURE REVIEW....................................................................................................14

Introduction......................................................................................................................14
Educational Production Functions................................................................................15
Screening Models..........................................................................................................16
Simultaneous Equations Models..................................................................................18
Value-Added Model........................................................................................................19
Coleman Report 1981....................................................................................................23
Additional Student Characteristics...............................................................................27
Summary..........................................................................................................................31

III. DISSERTATION MODEL...............................................................................................33

Introduction......................................................................................................................33
Theoretical Model.........................................................................................................34
Public High School Model Specified..........................................................................38
Catholic High School Model Specified.......................................................................39
Identification of Models..............................................................................................39
Remarks..........................................................................................................................41
IV. DATA.................................................................................................................................44

Introduction.......................................................................................................................44
NELS:88 Base Year Survey (Eighth Grade)........................................................................45
NELS:88 Second Follow-Up Survey (High School Seniors).............................................49
Composite Variables.........................................................................................................56
Cognitive Test Variables.....................................................................................................58

V. PROBIT REGRESSION RESULTS FOR PUBLIC HIGH SCHOOL CHOICE...........63

Selectivity Bias....................................................................................................................63
Student Characteristics.......................................................................................................67
Grade School & Teacher Background Characteristics.......................................................70
Eighth Grade Peer Group Characteristics..........................................................................72
Family Characteristics.......................................................................................................73

VI. PUBLIC HIGH SCHOOL STUDENT GAIN SCORE REGRESSION RESULTS........76

Selectivity Bias....................................................................................................................76
Change in Public High School Student Mathematics Test Scores.................................80
  Student Characteristics....................................................................................................80
  School Characteristics....................................................................................................89
  Peer Group Characteristics............................................................................................93
  Family Characteristics....................................................................................................95
  Mills Ratio.......................................................................................................................97
Change in Public High School Student Reading Test Scores.......................................98
  Student Characteristics..................................................................................................101
  School Characteristics..................................................................................................106
  Peer Group and Family Characteristics.........................................................................109
Change in Public High School Student Science Test Scores........................................109
  Student Characteristics.................................................................................................112
  School Characteristics..................................................................................................117
  Peer Group Characteristics............................................................................................119
  Family Characteristics..................................................................................................120
Change in Public High School Student History Test Scores.........................................121
  Student Characteristics.................................................................................................122
  School Characteristics..................................................................................................125
  Peer Group Characteristics............................................................................................127
  Family Characteristics..................................................................................................127
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Probit Regression Results for Public School Choice on Background Variables</td>
<td>65</td>
</tr>
<tr>
<td>2. Public High School Student Mathematics Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification</td>
<td>80</td>
</tr>
<tr>
<td>3. Public High School Student Reading Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification</td>
<td>102</td>
</tr>
<tr>
<td>4. Public High School Student Science Gain Score Regression Coefficients, Standard Errors, and t-ratios by specification</td>
<td>113</td>
</tr>
<tr>
<td>5. Public High School Student History Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification</td>
<td>124</td>
</tr>
<tr>
<td>6. Probit Regression Results of Catholic School Choice on Background Variables</td>
<td>135</td>
</tr>
<tr>
<td>7. Catholic High School Student Mathematics Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification</td>
<td>148</td>
</tr>
<tr>
<td>8. Catholic High School Student Reading Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification</td>
<td>160</td>
</tr>
<tr>
<td>9. Catholic High School Student Science Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification</td>
<td>167</td>
</tr>
<tr>
<td>10. Catholic High School Student History Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification</td>
<td>171</td>
</tr>
</tbody>
</table>
# LIST OF CHARTS

<table>
<thead>
<tr>
<th>CHART</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Specification of Probit Model Estimated for choice of Public High School</td>
<td>66</td>
</tr>
<tr>
<td>2. Specification of Public High Student Gain Score Regressions Estimated</td>
<td>78</td>
</tr>
<tr>
<td>4. Specification of Catholic High School Student Gain Score Regression Estimated</td>
<td>146</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Background of the Study

Student academic achievement is generally acknowledged as the primary goal of education. Much of the research conducted on the effectiveness of schools has been aimed at measuring the output of our educational institutions in terms of student test scores, high school graduation rates, college admissions, and other measures of achievement. There are several reasons for this interest in educational output. First, it has long been established that schooling enhances individual productivity and earnings.\(^1\) Since quality education can have a significant impact on an individual’s productivity and earnings, which are both public and private goods, the overall social welfare may be well served by understanding the influence our system of schools has on student productivity. Second, education contributes to an individual’s trainability, health, efficiency in consumption, access to information, social status, literacy, and a variety of other private outcomes. These primarily affect the quality of life individuals of a social economy can enjoy as a result of an effective system of schools. Third, beyond the private ends which improve the lives of children and their families, schooling also serves the nation, region, and community by providing students with a common set of values and knowledge which enable them to function democratically. Although there has been much debate over which values should be instilled in our children, there is a fundamental core of principles which is essential to any functioning democracy. It is essential that students be

---

instructed about the need to vote, the need to be informed about choices and their consequences, and the need for laws in order to maintain social stability. Finally, education contributes to cultural and scientific progress, economic growth, full employment, and the defense of the nation. Since quality education contributes to such a wide variety of private and public goods, the effects of schooling on individual achievement gains will be the central topic of this research.

During the past three decades, there has been an increasing concern about the costs and effectiveness of our system of schools on individual student achievement. The escalating costs of education over the last twenty years, accompanied by the lack of any significant improvement in student achievement, is the first reason for concern. Between 1970 and 1990, real expenditures on education in the public sector increased by 80%, from $3,000 to $5,400 per student. The recent estimate of $5,400 is an average measure of the per-pupil costs of public elementary and high school education. These costs vary from a high of $8,680 in the state of New York to a low of $2,767 in the state of Utah. Despite the increase in real U.S. expenditures on public sector education, performance on the National Assessment of Education Progress (NAEP) has been essentially flat since 1970, indicating no significant improvements in student achievement. During the early 1970s, student achievement on national SAT tests was on the decline. This decline was mainly attributed to the expanded clientele for these tests, namely a higher minority enrollment from slightly lower socioeconomic status. This outcome was often interpreted as a positive move in American education insofar as it became more accessible to a variety of students. Nevertheless, concerns over increased costs of education, with little if any general improvement in student achievement levels, have persisted to the present.

---

The fundamental problem with the vast majority of national studies regarding student achievement at the primary and secondary levels is the serious limitations of the data bases used.\(^5\) The National Assessment of Educational Progress (NAEP) was established by the federal government in the late 1960s and funded by the U.S. Department of Education, to take the academic pulse of the nation's in-school population. The NAEP’s main purpose was to chart academic progress in order to inform educators and policy makers about the current status and change over time in the competency levels of American students. Large samples of schools and students were selected randomly across the nation to assess student achievement in a wide range of curricular areas such as reading, writing, mathematics, science and social studies. The data provided by these surveys were used by researchers to compare various subgroups of American students based on race, ethnicity, gender, region, parental education, and school sector. One of the early results of these studies, designed to compare student achievement levels across various school sectors, indicated that student achievement levels were higher in the private sector than in the public sector. Some researchers used the difference in average achievement levels of students in different grades as a proxy for measures of academic gains to see if they were large enough or significant enough to be of importance to educators and policy makers. The results of these studies were often contradictory due to the methods used and the limitations in the data sets. In order to chart true academic achievement, researchers need to follow and test the same student cohort throughout their academic careers. The data set must be longitudinal rather than cross-sectional in nature. Since the NAEP surveys were cross-sectional, any conclusions about student achievement gains needed to be seriously qualified.

**Coleman Report**

One series of studies conducted in the early 1980s by James Coleman, Thomas Hoffer and Sally Kilgore for the National Center of Education Statistics (NCES) produced results, based on national cross-sectional data, that mean test scores of students at the secondary level were higher in the private school sector than in the public school sector, even after controlling for differences in student characteristics. These results received national attention and spurred researchers to investigate the possible causes for the differences in achievement outcomes for each sector. They often modeled academic achievement as the output of an educational production function. These production functions included the basic inputs to the educational process such as family background, teacher input, school facilities, peer group effects, and individual student characteristics. Researchers attempted to control for as many student, school, and family background characteristics as the data allowed in order to test for their significance on achievement outcomes. Unfortunately, the majority of these studies were based on national or regional cross-sectional student data, which did not allow for causal statements to be made about their findings. However, they were able to identify many variables which were significantly related to student achievement and thus had some explanatory value. One of their central concerns was to identify those variables or inputs to the educational process which could account for the superior achievement levels of students in the Catholic and private sectors.

Coleman claimed that students who attended private schools received higher test scores than their public school counterparts in mathematics, reading, science, and social studies. These results were attributed to the higher academic demands and the stricter discipline imposed by the private sector schools. Many researchers pointed out that these results were plagued with selectiv-

---

Selectivity bias, in this case, refers to the possibility that students attending private sector schools outperform students in the public sector because they are the children of socially established and well-educated parents. The heart of the issue concerns the difficulty in determining whether student achievement differences are due to the effectiveness of public and private sector schools or the abilities of students who attend them. Unless the influence of student ability and socioeconomic status on the choice of school is controlled, estimates of school sector effects will be biased.

**Selectivity Bias**

Many studies have identified the student's socioeconomic status, college expectations, and family income as being characteristics significantly related to student ability and achievement outcomes. It has been argued that controlling for these characteristics eliminates the statistically significant impact of private sector schools on student achievement outcomes. For example, well-educated and socially established parents might select private sector schools for their children because of the quality or type of educational services they provide. Through this process of self-selection, private and Catholic schools are often described as primarily serving students from relatively affluent families whose socially advantaged children typically score higher than the national average student. Furthermore, private sector schools are able to screen the students they admit by way of test scores and parental interviews. This screening process further enables schools in the

---

private sector to control for certain student and family characteristics often associated with ability and high academic achievement.

Private sector schools which exercise control over the selection of prospective students are considered to be directly selective. The superior achievement levels of students attending high schools in the private sector, and Catholic school students in particular, have been attributed in large part to the ability of private schools to select students on the basis of family characteristics and previous academic achievement. However, from data gathered on every Catholic high school in America, it is known that 88% of all students who apply are admitted, and very few are suspended or expelled.¹² For example, the average Catholic high school expels less than one student per year. Although Catholic schools within the private sector are selective, the major selection criteria is exercised by the families and students who attend them. If families do not want the type of education delivered by private sector schools and students do not want the order and discipline which often characterize them, they are unlikely to apply. This selection process, exercised by families and schools, might lead to a relatively homogeneous group of students with common values and shared expectations which may not be present in the nation’s schools as a whole.

On the other hand, this process of self-selection is not limited to schools in the private sector only. Public schools are indirectly selective insofar as the process of selection is exercised indirectly through the price of real estate in the districts they serve. Since family income and student academic achievement are often related, high-priced houses, located in wealthy suburbs, often house high-scoring students from well educated families of high socioeconomic status. The high

priced schools found in wealthy suburbs may be part of a public goods package which attracts upper income families who prefer the kind of educational services these schools supply.13

There is evidence that the process of self-selection into relatively homogeneous groups of student achievers of similar socioeconomic status might not entirely account for the achievement levels of students in the Catholic sector. For example, the previous argument implies that Catholic schools primarily serve students from relatively affluent families. However, in the 1992-1993 academic year, 71% of the students in Catholic high schools were from families of modest ($25,000 - $35,000) to moderate ($35,000 - $50,000) incomes compared to the national distribution of 54% in the public sector.14 There was significant under-representation in the Catholic sector schools of families with incomes exceeding $50,000 (23%) compared to those in the public sector (30%). Comparative statistics indicate that Catholic schools do not serve primarily students from affluent families, but rather those from moderate income backgrounds, and these students out-perform their counterparts in the public sector.

One must keep in mind that there are considerable differences between the types of schools students attend in the private and public sectors which might account for the differences in academic achievement. A portion of the Catholic school advantage is probably attributable to factors directly or indirectly related to the schools themselves, in particular policies on homework, curriculum, graduation requirements, guidance towards post-secondary education and the like.15 The NELS:88 base year longitudinal survey will allow researchers to control for these and many other variables so as to specify which ones are related to academic achievement. Furthermore, they will

be able to test for whether the relationships are significant enough and certain enough to be of relevance to policy makers considering school choice proposals.

**Efficiency in Education**

Differences in academic achievement between students in public and private sector schools have led to discussions concerning the efficiency of the nation's school system. According to educational theory, economic efficiency requires that higher quality inputs (and more of them) should result in an education process which promotes high levels of student achievement. For example, economic efficiency in the educational process would require that increased expenditures on special education programs result in higher levels of academic achievement for the students enrolled in them. The escalating costs of education over the last twenty years, accompanied by the lack of any significant improvement in student achievement, may indicate the inefficient use of resources in the educational process. If the nation's system of schools is inefficient in its use of educational resources, it may be possible to increase student achievement levels without increasing school expenditures on educational resources. Such an outcome would be considered "pareto efficient" insofar as students would be made better off with no additional burden to tax payers. Discussions of economic efficiency in the educational process are limited since there is little if any empirical knowledge of the production possibilities frontiers of the various types of schools. Although researchers lack sufficient data on the various types of schools required to trace out a production possibilities frontier and obtain some pareto optimal solution, they have attempted to identify certain student and school characteristics which are positively related to student achievement outcomes.

Early studies, using aggregated data, identified teacher salary and per-pupil expenditures as being significantly positively related to student achievement in public sector schools.\(^\text{16}\) This re-

sult is plausible since schools which offer higher salaries should attract larger pools of applicants from which to choose the more experienced or qualified teachers. If teachers are relatively mobile and wages reflect ability or experience, it would not be surprising to find a significant positive relationship between teacher salary and student achievement. Similarly, high per-pupil expenditures (also a function of teacher salaries) would imply superior educational facilities and programs designed to promote student achievement. However, when factors such as socioeconomic characteristics of the school communities were added to the model they were also found to be significantly positively related to student achievement. It is possible that schools with large per-pupil expenditures and high teacher salaries are also located in well established communities which harbor high achieving students. Socioeconomic status and the educational level of the parents might be more relevant to student achievement than per-pupil expenditures and teacher salaries.

Value-Added Approach

Concerns about school efficiency led other researchers to control for student and school background characteristics by using a value-added approach to student achievement. According to this approach, the change in academic achievement, measured by standardized tests administered to each student over a period of years, is the primary measure of output in the educational process. Calculated changes in achievement over time for each student were then regressed on a list of pupil-specific teacher, school, family and peer characteristics. This value-added approach differed from previous empirical investigations in two ways. First, pupil-specific data is used to measure student achievement rather than aggregated data on school and district student grade point averages. Student achievement measured by averaging student grades across an entire school district did not allow researchers to control for specific school and teacher effects on academic achievement. Second, previous studies used one achievement outcome as the primary measure of student
output rather than a change in achievement over time. The use of achievement level as an output measure, without controlling for the beginning level, made it difficult for researchers to measure the effects of the school on student learning. A value-added approach to student achievement may find that schools with high average scores contribute very little to students’ performance.

This value-added method was used to test the hypothesis that the use of per-pupil data would reveal more significant school inputs to the educational process than would the use of aggregated data.\textsuperscript{17} Data was gathered on 1,896 randomly selected students from grade schools, middle schools, and high schools in the Philadelphia School District. Change in student grades over a three-year period were regressed on a list of student, teacher, school, family, and peer characteristics first by using student-specific data, then by using school averages. The results of this study indicated that more impact from school inputs is revealed when pupil-specific data was used in the student achievement regressions. For example, it was found that teacher experience, teacher exam scores, and student class size were not significantly related to student achievement when averages were used. However, when using pupil-specific data in the regression it was found that teacher experience (and therefore salary), teacher exam scores, and class size were negatively related to student achievement gains. Furthermore, the regression results indicated no significant relationship between student achievement outcomes and the physical facilities of the school, school principal experience or extra degrees, and teacher degrees beyond the undergraduate level. These teacher, school, and staff inputs to the education process represent the largest portion of its costs and seem to have the least impact.

The results of past research using educational production functions have provided clearer insight into which school inputs are significantly related to student achievement and which are not.

Some inputs cannot be easily manipulated, such as student IQ or socioeconomic background. Other variables such as class size, teacher qualifications, and school enrollment can be changed in ways that can make the educational process more efficient. Policy makers could do this by funneling economic resources away from inputs which show no significant relationship to student achievement and towards those that do. Greater efficiency can be realized by increasing student cognitive gain without altering the total cost of inputs.

**National Education Longitudinal Study, Base Year 1988**

The results of past research have been severely limited due to the fact that much of it was based on national or regional cross-sectional data that did not adequately measure the inputs to the educational process or student achievement outcomes. In 1988, the United States Department of Education, through the National Center for Education Statistics (NCES), funded the National Education Longitudinal Study, Base Year 1988 (NELS:88). This survey collected information on a random sample of approximately 25,000 students from across the nation. Questionnaires were administered to each of the students selected for the study, their teachers, parents, and school administrators. These questionnaires were designed to provide trend data about critical transitions experienced by students as they leave elementary school and progress through high school and post-secondary institutions or the work force.¹⁸ Information on student, family, school, teacher, and peer characteristics, beginning in the eighth grade and continuing through the present, will enable researchers to control for the influence of student background and ability on academic achievement. Furthermore, researchers will be able to test for the presence of significant relationships between various school characteristics and student achievement.

Academic achievement levels of the base year student cohort (1988-1989) will be the primary focus of this study. The same group of students is surveyed and tested every two years beginning in the eighth grade and continuing through post-secondary school or the work force. Their teachers, parents, and school administrators are also surveyed at two year intervals throughout the students’ academic careers. The results of the first three waves of interviews, beginning in 1988 and ending in 1992, will serve as the data base for this study. Since various students dropped out of the original cohort over the years, the sample was freshened with new students in the second and third rounds of the survey. Only those students present in all three rounds and making normal progress will be included in this study.

Using the results of the first three waves of the NELS:88 survey, a model constructed to correct for selectivity bias will be used to test the student personal, family, teacher, peer, and school characteristics significantly related to student gain scores. This value-added approach will regress separately the change in mathematics, reading, science, and history scores between the eighth and twelfth grades on a set of school and student characteristics. This value-added model will use data collected from students in each of two types of high school, public and Catholic. The value-added regression results for each type of school will be compared in order to identify which school characteristics might account for the differences in student achievement outcomes associated with public and Catholic schools.

This value-added approach to student achievement is of particular interest to both educators and administrators in the private and public sectors. Once researchers can control for student ability and family characteristics, then the school specific contribution to the educational process can be measured. Although Catholic high schools have higher proficiency scores in math, science, reading, and writing to begin with, it is possible they might add very little to student achievement
gain over time. Yet, it is the school specific contribution to the educational process which is of in-
terest to educators and administrators since this would provide them with some measure for the
effectiveness of their school programs.
CHAPTER II
LITERATURE REVIEW

Introduction

The issue of educational choice, which may be defined as providing parents with greater choice among various school alternatives, continues to draw national attention. The heart of the issue concerns whether or not private schools are more effective than public schools in promoting student achievement at the primary and secondary levels. The academic and political debates are particularly heated when proposals are made for public subsidies to private sector schools. The outcomes of these debates will have ramifications on a variety of issues which focus on the schools themselves as well as on equal educational opportunity, future wage determination, status achievement, school financing, urban location, and choice of housing.\(^{19}\)

These debates have been and continue to be based on input-output analyses designed to identify which school inputs to the educational process might be related to desired student achievement levels. The first, and perhaps most influential, study was the *Equality of Educational Opportunity*, known as the “Coleman Report” of 1966.\(^{20}\) This study was mandated by the Civil Rights Act of 1964 and sparked the national debate on school quality. The Coleman Report was influential since it was the first time survey information was collected on the characteristics and

---

19 Hameshek, E. (1978). “Conceptual and Empirical Issues in the Estimation of Educational Production Functions,” The Journal of Human Resources, XIV, 3, pp. 351-388. Much of the following discussion on educational production functions is adapted from this article since it is an excellent summary of the research done prior to 1978.

achievement levels of over a half million students from 3,000 schools across the nation. This information allowed researchers in the social sciences the opportunity to focus attention on the relationship between school inputs and student achievement.

**Educational Production Functions**

As economists entered this field of study, the usual input-output analysis was replaced by educational production function models, relating educational inputs to student output or achievement. Most of the early empirical research following the Coleman Report estimated a reduced form equation similar to equation (2.1).²¹ According to this equation, student achievement (A)

(2.1) \[ A = F(GSES, TQ, SQ, PG) \]

is a function of genetic endowment and socioeconomic status (GSES), teacher quality (TQ), non-teacher school quality (SQ), and peer group effects (PG). Most frequently student achievement outcomes were measured by using the results of various standardized test scores, although attitudes, college continuation, dropout rates, and attendance rates have also been analyzed.²²

Much of the public's interest in academic achievement, and the role of school inputs, relates to the perceived effect they have on future capabilities and the success of students after they have left school. Early studies concerning the effects of student achievement on subsequent achievement in the labor force were inconclusive.²³ There was, however, a positive and significant relationship between test scores and the quantity of education that students obtained. Economists continued to study the relationships between the quantity of education and earnings as well as labor market performance. Researchers found, in many cases, a significant and positive relationship be-

---


tween them. 24 Yet it was also found that intellectual skills, measured primarily by test score results, were not the most important outcome of schooling in determining future individual success. In certain types of jobs where cognitive skills are important, more education would be desirable, but in other types work requiring manual dexterity, less education might actually be desirable. These early studies led economists and social scientists to theorize that test scores were a signaling device which helped to select students for further education rather than an outcome of quality schooling.

**Screening Models**

Concomitant to the production function model of educational achievement, a new model was developed which considered education as a sorting mechanism providing information about student endowments. 25 According to sorting models, the educational system provides individual students with information about their own capabilities by reporting their performance on examinations. Students then select the educational track most suited to their ability level in order to capture their future "ability rents", or economic returns to their investments in education. These models typically assume that schools provide individuals with information about their natural abilities as opposed to actually improving their skills. School quality or efficiency is measured by how well the institution sorts students into the areas of their comparative advantage rather than measuring the improvement in academic skills. Finally, schools are seen as providing education as primarily a private consumption good which improves student ability to capture future ability rents.

Further specifications of the screening model included the choice of private school education as an alternative to the public school option. These models theorized that individual families self-selected into school districts which provided a level of educational quality determined by the

---

average voter. Educational quality was often assumed to be interchangeable with the quantity of educational inputs provided. School districts with greater tax revenues, due to higher district property values, were believed to provide better quality public school education than was provided by lower income school districts. Wealthy suburbs, for example, would be seen as having the financial resources necessary to provide more educational inputs taking the form of low student-teacher ratios, larger libraries, more computers, and more programs for students with special needs.

Individuals with incomes greater than the median income, or average family income for the local public school district, often demand more education than what is provided publicly and opt into the private school sector. The underlying assumption in these models is that private schools provide a greater quantity of education than the public sector schools and therefore a greater quality education. Although empirical evidence indicates that private school expenditures are significantly lower than those of the public sector, students in private sector schools tend to score higher on standardized tests. This points to the basic weakness of this model, which lies in the assumption that quantity of educational inputs is significantly positively related to the quality of educational outcomes.

In addition to the educational production function and screening models, a third model using simultaneous equations was developed to measure multiple outputs in the educational process which jointly determine one another. Many of the earlier studies, concerned with the determinants of student achievement, used single-equation production functions which related educational inputs such as individual, home, teacher, and school characteristics to scores on achievement tests. Simultaneous equation models explicitly recognize that the educational process has multiple outputs.

26 Guerra, Michael J., "Dollars and Sense" p. 23.
Simultaneous Equations Models

One study in particular used a model consisting of six simultaneous equations.28 Six endogenous variables—including student’s achievement, motivation, expectations, self esteem, parents and teacher expectations—were regressed separately on a set of forty-eight exogenous variables and five of the six endogenous variables. Among the exogenous variables were measures of (i) student demographics such as sex, ethnic group, and age; (ii) family background characteristics such as socioeconomic status and information on the home; (iii) school peer group characteristics such as racial composition of the school and classroom; and (iv) school characteristics such as problems in the school, and regional location. Using a sample of over sixteen thousand twelfth grade students from the Equality of Educational Opportunity survey, the system of equations was estimated using Two-Stage Least Squares regression analysis.

The results of the study indicated that many educational outputs jointly determine one another. When the variable representing student achievement was regressed on student motivation and other exogenous variables, a significant positive relationship was found between achievement and motivation. The results of the regression equation using self-esteem as the dependent variable indicated significant positive relationships between it and student achievement as well as it and motivation. Similarly, when student motivation was regressed on student achievement, a significant positive relationship was found. The results of the simultaneous equation model indicate a complex interrelationship between student motivation, self-esteem, and achievement. All three variables are inputs as well as outputs of the educational process. The significant relationships between the coefficients of these variables suggest that high achieving students are more likely to have high self-esteem which, in turn, contributes to achievement.

28 Boardman et al., The Educational Process.
Typical production functions in the fields of Industrial Organization or Labor Economics incorporate two or three inputs to the production process such as capital, labor, and possibly education (human capital). In the Economics of Education literature, the relatively fixed inputs of labor and capital in the educational process (one teacher per classroom with small variance in class size) could explain little about student achievement. Most of the empirical literature in this field, since the Coleman Report of 1966, suggests that achievement is a function of a student’s genetic endowment, socioeconomic status, teacher quality, school quality, and peer group characteristics.\textsuperscript{29} Since this equation is a reduced form, attitudinal differences such as self-esteem and motivation are themselves determined by the four factors listed above. Conceptually, there has been little controversy over the specification of the model. Controversy enters when more details about the definition and measurement of variables are introduced.

**Value-Added Model**

Early attempts at estimating equation (2.1) have represented inputs by school, district, or nation-wide averages rather than by the more student-specific data.\textsuperscript{30} Typically, data gathered for the Coleman Report (1966) on individual students in the first, third, sixth, ninth, and twelfth grades were merged with information on their peer groups, a “representative teacher” (usually the school average), school quality (as reported by the principal), and socioeconomic status of the school (often using block house values of the school district as a proxy). Most analyses were purely cross-sectional and included only contemporaneous measures of the inputs. Many educational inputs, such as learning capacity and parental involvement, were not measured directly but were proxied by other observable attributes such as socioeconomic status estimated by the value of the student’s


home. Measurements of school and teacher characteristics were based on data collected for administrative purposes and provided average measures of school or teacher characteristics independent of the individual student. The fact that schools are heterogeneous institutions with considerable variance in teacher experience and education introduces considerable errors in measurement when testing for significant relationships between teacher attributes and individual student achievement. Measurement error resulted in biased coefficients roughly proportional to the variance of the measurement error. 31 Educational production functions assumed that the included variables were accurately measured, when in fact they often were not.

Student achievement levels were often measured by test score results which were aggregated at each grade level and averaged out for each school. Achievement gain was imputed by comparing the difference in mean test score results for each grade for all the schools in the survey. As a result, early empirical investigations failed to find significant relationships between school characteristics and student achievement because the aggregated data disguised the school’s true impact. In most of the early studies little attention was paid to the underlying dynamic structure of the educational process and how the effects of different educational inputs accumulate. For example, comparing the mean differences in test scores between two cohorts of students does not allow for the exploration of how various teacher or school characteristics are related to individual student academic achievement over a period of time. Researchers’ attempts to identify which student, family, teacher, school, and peer group characteristics were significantly related to student achievement led to the development of a new approach of estimating educational production functions known as the value-added method.

According to this method, the change in academic achievement, measured by standardized tests administered to each student over a period of years, is the primary measure of output in the educational production function. This approach became known as the value-added method because it focuses on the improvement in academic skills which students accumulate during the educational process (the value added to the student) and how this relates to various student, family, and school characteristics. Calculated changes in achievement over time for each student are then regressed on a list of pupil-specific teacher, school, family and peer characteristics. This value-added approach differed from previous empirical investigations in two ways. First, pupil-specific data is used to measure student achievement rather than aggregated data on school and district student grade point averages. Student achievement measured by averaging student grades across an entire school district did not allow researchers to control for specific school and teacher effects on academic achievement. Second, previous studies used one achievement outcome as the primary measure of student output rather than a change in achievement over time. The use of achievement level as an output measure, without controlling for the beginning level, made it difficult for researchers to measure the effects of the school on student learning. A value-added approach to student achievement may find that schools with high average scores contribute very little to students’ performance.

This value-added method was used to test the hypothesis that the use of per-pupil data would reveal more significant school inputs to the educational process than would the use of aggregated data. One such study conducted by Anita Summers and Barbara Wolfe used panel data from over 1,500 randomly selected students from 150 primary and secondary schools in the Philadelphia School District.32 The panel data included a three year personal educational history for each student, information on the school-wide resources of the school attended, the estimated stu-

---

dent's family income, and information on each student's teachers. The dependent variable chosen was the calculated change in student composite achievement scores over a three year period from third to sixth grade. The independent variables were of four types: the genetic and socioeconomic character of the student (family income, IQ, race), student-specific school and teacher characteristics (class size, teacher quality), and peer group characteristics (% of high achievers, % Black). The calculated change in student composite scores was regressed on the independent variables-first using pupil-specific data, then using school averages. The two different regression results indicated that more impact from school and teacher characteristics is revealed when pupil-specific data is used.

The most interesting results related to the findings on teacher characteristics. When school averages were used, the rating of the teachers' undergraduate college, their teacher exam score, and their years of teaching experience were not significantly related to student achievement gain. When student-specific data was used, the rating of the teachers' undergraduate college was significantly positively related to student achievement gains while teacher experience and teacher exam scores had a small but significant negative relation to student achievement gain. Equally interesting were the results that third grade test scores were positively related to student achievement gain, whereas class size, school enrollment, library books per student, and percent of low achievers were significantly negatively related to student achievement gain.

The value-added method of estimating the relationship between student academic gains and various teacher and school characteristics might provide insight into which characteristics significantly impact the educational process. Certain student and family characteristics such as student IQ or socioeconomic status cannot be easily manipulated. However, other variables representing

---

33 This measure of family income was proxied by using Census data to estimate the value of the house or apartment the family owned or rented.
class size, teacher qualifications, and school enrollment can be changed in ways that make the educational process more efficient. Economic efficiency in the educational process is improved when significant school and teacher characteristics are altered in such a way as to improve student academic outcomes without increasing educational costs. According to the regression results of this study, the coefficients of the variables representing the physical facilities of the school, playground space, a new or old building, school principal experience or extra degrees, and teacher degrees beyond the undergraduate level were not related to the improvement of student skills, or learning. On the other hand, the variables representing beginning teachers, small classes, and small schools were each significantly related to student learning. The value-added method of estimating the relationships between student learning and various teacher and school attributes might provide insight to educators, economists, and sociologists attempting to identify the optimal package of school and teacher inputs to the educational process required for economic efficiency and high student achievement.\textsuperscript{35}

\textbf{Coleman Report 1981}

Concurrent to this development in the Economics of Education literature, the issue of public support for private education was receiving national attention.\textsuperscript{36} At that time (late 1970s and early 1980s) there were two broad policy directions. The first concerned the facilitation of private education, exemplified by proposals for school vouchers at the state level. The second concerned the restriction of private schooling, exemplified by an IRS proposal to remove tax exemption

\textsuperscript{34} Composite measures were used because the raw scores on different tests cannot be directly compared.  
privileges from private schools which did not meet a racial-balance criterion. It was in the context of addressing this issue of policy direction that the Federal Government funded a series of national studies, conducted in the early 1980s by James Coleman, Thomas Hoffer, and Sally Kilgore for the National Center of Educational Statistics (NCES), on student achievement in the nation’s public and private sector schools.

The first study used data collected on students from a random sample of approximately 1000 public and private high schools. The study consisted of three waves of interviews, testing and data collection. In 1980, 36 sophomores and 36 seniors were randomly selected in each school, and were extensively interviewed. Their records were analyzed, and they were given six standardized achievement tests. Considerable data were also collected on the district and school. Two years later, a second study was conducted in which the 1980 sophomores were interviewed and re-tested, including those sophomores who had dropped out of school. The 1980 seniors were also tracked down but only re-interviewed to find out what they were doing. These two studies provided researchers with a data base which included student information on test scores, family income, personal background characteristics, and type of high school attended. In 1984, a third study was conducted on a subsample of 530 of the original schools. Information was gathered on each school using in-depth interviews with teachers and principals concerning school organization, governance, and attitudes towards students and teaching. This study, entitled *The Administrator and Teachers Survey* (ATS), provided researchers with information concerning school characteristics from a nationally based sample which included variables other than the standard measures of resources, size, and the like. Some of these variables were used by researchers and educators to study the relation-

---

ship between student academic outcome and various school characteristics regarding school policy, personnel practices, leadership, teacher and principal expectations.

The first wave of research was conducted on the results of a cross-section of sophomore and senior test scores in reading, vocabulary, and mathematics gathered in the 1980 study. The most important results of this wave of research was to point out the different characteristics between the private and public sector schools. The private schools were mostly parochial, suburban, and smaller; they enrolled fewer Blacks and Hispanics, and the students came from more educated parents with median incomes 24% greater than families in the public schools. The school environments were very different as well. Private schools offered, and students took, more advanced academic courses, experienced higher expectations from teachers and administrators, had more homework, experienced more strict and fair discipline, experienced less fighting, less violence, less drug or alcohol abuse, more school spirit, and more involvement in school activities. Students in the private sector schools scored higher on average than did the public school students at the sophomore and senior levels. This report added to the Economics of Education literature in so far as it highlighted a difference between student achievement outcomes in the public and private sectors, namely that private sector students, on average, scored higher on standardized reading, vocabulary, and mathematics tests than their public school counterparts.

Most critics agree that there is strong evidence in the study indicating that private sector students outperformed students in the public sector, but the reasons given for this difference have sparked a national debate which has persisted to the present time. According to the Coleman Report, private school characteristics relating to higher academic demands and stricter discipline than were imposed on students in the public sector high schools, were responsible for the higher student

---

test scores. This conclusion was based on the results of two models which regressed separately student test scores on seventeen family background variables, and several school policy variables. The first model used data on students attending private high schools while the second model used data on students attending public high schools. Critics of the models have argued correctly that unless all relevant factors are included in a regression equation, the coefficients estimating the impact of the factors will be biased. In this case, if any significant background variable related to both Catholic school choice and student achievement is not included, the significance of the omitted background variable could be attributed to private school characteristics and its policies. The main issue concerns whether private school policies or the presence of certain shared private school student background characteristics, possibly omitted from the regression equation, account for the superior academic achievement of the students who choose to attend them. This is referred to in the Economics of Education literature selectivity bias.

Early researchers focused attention on the possible existence of selectivity bias. They tested for the significance of variables not included in the Coleman models, variables which might account for both the choice of private sector high schools and superior academic achievement on standardized tests. Goldberger and Cain were among the first to identify the selectivity problem and noted that much of the material used to test the sophomores and seniors in the Coleman study contained material which was primarily taken from the elementary school level, and not from an explicit part of the high school curriculum. They hypothesized that the tests might be revealing what the students learned at the elementary, rather than at the high school, level. They also suggested the possibility that high achievers, or students of greater ability, were choosing to attend private sector high schools. In order to test for the true significance of private school characteris-

---

tics as they relate to achievement outcomes, it is necessary to control for prior academic achievement and student ability.

**Additional Student Characteristics**

Noell added to the literature by identifying four measured student background characteristics, available but not incorporated into Coleman’s model, which could account for higher private school student test scores. The addition of the four variables-sex, handicap status, region, and eighth grade college expectation-reduced the relationship between private school attendance and student test scores outcomes to insignificance. The one exception was the sophomore reading test scores, which remained significantly related to private high school attendance. The results of the model indicated that the inclusion of four added variables (sex, handicap status, region of residence, and eighth grade college expectations) eliminated the statistically significant relationship between Catholic school attendance and student achievement outcomes for all test scores except the sophomore reading scores. Here there was still a small but significant positive relationship.

It is worth emphasizing that the first wave of studies used cross-sectional data from the Coleman survey (1980), which did not include measures for prior student achievement. Without a direct measure of prior achievement, it is difficult to infer anything about student achievement gain or school sector effects, even when controlling for student and family characteristics. A great deal of research using data from Coleman’s follow-up survey in 1982, which re-tested the 1980 sophomores including those who dropped out, indicated that the latter test scores were strongly related to the prior ones. Research also indicated that after controlling for prior achievement, other variables such as student and family characteristics, course taking and tracking remained significant. Since researchers had access to student panel data, the focus of their studies centered on student

---

achievement gains in the public and private sectors as well as the student, family, and school characteristics possibly related to them.

One of the most striking results of the second wave of research concerned how little students seemed to learn during the last two years of high school.41 For example, of the 115 questions on reading, vocabulary, general and advanced mathematics, writing, and science tests, the average gain for all students across sectors was only 6.17 standardized points. That is, students on average answered 6.7 more questions correctly on the senior test than on the sophomore test. The average gain in test scores for students in the Catholic schools was 7.83 compared to 6.05 for the students in the public schools, both very modest gains. Many have concluded that the learning which took place at the secondary level was far less than that which occurred at the elementary level. Such conclusions need to be qualified since it is questionable whether the tests measured student improvement in the appropriate subjects of high school learning, such as Biology, Chemistry, Physics, Algebra, or the Humanities. It is plausible that high school students might gain very little in the skills learned at the primary level but might significantly improve their skills in the studies appropriate to the high school level. Unless tests are designed to measure competency in subjects appropriate to the secondary level, it will be difficult for researchers to determine the effectiveness of educational programs in the private or public sectors.42

Despite serious limitations to the available data, researchers continued to find that sophomore test scores and many student background characteristics were significantly positively related to senior test scores. In one study by J.D. Willms, senior test scores were regressed on a set of family background characteristics, socioeconomic status, school type, gender, race, and handi-

42 I will examine the NELS:88 data base to see what tests are administered at each level.
capped status. In all of the four models estimated, sophomore test results were significantly related to senior test results. Student Catholic school attendance was also significant and positively related to senior test scores, but the relative magnitude of the variable’s coefficient was small. It is reasonable to expect that as researchers add more significant explanatory variables to models of student achievement, the significance of Catholic school attendance will diminish. However, in this study by Willms, it is interesting to note that many student composition variables, such as school mean socioeconomic status, and the percent of student enrollment composed of minorities, were reduced to insignificance once sophomore achievement was controlled. Eventually these variables were removed from his final model.

As some researchers studied the relationship between prior student achievement and senior test scores, others explored the relationship between senior test scores and tracking as well as course taking. The results of these studies might be of particular interest to administrators and policy makers since course taking and student tracking can potentially be manipulated through the implementation of school policies. One study by Gamaron (1987), using only public school students, designed a model similar to the one used by Willms, but added variables to control for each student’s academic track, whether the student dropped out, and the number of academic courses the student took. Three major results emerged. First, as expected, the sophomore test scores were significantly positively related to senior test scores. Second, the coefficients on each of the student tests were almost identical to those in Willms’ study. Since Willms’ model included only public school students, sector differences seemed to have little if any significant relationship to student achievement. Third, when the number of courses taken, academic track, and whether the student

---

dropped out were controlled, the variable measuring socioeconomic status (a composite of family income, mother’s and father’s education, father’s occupation, and home educational resources) became insignificantly related to all student test scores. Finally, academic track and course taking were significantly positively related to student test score results, but of the two, the coefficient for academic track was larger.

Lee and Bryk (1988) added to the literature by specifying a complex model of four regression equations estimated separately for students in the Catholic and public schools. The first equation regressed academic track on academic background (including eighth grade college expectations), social class, and minority status. The second equation regressed sophomore mathematics scores on a set of student background characteristics and placement in an academic or remedial track. The third and fourth equations regressed separately the number of advanced mathematics courses and senior mathematics courses taken respectively, on all prior variables. Their results indicated that there was a significant and positive relationship between sophomore test scores and both the number of advanced mathematics courses taken and academic track. These results were consistent with those of Gamaron and Willms. There was also a positive and significant relationship between Catholic sector and senior mathematics test scores, but the sector relationship was less significant than that of the sophomore test scores. The main goal of this analysis was to compare sector differences in tracking and course taking, and to test for the significance of these in relation to student prior social and academic backgrounds.

Gamaran and Willms were also interested in assessing whether the student test score results were significantly related to the different types of students in the Catholic and public schools or to differences in school policies. Numerous studies indicated that a greater proportion of Catho-

---

lic school students were enrolled in an academic track and took more academic courses.\textsuperscript{46} Lee and Bryk conclude that it was the strong institutional pull towards academic pursuits which was primarily responsible for this phenomena. This conclusion needs to be seriously qualified since their results do not support it. First, sophomore test scores had a greater impact on senior test results than did Catholic school attendance. Second, the sophomore test scores can be interpreted as prior to high school academic achievement since most of the questions dealt with material taught at the elementary level. Finally, they dealt with mathematics scores only, which were significantly related to academic track and the number of advanced courses taken, rather than to school sector. The impact of Catholic sector policies on student achievement levels is still an unresolved issue.

\textbf{Summary}

The following conclusions emerge from the studies using data gathered for the Coleman Reports. First, the amount of achievement gain between the 10th and 12th grades appear to be very modest. Second, prior achievement, as measured by the sophomore test scores, was significantly positively related to later student achievement reflected in senior test scores. Third, family and student background characteristics were significantly related to student achievement. The coefficients of the variables representing these characteristics were consistently among the largest relative to the other variables in the regression equations. Fourth, students enrolled in an academic track-that is students taking more advanced level courses-consistently learn more, but the estimates vary according to modeling specifications. Finally, the relationship between Catholic school attendance and student achievement is statistically significant, but the size of the coefficient for the Catholic school variable small when models control for prior achievement, student background, tracking, and course taking.

\textsuperscript{46} Coleman et al. (1987). \textit{Public and Private High Schools}.
This review of the literature concerning various inputs to the educational process and their relationship student academic achievement is not exhaustive, but is intended to highlight general trends in the development of the Economics of Education literature. Based on the results of studies conducted in the last decade, some researchers have concluded that some of the practices and school policies which characterize and differentiate private from public schools could lead to greater student achievement in all schools. However, there is little evidence to date that these differences would be as great as popularly assumed. Due to measurement problems and selectivity bias, it is clear that further studies utilizing information gathered for the Coleman Reports will add little to what is known about the significant inputs to the educational process. In order to identify which student, school, teacher, family, and peer characteristics are significantly related to student achievement gains, researchers will need access to data collected over a period of time with repeated measures of achievement appropriate to each level of schooling in each sector. A value-added model of achievement gain can be estimated in order to specify which student, family and school characteristics are significantly related to student learning. The results of this type of study will provide educators and policy makers with information regarding inputs to the educational process which have the strongest impact on student achievement.
CHAPTER III
DISertation MODEL

Introduction

The primary focus of this study is to study the relationship between student achievement and various student, family, school, and peer characteristics. Earlier studies of student achievement often used single equation production functions, which regressed measures of student achievement on a set of individual, family, teacher, school, and peer group characteristics. As economists entered the field, these single equation educational production functions came to be known as "static models" of student achievement. The term "static" referred to the fact that the test scores, used as independent variables in the production function equations, were the results of one set of student tests administered at one point in time.

Static models have been valuable to researchers insofar as they have provided insight to questions concerning the average size and distribution of achievement levels for particular subgroups in the student population or the characteristics of students within these subgroups. While static models have provided descriptive statistics about the number and type of students who are likely to do well or drop out of the educational system, the aggregate nature of the data employed does not allow them to conclusively determine the student background and school characteristics significantly related to student achievement. Furthermore, single equation production functions have been unable to control for previous levels of student achievement or selectivity bias. This has

---

resulted in biased and inconsistent estimators of the explanatory variables, which obscures their true impact.

**Theoretical Model**

In order to identify and test for the significance of various student background and school characteristics as they relate to achievement, a dynamic multiple regression model, which controls for selectivity-bias and previous student achievement, is required. The model is "dynamic" insofar as student achievement is measured using test score results for the same subject, at different periods in each student's academic career. This enables researchers to control for previous student achievement. In this particular case, student eighth grade standardized tests scores in mathematics will be subtracted from the same student's twelfth grade standardized test scores in mathematics. Calculated gain scores, measuring student learning over time, is then regressed on a set of individual, family, teacher, school, and peer group characteristics to test for which are significantly related to student achievement over time. The model is "multiple regression" in so far as it contains a probit regression as well as the student gain score educational production function described above. The probit regression measures the probability that a student will choose a certain type of high school, i.e. public or non-public. The results of the probit regression are used to construct a variable called the "Mills Ratio", which controls for selectivity bias and is included in the educational production function as one of the independent variables. Conceptually, the two-equation system can be modeled in the following way.

\[(3.1) \quad P_t = F(Z_t)\]

\[(3.2) \quad Y_{jlt} = f^*(F_{lt}, S_{lt}, T_{lt}, P_{lt}, I_{lt})\]

In the first stage of this model, equation (3.1), the choice of public or non-public high school, is explicitly analyzed. Essentially, the choice of school (public = 1 or non-public = 0) is a
function of the value a student places on public high school education, represented by the continuous variable $Z_i$. Researchers do not have direct observations on the continuous index of value $Z_i$, but they do have data which distinguishes whether an individual is in one category (high value of $Z_i$) or a second category (low value of $Z_i$). Past studies indicate that the choice of school sector is a linear function of individual student, family, and school characteristics.\textsuperscript{48} The value a student places on public high school education is modeled by equation (3.3).

$$Z_i = \alpha_i + \beta_i X_i + \nu_i$$

In this equation $\alpha_i$ and $\beta_i$ are regression coefficients, $X_i$ is a vector of individual student, family, school, and other background characteristics, and $\nu_i$ is a normally distributed random error term. It is assumed that for each individual student, $Z_i^*$ represents the critical cutoff value which translates the underlying index into a school choice decision. Specifically, a student will choose to attend a public school if $Z_i > Z_i^*$ or a non-public school if $Z_i \leq Z_i^*$. The probit model assumes that $Z_i^*$ is a normally distributed random variable, so that the probability that $Z_i^*$ is less than or equal to $Z_i$ can be computed from the cumulative normal probability function. Combining equations (3.1) and (3.3) we can write,

$$P_i = F(Z_i) = F(\alpha_i + \beta_i X_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_i^*} e^{-s^2/2} ds$$

where $s$ is a random variable normally distributed with mean zero and unit variance. By construction the variable $P_i$ will lie on the $[0,1]$ interval, and represents the probability that an event, such as choice of public high school, occurs. Since this probability is measured by the area under the standard normal curve from $-\infty$ to $Z_i$, the event is more likely to occur the larger the value of $Z_i$.

When using a probit model with individual observations, the most suitable estimation technique is maximum likelihood. Given the fact that there are no individual student measurements for $P_i$, and only low or high measurements for $Z_i$ ($Z_i = 1$ for public high school choice, otherwise $= 0$), there needs to be found parameter estimators for $\alpha_1$ and $\beta_1$ which make it most likely that the choices in the sample would have occurred. If it is assumed that the public high school option is chosen $n_1$ times and the non-public school option is chosen $n_2$ times ($n_1 + n_2 = N$), and if the data is ordered in such a way that the first $n_1$ observations are associated with the public school choice, then the likelihood function has the form

$$l = P_1 \cdot P_2 \cdots P_{n_1} (1 - P_{n_1+1}) \cdots (1 - P_N) = \prod_{i=1}^{n_1} P_i \prod_{i=n_1+1}^{N} (1 - P_i)$$

(3.5)

$$l = \prod_{i=1}^{N} P_i^{y_i} (1 - P_i)^{(1-y_i)}$$

The last expression holds since $Y_i = 1$ for the first $n_1$ observations and 0 for the last $n_2$ observations. $l$ is maximized by substituting for the probability function in (3.5) and then obtaining the estimators $\alpha_1$ and $\beta_1$ by differentiating $l$ with respect to $\alpha_1$ and $\beta_1$, and setting them equal to zero.

$$\frac{\partial l}{\partial \alpha_1} = 0 \quad \text{and} \quad \frac{\partial l}{\partial \beta_1} = 0$$

(3.6)

The vector of $\beta_1$ estimators can then be tested to identify which ones significantly impact school sector choice. This completes the first stage of the multiple regression model.

The second stage of the model specifies a modified form of equation (1) to be

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

(3.7)

where $Y_i$ is the gain in test scores over a specified interval of time, $\alpha$ and $\beta$ are regression coefficients, $X_i$ is a vector of individual student, family, school, and other characteristics, and $\varepsilon_i$ is a
normally distributed random error term. Ordinary Least Squares (OLS) estimation of the production function in equation (3.7) will yield biased and inconsistent estimates of $\alpha$ and $\beta$ if there are omitted variables, such as scholastic aptitude, which could affect school choice as well as achievement gain.\(^49\) In order to correct for the possibility of selectivity bias, the estimators $\alpha_1$ and $\beta_1$, calculated in equation (3.6) using the results of the probit equation (3.4), can be used to construct a special public school choice variable that corrects for potential self-selection based on unmeasured factors. James Heckman has devised a two-stage estimation process that will yield consistent and unbiased estimates of $\alpha$ and $\beta$.\(^50\) Applying the Heckman two-stage process to the results of the probit equation leads to the following production function corrected for selectivity bias.

\[(3.8) \quad Y_i = \alpha + \beta X_i + \omega P_i^* + \mu_i\]

In this equation $Y_i$ is a measure of the test score gain for student $i$, $\alpha$, $\beta$, and $\omega$ are regression coefficients, $X_i$ is the vector of student characteristics, $P_i^*$ is the public school choice variable known as the Mills Ratio which corrects educational production function for selectivity bias, and $\mu_i$ is the residual where

\[(3.9) \quad E(\mu_i|X_i, Z_i \geq Z_i^*) = 0\]

Using the new public school choice variable $P_i^*$ which corrects for selectivity bias, test score gains can be regressed on it and other background variables yielding consistent estimates of $\alpha$ and $\beta$ free of bias.


Public High School Model Specified

The two equation system described above will be applied separately to students in each of two school sectors, public and Catholic high schools. In the first specification of the model, the probit equation will regress a dummy variable representing the choice of public school (PUBLIC = 1 if public high school is chosen, otherwise PUBLIC = 0) on all students in the sample. The results of the probit equation will be used to calculate the Mills Ratio for public school choice, which in turn will be included as an independent variable in the educational production function in the second part of the model. Calculated standardized public school student gain scores in mathematics, reading, science and history will be regressed separately on the Mills Ratio and other student, family, school, and peer group characteristics. Ordinary Least Squares will be used to test for the presence of significant relationships between student achievement gains and various background characteristics. The two regression model for public school student gain scores in mathematics, adapted from equations (3.3) and (3.8), may be written in the following way:

Probit Regression For Public High School Choice

\[ P_i = \alpha + \beta X_i^+ + \nu_i \]

- \( P_i \) = dummy variable indicating the choice of public school = 1, otherwise = 0.
- \( \alpha, \beta \) = regression coefficients
- \( X_i^+ \) = vector of student and family, teacher, school, and peer characteristics, plus one variable significantly related to the choice of public high school but not related to student achievement gains, and not included in the educational production function.
- \( \nu_i \) = residual

Educational Production Function For Public High School Students, Mathematics Gain Score.

\[ \Delta M_i = \alpha + \beta X_i^+ + \omega P_i^+ + \mu_i \]

- \( \Delta M_i \) = Calculated student mathematics gain score between the 8th and 12th grades.
- \( \alpha, \beta, \omega \) = Regression coefficients.
- \( X_i \) = Vector of student background characteristics except for the public school choice.
- \( P_i^+ \) = The Mills Ratio for public school choice.
- \( \mu_i \) = residual
Catholic High School Model Specified

The two regression model for Catholic high school student gain scores in mathematics, adapted from equations (3.3) and (3.8), also may be written in the following way:

**Probit Regression For Catholic High School Choice**

\[ C_i = \alpha + \beta X_i^* + \nu_i \]

- \( C_i \): dummy variable indicating the choice of Catholic school = 1, otherwise = 0.
- \( \alpha, \beta \): regression coefficients
- \( X_i^* \): vector of student and family, teacher, school, and peer characteristics, plus one variable significantly related to the choice of Catholic high school but not related to student achievement gains, and not included in the educational production function.
- \( \nu_i \): residual

**Educational Production Function For Catholic School Students, Mathematics Gain Score.**

\[ \Delta M_i = \alpha + \beta X_i + \omega C_i^* + \mu_i \]

- \( \Delta M_i \): Calculated student mathematics gain score between the 8th and 12th grades.
- \( \alpha, \beta, \omega \): Regression coefficients.
- \( X_i \): Vector of student background characteristics except for Catholic school choice.
- \( C_i^* \): The Mills Ratio for Catholic school choice.
- \( \mu_i \): residual

Equations (12) and (14) below are easily modified to include student gain scores in reading, science, and history.

**Identification of Models**

In order to obtain reliable parameter estimates from the reduced form equations, the system must be properly identified. A necessary and sufficient condition for identification requires that the probit regression and educational production function are independent of each other.\(^{51}\) This condition will be satisfied if the number of excluded exogenous variables in the system of equations equals or exceeds the number of included endogenous variables minus one. Since this model has
two included endogenous variables (school choice and student test score gains), there must be one excluded exogenous variable in the system. The two-equation system will be identified if one exogenous variable (a variable determined outside of the two-equation system) is significantly related to school choice and included in the probit equation, but not significantly related to student achievement gain and therefore excluded from the educational production function. In the public and Catholic school models, the Catholic religion dummy variable will be used to identify each two-equation system, the logic being that religious affiliation is likely to be related to the choice of high school but not achievement gain.²²

The effects of religious affiliation on the returns to human capital has received some attention by researchers over the past few decades, but most of the research concerned the impact of religious affiliation on earnings differences controlling for religious upbringing, education, and various family background characteristics.²³ A number of authors have noted that earnings and rate of return to education vary across religious groups. For example, Taubman found that Jews have higher earnings and higher rates of return than do non-Jews.²⁴ However, there is also evidence that religious denomination alone does not affect behavior, but rather the strength of religious belief. For example, the results of studies conducted on teen sexual activity have indicated that when religious service attendance is controlled, religious affiliation was not significantly related to their probability of teenage women getting pregnant. This hypothesis can be tested by including measures of student religiosity in the educational production functions. If the regression results indicate no significant relationship between student gain scores and student religiosity, but if there is a sig-

significant relationship between religiosity and school choice, then it could be a valid identifying variable in the two-equation models.

Remarks

The use of instrumental variables to control for self-selection into the Catholic sector will provide unbiased and consistent estimates of the determinants of educational growth. If there are differences in student achievement significantly related to school type, this model might provide insight in identifying which inputs to the educational process are significant and whether they are a function of school policy. For example, many researchers have noted that there is a strong institutional pull within the Catholic high schools towards academic pursuits (Lee and Bryk, p. 91). Studies have indicated that not only was there a larger proportion of students in an academic track in the Catholic sector, but the students’ background were less strongly related to this placement than in the public sector. It is plausible that although family background variables might be significantly related to academic achievement, school policy may be at least as important. The inference is that if public schools implement policies regarding course taking which are directed towards academic pursuits, significant achievement gains could result. The ease of implementing academic policy changes in the public school system might be very difficult given its complex set of regulations at the federal and local level. The independent and parochial nature of the “system” of Catholic schools allows for more flexibility and probably enables them to more easily implement changes which favor student achievement.

The impact of tuition on the choice of school and student achievement requires further empirical investigation. Due to selectivity factors, OLS estimation of educational production functions which include a variable for tuition will result in biased and inconsistent estimators. For ex-

ample, families whose reservation tuition (the maximum tuition a certain family would pay for private education) is below that charged by the private school will opt to send their children to the public sector schools. Since there is no information on the tuition those families are actually willing to pay for education, testing for the impact of tuition on cognitive achievement or school choice using OLS estimation techniques will result in biased and inconsistent estimators. In order to correct for selectivity-bias, the Mills Ratio would need to be included in the educational production function in order to insure that the coefficients of the tuition and other variables are unbiased and consistent.

A second problem with the use of a tuition variable in regression analyses involves the fact that it does not capture the total value a family places on public or private education. As mentioned above, families often self-select into relatively homogeneous residential communities partially because of the package of public goods offered there. For instance, the perceived quality of an area’s public schools could significantly impact a family’s residential choice. Their reservation tuition would then be reflected in housing tax expenditures which help support the schools. A different family living in the same community might choose to send their children to private schools because their reservation tuition might be greater then the combination of their tax expenditures on housing and the private school tuition. If these are in fact some of the underlying forces contributing to school choice, then it would be reasonable to conclude that the students in the private sector consistently outperform their public school counterparts because their family’s valuation of educational progress is greater then those in the public sector. Testing this hypothesis would require information on family tax expenditures which might not be available. Furthermore, even with this information, the data would be discontinuous for families who selected the private school option.
Perhaps the best way to deal with the discontinuity in the tuition data would be to use the Heckman two-step procedure to calculate the Mills Ratio and correct for the discontinuity.

If the independent parochial nature of the Catholic school system does significantly improve student achievement, and if this school system is to serve as a kind of model for the public sector, then de-regulation at the federal level must be accompanied by increased regulation at the state and local levels. Catholic elementary school policies are usually set by the pastor of the parochial school with input from the bishop, teachers, and parents of the students who attend them. If the pastor values strong academic policies, and most of them do, then high academic standards will be demanded. If the parents attend the parochial school church regularly, the values of a good education as well as a strong work ethic might be reinforced on a regular basis. The weekly reinforcement of a strong work ethic and the value of a good education might be the most significant factor motivating parents to be involved in the educational formation of their children. Testing for a significant relationship between parental church attendance and academic achievement of their children might provide insight to factors which promote learning. If the relationship is significant, then simply changing public school policies will not have an impact on achievement without some further reinforcement by way of adult education.
CHAPTER IV

DATA

Introduction

Beginning in 1988, with a cohort of 25,000 eighth graders attending 1,000 public and private schools across the country, the National Education Longitudinal Study of 1988 (NELS:88) was designed by the National Center for Education Statistics (NCES) to provide longitudinal data about critical transitions experienced by students as they leave eighth grade settings, progress through high school (or drop out), enter and leave post-secondary institutions, and enter the workforce. The 1988 eighth grade student cohort was interviewed and tested in the spring of 1988, 1990, 1992, and 1994. The fourth follow up is scheduled for the spring of 1996, the results of which will be made available sometime in the spring of 1997. Policy relevant data about educational processes and outcomes were collected during the four survey periods, especially as it pertained to student learning, indicators of dropping out, and opportunity to learn. The longitudinal design of NELS:88 permits researchers to examine the change in students' lives and the role of schools, teachers, peers, and family in promoting positive academic growth. In particular, this study intends to use the relevant data collected from the students in the 1988 (eighth grade) and the 1992 (senior year) surveys to investigate students' academic achievement over time and how this is related to various school, student and other background characteristics.

---

NELS:88 Base Year Survey (Eighth Grade)

Student Questionnaire

The data collection instruments for the 1988 base year study consisted of four separate questionnaires and a battery of eighth grade tests. A forty-five minute self-administered student questionnaire was completed by eighth grade students in the classrooms of their schools. The student questionnaire was designed to collect information about a wide variety of topics, including the student’s and parents’ background, language use, family background, perceptions of self, plans for the future, jobs and household chores, school life, school work, and school activities. Students completed a series of cognitive tests, which were administered in a single group session. The combined tests included 116 items to be completed in eighty-five minutes. The Educational Testing Service (ETS) developed the cognitive test battery. The eighth grade tests are described briefly below:

Reading (21 items, 21 minutes): consists of five short passages followed by comprehension and interpretation questions.

Mathematics (40 items, 30 minutes): consists of quantitative comparisons and other questions assessing mathematical knowledge.

Science (25 items, 20 minutes): consists of questions assessing science knowledge and scientific reasoning ability.

History/Social Studies (30 items, 14 minutes): consists of questions assessing knowledge of U.S. history, civics, and government.

Parent Questionnaire

A self-administered thirty-minute questionnaire was completed by one of the student’s parents on about the same date that the student questionnaire and eighth grade tests were administered. The instructions in the questionnaire and accompanying letter directed the most knowledgeable parent (or guardian) to complete the questionnaire. The most knowledgeable parent was defined as
the parent who knows the most about the student's educational activities and related behaviors. In accordance with this definition, the respondent was self-selected.

The parent questionnaire was designed to collect information from parents about factors which might influence educational attainment and participation. The questions focused on family background and socioeconomic characteristics, and on the character of the home educational support system. In addition, the parent instrument collected data relative to parental behaviors and circumstances with which the student may not be familiar, such circumstances as parental education and occupation, and items relating to income and religious affiliation. English and Spanish language versions of the questionnaire were made available to parents.

Teacher Questionnaire

A self-administered teacher questionnaire was completed by selected teachers responsible for instructing sample students in two of the four test subjects (mathematics, science, reading, and history). The teacher questionnaire was designed to collect information in three areas: teachers' perceptions of the sampled students' classroom performances and personal characteristics; curriculum content of areas that they teach; and teachers' background and activities. Teachers were asked to respond to the questionnaire items in relation to a specific list of sampled eighth students enrolled in their classes. Part I, concerning student information, asked the teachers to indicate which of the sampled students they had during the 1987-1988 academic year, and whether or not the student had various school-related problems and handicaps. Part II, concerning class information, required the teacher to respond to a series of course-related questions regarding the set of classes they taught to one or more of the sampled students. Part III, concerning teacher background and activities, requested teachers to provide general background information about themselves and their school. Westat, NORC's (National Opinion Research Center) subcontractor, pre-
pared the teacher questionnaire under the direction of NORC and the National Center for Education Statistics (NCES).

**School Administrator Questionnaire**

A self-administered, forty minute school administrator questionnaire was completed by the school principal, headmaster, or other knowledgeable administrator designated by the principal. The questionnaire was designed to collect information about school, student, and teacher characteristics; school policies and practices; the school’s grading and testing structure; school programs and facilities; parent involvement in the school; and school climate. The primary purpose of the school administrator questionnaire was to gather descriptive information about the educational setting and environment in terms of enrollments and educational offerings, as well as specific school policies. The information obtained through the school administrator questionnaire provide supplemental information to that provided by the student questionnaire so that student outcome and achievement data can be considered in terms of the educational setting. NORC and its subcontractor Westat collaborated in designing the instrument.

**Sample Design**

The base year eighth grade survey employed a two-stage, stratified sample design, with schools as the first stage unit and students within the schools as the second-stage unit. Within each stratum, schools were selected with probabilities proportional to their estimated eighth grade enrollment. Within each school approximately 26 students were randomly selected. In schools with fewer than 24 eighth graders, all eligible students were selected. From a national frame of about 39,000 schools with eighth grades, a total of 1,734 schools were selected, of which 1,057 participated.
In designing a sampling frame for a survey one can either use an explicit or an implicit list of the students to be sampled. The creation of an explicit list of all eighth grade students in the United States would have been a very difficult, if not impossible task. NORC therefore elected to use an implicit list of public and private schools in the U.S. The most readily and available source for a complete and accurate frame was the database compiled by Quality Education Data, Inc. (QED) of Denver Colorado. This database includes public, private parochial and non-parochial schools. The QED records were successfully employed in five test states and proved to be highly accurate.

The study excluded certain kinds of students: specifically, mentally handicapped students and students not proficient in English; and students having physical or emotional problems that would make participation in the survey unwise or unduly difficult. Just as certain students were considered to be ineligible, so too certain kinds of schools were ineligible for selection. The eligible populations of schools are restricted to “regular” schools in the United States, private as well as public. Excluded from the sample are Bureau of Indian affairs (BIA) schools, special education schools for the handicapped, area vocational schools that do not enroll students directly, schools for dependents of U.S. personnel overseas, and students educated at home or in private tutorial settings. Since this survey was designed to minimize overlap with the National Association of Educational Progress (NAEP) sample for the 1987-1988 school year, any schools selected for the NAEP survey were not eligible for NELS:88 selection. Exceptions to this principle could have occurred in practice since not all schools originally selected for NAEP agreed to participate, and therefore sub-

stitute schools were selected. While NORC was able to eliminate the originally selected NAEP schools from the survey, it was not able to screen out NAEP substitute schools.

**Missing Data**

Although the sample design yields, in theory, a sample that mirrors the population within the sample, non-response can introduce distortions in practice. In the NELS:88 base year survey there were two stages of sample selection and therefore two stages of potential nonresponse. During the base year survey, schools were asked to permit the selection of eighth grade students from school rosters and to hold makeup days for the collection of student data. Not all selected schools agreed to take part in the study. In addition, not all of the individual students selected for the sample within cooperating schools (or the teachers or parents linked to these students) provided the data sought from them. Shortened versions of the NELS:88 school administrator questionnaire were sent to nonresponding schools in the poll of original selections. Almost all of these schools provided data. There were some cases when information not provided by the school administrator or the student was obtained from other sources.\(^{57}\)

**NELS:88 Second Follow-up Survey (High School Seniors)**

The data collection instruments for the NELS:88 second follow-up were similar in content and form to those utilized in the prior waves.\(^{58}\) The second follow-up instruments consisted of a student questionnaire and cognitive tests, dropout, parent, teacher, and school administrator questionnaires. The new student supplement, added in the first follow-up survey to elicit demographic information from newly freshened students, was again administered in the second follow-up. Al-

---

\(^{57}\) National Center for Educational Statistics. (1990). *1988 Base Year: Student Component Data File User's Manual*, p. 25. One example of this is when information from the QED datafile, used to create the data frame, was also used to fill in missing information about the grade range of the school.

though this supplement was included in the data base, it will not be used in this study. Only students present in all three waves, and making normal progress will be included in this study.

*Student Questionnaire*

Sample members who attended school during the spring term of the 1991-1992 school year were administered a student questionnaire either at an in-school or off-campus survey session. The self-administered questionnaire, which took approximately one hour to complete, collected information on a wide range of topics, including students’ background, language use, home environment, perceptions of self, occupational or postsecondary educational plans, jobs and household chores, school experiences and activities, work, and social activities. The second follow-up student and dropout questionnaires were available in both English and Spanish.

In addition to the student questionnaire, students completed a series of cognitive tests, also administered at in-school or off-campus survey sessions. The combined tests, covering four subject areas, included 116 items to be completed in 85 minutes. The cognitive tests are described briefly below:

**Reading Comprehension** (21 questions, 21 minutes): The subtest contained five short reading passages or pairs of passages, with three to five questions about the content of each. Questions encompassed understanding the meaning of words in context, identifying figures of speech, interpreting the author’s perspective, and evaluating the passage as a whole.

**Mathematics** (40 questions, 30 minutes): Test items included word problems, graphs, equations, quantitative comparisons, and geometric figures. Some questions could be answered by simple application of skills or knowledge, others required the student to demonstrate a more advanced level of comprehension and/or problem solving.

**Science** (25 questions, 20 minutes): The science contained questions drawn from the fields of life science, earth science, and physical science/chemistry. Emphasis was placed on understanding of underlying concepts rather than the retention of isolated facts.

**History/Geography/Citizenship** (30 questions, 14 minutes): American history questions addressed important issues and events in political and economic history from colonial

---

59 Ibid., pp. 18-28.
times through the recent past. Citizenship items included questions on the workings of the federal government and the rights and obligations of citizens. The geography questions touched on patterns of settlement and food production shared by other societies as well as our own.

NORC's subcontractor, the Educational Testing Service (ETS), developed the cognitive test battery for the second follow-up survey. Six forms of the cognitive test battery were produced in the second follow-up, each comprising a different combination of mathematics and reading difficulty levels. Each sample member's test form was determined by his or her scores on the base year and/or first follow-up mathematics and reading test results. The purpose of the multilevel design of the second follow-up cognitive test battery was to guard against ceiling and floor effects which may occur when testing must span four years of schooling. This adaptive approach tailors the difficulty of the reading and mathematics tests to the ability of the respondent, thereby leading to a more accurate measurement than a single level design.

In the second follow-up survey, data collection procedures involved mailing a self-administered questionnaire to school principals, teachers, and parents.60 Two weeks after the initial mailing, a postcard reminder was mailed to respondents who had not yet returned a completed questionnaire. Two weeks after the postcard was mailed, telephone interviewers called the respondents to prompt them for the return of the completed questionnaire. Three weeks after the telephone prompt, telephone interviewers began calling any respondents who had not yet completed a questionnaire to attempt to complete an interview over the telephone. For the course offerings and transcript surveys, data collection forms were mailed to principals and other school staff, with follow-up over the telephone and in person.

---

60 Ibid., pp. A-7 to A-10.
**Parent Questionnaire**

In the second follow-up parent survey, a forty minute questionnaire was mailed to the parent or guardian of the NELS:88 student in May 1992. Like the base year parent survey, the instructions in the questionnaire and accompanying letter directed the parent or guardian who was most knowledgeable about the teenager's current situation to complete the questionnaire. Since there were questions in the parent survey which concerned postsecondary educational costs, parent questionnaires were mailed in May 1992 in order to ensure that the parents or guardians would have the necessary information regarding financial aid (for college) to answer these questions fully. Although the parent instrument was designed as a self-administered questionnaire, many completed the survey over the telephone with a trained interviewer.

**Teacher Questionnaire**

In the second follow-up teacher survey, one teacher report was collected for each student attending a NELS:88 school if the student was enrolled in a mathematics or science class. For students enrolled in both a mathematics and a science class, only one teacher report was collected. The subject area of the teacher report collected for students enrolled in both a mathematics and science class was the same subject area of the teacher surveyed for the student in the base year teacher surveyed for the student in the base year teacher survey. Teachers at schools at which the students were surveyed before April 1, 1992 were mailed a questionnaire in early February 1992. Teachers at schools at which the students were surveyed on or after April 1, 1992, were mailed a questionnaire in early March 1992. Two weeks after the teacher questionnaires were mailed, non-responding teachers were prompted for the return of the questionnaire with a postcard reminder. Two weeks after the postcard reminder was mailed to teachers, non-responding teachers were
prompted for the return of the questionnaire over the telephone. Teachers who did not respond after the postcard and telephone prompts were interviewed over the telephone by trained interviewers.

For most students a teacher report was collected from the fall term teacher in the selected subject. However, if students at a school were surveyed on or after April 1, 1992, then the teacher questionnaire was mailed to the spring term teacher for the student. This instrument design was based on the assumption that early in the spring term, the fall term teacher was the most familiar and could most fully assess the student. After April 1, a teacher report was collected from the spring term teacher because at that time the spring term teacher was more likely to have sufficient interaction with the student to make a full assessment of the student in the teacher questionnaire, and the fall term teacher might have difficulty recalling a student he or she had not instructed in several months.

*School Administrator Questionnaire*

In February 1992, school administrator questionnaires were mailed to the principal or headmaster of selected NELS.88 schools with second follow-up sample members still in attendance. Completed self-administered questionnaires and telephone interviews were collected from February through early July 1992. For any interviews conducted after the end of the 1991-1992 academic year, school principals were asked to refer to the previous academic year. As in the base year survey, the school principal or headmaster could delegate all but one of the sections to another knowledgeable school official.

Because questionnaires from school principals were collected in two different modes of data collection, by self-administration and over the telephone, a number of steps were taken to minimize any mode effects. Telephone interviewers were trained to adapt questions in a way which made sense when asked over the telephone. If principals had a copy of the questionnaire, they were
encouraged to read along in the questionnaire as the interviewer asked the questions over the telephone.

*Course Offerings and Transcript Questionnaires*

Course offering documents were collected from selected NELS:88 schools in the fall of 1991. Additional documents were collected as necessary during transcripts collection and processing. The majority of schools provided catalogs with descriptions of the courses offered during the 1991-1992 school year. Various elements were abstracted from the course offerings documents such as course title, course number, duration of the course, credits awarded for the successful completion of the course, and term offered.

In August 1992, transcript survey materials were mailed to the principals of the NELS:88 schools attended or most recently attended by sample members eligible for the survey.\(^{61}\) Because of the variability in transcript format across schools, explicit instructions for transcript preparation were provided. School staff were asked to retrieve from alternate sources any data elements that were not included on the school's transcripts. Transcript preparers were also asked to note any in-school survey sessions day transfers on survey documents, to facilitate the pursuit of additional records from transfer schools.

Two weeks after survey materials were mailed, nonresponding principals were prompted for the return of transcripts with a postcard reminder. Principals who did not return transcripts within three weeks of the postcard prompt were prompted over the telephone. Telephone prompting of nonresponding principals continued from October 1992 through February 1993. Field visits to schools requesting assistance in the preparation of transcripts were collected in February and March.

\(^{61}\) Ibid., pp. A-9 to A-10.
Abstraction of student and course level data from transcripts began in October 1992 and continued through March 1993. Retrieval of missing critical items from school staff occurred concurrently. Courses were coded using the course catalog for the school or district, in accordance with the Classification System of Secondary Courses, updated for the 1990 NAEP High School Transcripts Study. When a school or district catalog was unavailable, courses were coded by title alone.

The student data files released in the second follow-up of NELS:88 may be combined with data from second follow-up surveys of parents, teachers, and school administrators. The most powerful analyses are possible when students are viewed in the context of these fundamental influences across the four year time frame that is available. The NELS:88 files are designed to be merged and used to examine how different student outcomes are related to various structural patterns, as measured by parental, teacher, and other influences, and/or the way these change over time. The contextual data files on teachers, schools, and parents cannot stand alone but are dependent upon and subsidiary to the student files in NELS:88. For example, the second follow-up school component reflects characteristics of the secondary schools to which students in the contextual sample migrated after the eighth and tenth grades. Since these secondary schools were not selected as a representative sample, but are the product of student dispersion patterns, the second (as well as the first) follow-up school data must be used only in conjunction with student data.

Several types of student sample members are included in the second follow-up data files. Among the types of sample members in the student second follow-up data set are: 1) students who were added in the first or second follow-ups to freshen the sample; 2) sample members who have participated in one, two, or three waves of the survey; and 3) Base Year Ineligible sample who were found to be eligible and subsequently included in the second follow-up survey. Five variables
have been constructed to indicate which second follow-up sample members responded to key survey documents since the base year of NELS:88. The constructed variable F2UNIV2A indicates how the second follow-up sample member entered the sample, i.e. as a base year eligible student or a freshened student. Since this study is primarily concerned with students who were sample members in the eighth and twelfth grades, this variable will be used to limit the working data file to fit this criteria (F2UNIV2A = 1). The first variable on all the data files, STU_ID, is a unique seven-digit student identification code. This number remains with the student throughout NELS:88. To merge student records across two or more waves of the survey or between survey components (student, teacher, parent, and transcript) STU_ID is to be used.

**Composite Variables**

Composite variables are constructed in order to enhance analyses. Since research questions frequently require independent or control variables such as urbanicity of the school, the socio-economic status of the family, or the gender of the individual, a large set of classification variables has been carefully constructed and added to the records. Most constructed variables are constructed from two or more sources. These may be combinations of questionnaire items from the same or different NELS:88 data files, in the same survey year or across survey waves. Some composites are drawn from an external sampling resource that is unavailable to users, or utilize an external conceptional scheme in order to rank order or otherwise recode survey data. Some values should change over time; for example, if a student transfers from one school to another, then school control type, urbanicity, region and so on may change as well. The most recent round in which such a variable appears contains the best information for students who participated in that wave of

---

62 Ibid., pp. 114-129.
63 Ibid., pp. 129-132.
NELS:88. Due to the large number of composite variables in the NELS:88 data set, only the ones used most often in this study will be described below.

**Demographic Composites**

Many of the NELS:88 composite variables are respondent characteristics. Composite variables representing the student’s gender, race/ethnicity, and birth date are included in the second follow-up survey. The values for each of these characteristics were taken directly from the second follow-up student surveys. Any cases which contained missing values had gender imputed from the respondent’s first name or if that could not be done unambiguously, the value for student gender was randomly assigned. Similarly, if the value for student race or ethnicity was missing, or was inconsistent with the base year parent survey reports, the values for the parent questionnaire were used. If the value for race was still missing, then the race as identified by the school roster was used. Two composite variables representing the student’s birth month and birth year are included in the second follow-up files and will be used to calculate the student’s age in the twelfth grade.

**Socioeconomic Status**

The second follow-up files contain a composite continuous variable “F2SES2” which indicates a sample member’s socioeconomic status in the twelfth grade. This variable was constructed using base year parent questionnaire data, the base year student questionnaire data, or the first or second follow-up student data. F2SES2 utilizes the original Duncan Socioeconomic Index (SEI) and appears only on the parent component data file of the second follow-up survey. The composite is constructed with the values of five standardized components: father’s and mother’s educational levels, father’s and mother’s occupations, and family income. For cases without parent data, stu-

---

64 Ibid., pp. H-1 to H-38.
65 Ibid., pp. H-10 to H-12.
dent data were used from either the base year student questionnaire or the first or second follow-up new student supplement. F2SES2Q is simply the quartile to which the student belongs with values ranging from one through four. This version of the composite variable will be used in the model’s regression equations.

**Composites of School-level Characteristics**

The composites of school-level characteristics provide information on key characteristics of sample members' second follow-up school. G12CTRL1 classifies the student's second follow-up school by type: public, Catholic or other private, with private schools divided into other religious, no religious affiliation, or affiliation unknown. The value for this variable is based on the report from the school administrator questionnaire. G12URBN3 is a three-category composite that reflects the type of place in which the student's public school district, Catholic diocese, or, for other private schools, county is located. The categories are urban, suburban and rural. The information was obtained from QED (Quality Education Data Inc.), or when missing, looked up in the U.S. Bureau of the Census. *Statistical Abstract of the United States: 1992* (112th edition), Washington DC, pages 896-904, and added to the files. G12STATE identifies the state in which each sample member's school was located. This variable is located only in the restricted use files. A federal license was obtained from the U.S. Department of Education to gain access to this and other restricted information.

**Cognitive Test Variables.**

The cognitive test battery administered to students in the eighth, tenth, and twelfth grades consisted of multiple choice tests in four subject areas: reading comprehension, mathematics, sci-

---

67 Similar variables representing school type and location were obtained from the restricted use files for base year eighth grade schools. These variables were also included in the model’s regression equations.
enience, and history/citizenship/geography. In the base year (eighth grade) all students received the same set of tests. The base year test results showed a wide range of student achievement. This diversity is expected to increase as students progressed through high school with some taking advanced courses and making substantial gains, while others remaining at a relatively low level. A single test form administered to all students in the follow-up surveys would have the potential for serious "ceiling" and "floor" effects, that is, many students getting all items correct because the test was too easy for them, while others could only guess at most of the questions because they lacked sufficient background. When this situation occurs, it is impossible to accurately assess the level of achievement for the highest and lowest scoring students.

The reading and mathematics tests were selected for development of multiple forms targeted to students' ability levels in the second follow-up. The reading test was chosen because the time burden of reading passages before questions about them could be answered meant that relatively few test items could be administered in the time allotted for the test. With the smallest number of items of any subject area, the reading test could least afford any "wasted" questions: those that were too hard or too easy for the test taker. Two forms of the reading test were developed; the easier form was administered to students who had scored below the mean in the first follow-up (sophomore) test, while those scoring above the mean received a set of items that was, on average, more difficult.

In the case of the mathematics test, the need for multiple forms was based on the diversity of exposure to course work that could be expected by senior year. Academic track students would have, by this time, taken courses in algebra, geometry, and higher-level mathematics. Those in general or vocational programs might have only taken general or business math, essentially arithmetic,

---

or none at all. Unlike science or history, where many topics might have been introduced at a lower level of sophistication in earlier grades, much of the material covered in advanced mathematics courses would be completely unfamiliar to students who had not taken these courses. Three mathematics test forms were administered to students in the second follow-up survey. The easiest and hardest forms were given to students who had scored in the low and high quartile, respectively, in the first follow-up tests; students in the middle half of the distribution received the middle-difficulty test.

**Item Response Theory (IRT) Scoring**

Raw scores achieved on tests which vary in average difficulty are not comparable to each other. For example, a student who took the middle difficulty mathematics form in the second follow-up would probably have answered more questions correctly if he or she had taken the easiest form, and fewer if the hardest form had been administered. Item Response Theory (IRT) is a standard procedure employed to calculate scores that could be compared regardless of which test form a student took. A core of items shared among the different test forms made it possible to establish a common scale. IRT uses the pattern of right, wrong, and omitted responses to the items actually administered in a test form, and the difficulty, discriminating ability, and guess-ability of each item, to place each student on an ability scale. Once student ability is estimated using an IRT model, it is possible to predict what score a student with given ability would have achieved on any subset of test items used to calibrate the IRT scale.69

---

69 Item Response Theory (IRT) rests on two basic assumptions: 1) The performance of an examinee on a test item or question can be explained by a set of factors called traits or abilities; and 2) the relationship between examinees' item performance and the set of traits underlying item performance can be described by a monotonically increasing function called an *item characteristic curve (ICC)*. This function specifies that as the level of the trait or in this case student ability increases, the probability of a correct response to an item increases. Many possible IRT models exist, differing in the mathematical form of the item characteristic function and the number of parameters specified in the model. All IRT models contain one or more parameters describing the item and one or more parameters describing the examinee. Log Likeli-
In each of the four subject areas, the IRT scale was calibrated using PARSCALE software. The test responses of those sample members who completed a test in the base year, first follow-up, and second follow-up surveys were used for the calibration of the IRT model. Estimates for the coefficients of the various parameters describing the item (level of difficulty, and the like) were computed for each item that appeared in any one of the test forms administered in the first three waves. There were a total of 54 items (questions) in reading, 81 in mathematics, 38 in science, and 47 in history. Holding the estimated coefficients for the parameters describing the test item fixed, Bayesian estimates of student ability were obtained for all test takers at all three points in time. The procedure used takes into account group membership (year and test form) in order to minimize floor and ceiling effects. These ability estimates were used in conjunction with the item parameters to compute the IRT scores in the database.

This score is only an estimate of how many correct responses a test taker would have given if he or she had taken the test in any of the three waves of surveys. The IRT-based estimate is the probability of a correct answer, given the student’s demonstrated ability and the parameters of the item, summed over all the test items. This sum is not an integer, but can be interpreted as an estimated probability of the number of correct answers. The highest possible score is not the total

hood functions are used to obtain maximum likelihood estimates for coefficients of parameters contained in the model, similar to the procedures used in probit regression estimation. For a detailed discussion of the various IRT models and their estimation see Hambleton, R., Swaminathan, H., Rogers, J., “Fundamentals of Item Response Theory” Sage Publications, Inc., 1991, pp. 7-52.

70 The problem of obtaining maximum likelihood estimates in some cases can be overcome if a Bayesian estimation procedure is used. The basic idea is to modify the likelihood function to incorporate any prior information concerning student ability parameters. For example, it might be said, based on past experience, that student ability 9 is distributed normally with mean $\mu$ and standard deviation $\sigma$. In this case the prior information can be expressed in the form of a density function and denoted as $f(\theta)$. Bayes’ theorem states that the probability of an event A given B is $P(A|B) \propto P(B|A)*P(A)$ where $P(A)$ is the prior probability of event A given B. The above relationship is also true for density functions, where A is 9 and B is the observed item response pattern $\mu$. Bayes’ theorem can be written as $f(\theta|\mu) \propto f(\mu|9)*f(\theta)$. If $f(\mu|9)$ is the likelihood function then $f(\theta|\mu) \propto L(\mu|9)*f(\theta)$. The revised likelihood function $f(\theta|\mu)$ is called the posterior density and its mode is the “most probable” value for 9 and can be taken as an estimate of 9. For more details concerning Bayesian estimates see “Fundamentals of Item Response Theory” pp. 38-39.
number of test items for the subject area. The lowest score is not zero, but is an estimate of how many test items a person of extremely low ability might have guessed correctly.

Student gain scores in each of the four subject areas will be calculated and regressed on the Mills Ratio, to correct for selectivity bias, and other student, family, teacher, school, and peer group background variables. Three sets of student gain scores will be calculated for each subject area. The three calculated gain scores will include measurements for student achievement gain between the eighth and tenth grade, the tenth and twelfth grade, and the eighth and twelfth grade respectively. These gains in overall achievement over time will be calculated using the IRT-estimated Number Right student scores included in the data base.
CHAPTER V

PROBIT REGRESSION RESULTS FOR PUBLIC HIGH SCHOOL CHOICE

Selectivity Bias

One of the key issues regarding the comparison of improvements in academic outcomes between public and non-public school students is the possible existence of selectivity bias. Selectivity bias might result from the inability to control for scholastic aptitude or other factors such as parental motivation and encouragement, factors which relate to student academic performance. It has been strongly contended that these and other unmeasured student variables could account for the differences in academic achievement between public and non-public school students. As described previously in Chapter III, a two equation model has been constructed for this analysis to control for a potential correlation between unmeasured determinants of public school choice and student achievement outcomes. In the first equation of this model, the choice of public or non-public school (public = 1 or non-public = 0) is regressed on a set of student, school, peer, teacher, and family characteristics (e.g., race, eighth grade sector, family type, etc.). Due to the discrete nature of the choice process, a probit regression is required. Since probit regressions operate on individual cases, a process called listwise deletion is used. This procedure confines analysis to those cases with complete data on all variables in the regression so that all relevant statistics are based on the same body of data.

Probit Regression For Public High School Students

(5.1) \[ P_i = \alpha + \beta X_i + \nu_i \]
\( P_i \) = dummy variable indicating the choice of public school = 1, otherwise = 0.
\( \alpha, \beta \) = regression coefficients
\( X_i^* \) = vector of student and family, teacher, school, and peer characteristics, plus the single-adult variable significantly related to the choice of public high school but not related to student achievement gains, and not included in the educational production function.
\( v_i \) = residual

The results of the probit equation (5.1) are then used to construct a special public school choice variable called the Mills Ratio. The Mills Ratio is constructed in order to control for the potential correlation between unmeasured determinants of public school choice and student achievement outcomes. Student gain scores, which are calculated by subtracting each student’s eighth grade score for each of the standardized tests from the same student’s twelfth grade score, are regressed on the calculated Mills Ratio and other background variables. Inclusion of the calculated Mills Ratio in each of the student gain score regressions for mathematics, reading, science, and history will insure that the coefficients of the background variables will be efficient and unbiased.

In order to insure that the two-equation model is properly identified, at least one variable which is related to the choice of school but not to improvement in student academic outcome needs to be specified. In the present case, a variable representing whether or not the student lives in a single-adult household was used as the identifying variable. This choice was based on the results of the probit and student gain score regression equations. On the one hand, the probit results indicate that public school choice was significantly negatively related to students living in a single-adult household. On the other hand, student gain score regression results indicated no significant relationship between student achievement and the number of adults (one or two) living in the student’s home. At first glance this result may be somewhat surprising since one would expect that economic
Table 1. Probit Regression Results for Public High School Choice on Background Variables:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Pr&gt;Chi</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE **</td>
<td>-0.183</td>
<td>0.039</td>
<td>0.0001</td>
</tr>
<tr>
<td>FEMALE</td>
<td>0.005</td>
<td>0.046</td>
<td>0.9096</td>
</tr>
<tr>
<td>BLACK</td>
<td>-0.035</td>
<td>0.082</td>
<td>0.6704</td>
</tr>
<tr>
<td>ASIAN **</td>
<td>0.407</td>
<td>0.106</td>
<td>0.0001</td>
</tr>
<tr>
<td>HISP **</td>
<td>0.168</td>
<td>0.078</td>
<td>0.0330</td>
</tr>
<tr>
<td>Native American</td>
<td>-0.156</td>
<td>0.230</td>
<td>0.4968</td>
</tr>
<tr>
<td>P8GRADES **</td>
<td>0.012</td>
<td>0.003</td>
<td>0.0004</td>
</tr>
<tr>
<td>PCOLL **</td>
<td>0.170</td>
<td>0.077</td>
<td>0.0268</td>
</tr>
<tr>
<td>PFINCOLL **</td>
<td>0.135</td>
<td>0.048</td>
<td>0.0052</td>
</tr>
<tr>
<td>PCATH</td>
<td>-0.091</td>
<td>0.060</td>
<td>0.1290</td>
</tr>
<tr>
<td>S8CATH **</td>
<td>-2.240</td>
<td>0.091</td>
<td>0.0001</td>
</tr>
<tr>
<td>S8PRIVAT **</td>
<td>-2.422</td>
<td>0.095</td>
<td>0.0001</td>
</tr>
<tr>
<td>S8NAIS **</td>
<td>-3.331</td>
<td>0.089</td>
<td>0.0001</td>
</tr>
<tr>
<td>S8RURAL</td>
<td>-1.170</td>
<td>0.079</td>
<td>0.0001</td>
</tr>
<tr>
<td>S8SUBURB **</td>
<td>-0.985</td>
<td>0.060</td>
<td>0.0001</td>
</tr>
<tr>
<td>S8SALARY **</td>
<td>-0.064</td>
<td>0.020</td>
<td>0.0015</td>
</tr>
<tr>
<td>SH2SUBBR **</td>
<td>1.861</td>
<td>0.060</td>
<td>0.0001</td>
</tr>
<tr>
<td>MAEXPCT</td>
<td>-0.007</td>
<td>0.037</td>
<td>0.8522</td>
</tr>
<tr>
<td>PAEHCPCT</td>
<td>0.009</td>
<td>0.037</td>
<td>0.8010</td>
</tr>
<tr>
<td>SINGLE *</td>
<td>-0.132</td>
<td>0.069</td>
<td>0.0556</td>
</tr>
<tr>
<td>FAMSIZ</td>
<td>-0.005</td>
<td>0.017</td>
<td>0.7887</td>
</tr>
<tr>
<td>SESQ **</td>
<td>-0.064</td>
<td>0.026</td>
<td>0.0130</td>
</tr>
<tr>
<td>CATCOST</td>
<td>0.006</td>
<td>0.051</td>
<td>0.9120</td>
</tr>
<tr>
<td>RELCOST</td>
<td>-0.014</td>
<td>0.008</td>
<td>0.0994</td>
</tr>
<tr>
<td>PRIVCOST **</td>
<td>0.049</td>
<td>0.013</td>
<td>0.0001</td>
</tr>
<tr>
<td>INTERCEPT **</td>
<td>4.561</td>
<td>0.749</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Log of the likelihood function: -2332.876
Degrees of Freedom: 25
Number of observations used in calculations: 13,698
** significant at the 5% level
CHART 1: SPECIFICATION OF PROBIT MODEL ESTIMATED FOR CHOICE OF PUBLIC SCHOOL

I. **Dependent Variable:**

Choice of Public School (public school = 1, 0 otherwise)

II. **Independent Variables:**

AGE = student’s age  
FEMALE = female  
BLACK = African American  
ASIAN = Asian Pacific islander  
HISP = Hispanic  
P8GRADES = student grade composite in eighth grade  
PCOLL = student intends to attend college  
PFINCOLL = student intends to graduate college  
PCATH = student is a member of the Catholic church  
S8CATH = student graduated from Catholic grade school  
S8PRIVAT = student graduated from non-Catholic grade school  
S8NAIS = student graduated from a National Association of Independent Schools grade school  
S8SUBURB = student graduated from suburban grade school  
S8SALARY = grade school teacher salary  
S8RURAL = student graduated from a rural grade school  
SH2SUBRB = Student chose to attend suburban high school  
MAEXPCT = mother’s future academic expectations for her children  
PAEXPCT = father’s future academic expectations for his children  
SINGLE = student’s parents are divorced, separated, widowed, never married  
FAMSIZE = size of student’s family  
SESEQ = student’s socioeconomic status measured in quartiles  
CATCOST = Catholic high school tuition  
RELCOST = religious high school tuition  
PRIVCOST = non-religious private high school tuition
constraints would lead single parents to choose the less expensive alternative, namely public schools. However, the presence of significant time constraints on single parents might serve as strong motivation to choose private schools which are flexible enough to meet their needs. Thus the variable representing single adult households was included in the probit equation but not in the four student gain score regressions. The results of the probit regression for public school choice are included in Table 1 whereas descriptions of the included variables are listed in Chart 1.

**Student Characteristics**

**Age**

According to the results of the probit equation, the choice of public high school is significantly negatively related to the age of the student. The variable AGE spanned a range from twelve to twenty-one years, with a mean value of 17.4. This suggests that the majority of twelfth graders were in the grade proper to their age group, whereas the younger students may have been ahead of their peers and the older students behind them. The negative relationship between a student's age and the choice of public high school might reflect a trend in education where older than average students, who might have been held back earlier in their academic careers, opted out of the public school system and into the private schools to receive the educational services offered there. Private schools are often limited in the services they offer, due in part to tuition budget constraints, and must focus on the basics of education such as mathematics, reading, science, and social studies. Students who have been held back in earlier grades might benefit most from the types of programs which private schools can afford to offer, mainly those which focus on academic achievement. As we shall see in Chapters 6 and 8, academic track is one characteristic of the educational environment which is most consistently associated with academic progress, the one element which private schools can afford to offer.

67
Gender

The probit regression results indicate that choice of public school showed no significant relationship to student gender. This result is consistent with the fact that a little over 47.6% of the students in the sample who chose public schooling were female. The different enrollment statistics for male and female students were not sufficient to result in any significant relationship between public high school choice and student gender.

Race

While gender was not significantly related to the choice of public school, the probit regression results indicate that Hispanic and Asian Americans tend to choose public schooling over its alternatives, and these results are both significant at the 5% level. Although the probit regression results do not provide sufficient evidence to establish firm conclusions regarding the reasons for this choice, one factor not included in this analysis which could account in part for their choice of public schools is the lack of information on alternative choices. First, Hispanic immigrants may lack sufficient information regarding the quality of public and private schooling to make an informed choice for their children's education. Second, Hispanic immigrants tend to settle and work in geographical areas occupied by other immigrants of similar ethnic background. Geographical boundaries could significantly limit student access to private high schools located in areas which are either far from home, or require the crossing of ethnic boundaries which could prove to be dangerous. Finally, the public school facilities which new immigrants might be initially exposed to could be far superior either to the schools they left behind or to the relatively Spartan facilities provided by the private schools. Lack of information concerning alternatives to public schooling, as well as the presence of geographical and ethnic boundaries, might account for the tendency of Hispanics to choose the public school system for their children.
The probit results indicate that the variable representing Asian ethnic background was significantly positively associated with the choice of public schools. Asian Americans comprise about 2.1% of the nation's population and represent the fastest growing minority in the United States.\(^{71}\) Approximately two-thirds of the American Asian population is composed of immigrants, a substantial majority of whom arrived with professional skills and or high levels of education. As for the American born Asians (second generation immigrants), men of Chinese and Japanese ancestry had mean levels of college attainment twice the national average of 1.7 years in 1980, as well as mean earnings 15% and 13% above the US average.\(^ {72}\) Although students of Asian ancestry might have an economic advantage over students of other minority status, their social religious background is not Judeo-Christian. The majority of the private schools in this sample are associated with the Christian religion. Catholic grade schools, for example, recruit primarily from the Catholic churches to which they are attached or associated, while Catholic high schools recruit from the Catholic grade schools located in their districts. Since the majority of Asian immigrants are not of the Judeo-Christian heritage, then it is unlikely they would choose to attend, or be recruited by, the local Catholic or Christian high school. This might account, in part, for their choice of public schooling.

The probit results indicate that the public school system educates students from a broad spectrum of religious backgrounds including Catholic, Protestant, Jewish, Eastern, and no religious background in particular. Of the five basic categories representing different religious backgrounds included in the survey, no one category was significantly related to the choice of public school. Once again, due to the lack of any significant relation to public school choice, the variables repre-


\(^{72}\) Farley, R. (1990). \textit{Blacks, Hispanics and White Ethnic Groups: are Blacks Uniquely Disadvantaged?} \textit{American Economic Review Papers and Proceedings, Vol. 80, p. 239 (Table 2).}
senting different religious backgrounds were removed from the probit regression analysis through the stepwise deletion process described above.

**Student Ability and College Aspirations**

Student ability and eighth grade college aspirations may have some direct impact on the decision to attend college and graduate with a four-year diploma. These characteristics would indirectly influence a student’s decision to choose the type of high school deemed most beneficial to achieving that end. In this probit analysis, public school choice is regressed on both; the student’s intention to attend college and the students’ eighth grade test composites which serve as a proxy to student scholastic ability. The results indicate that student eighth grade score composites and the intention to attend college were both significantly positively related to the choice of public high school. The latter variable, expressing the student’s future educational aspirations, was divided into five broad categories including the student’s intention to graduate from high school, attend several years of college, graduate from college, attend post-graduate school, or attend a post-high school vocational school. Of these five categories, two were significantly positively related to the choice of public high school-the student’s intention to attend college for some years, and the intention to graduate college.

**Grade School & Teacher Background Characteristics**

In order to test for the significance of school background variables on the choice of high school attendance, public school choice was regressed on a total of nine variables describing various characteristics of the students’ grade school. These characteristics included the type of school, designated as private, Catholic, and public, the location of the school designated as rural, suburb, and urban, the total eighth grade student enrollment, and the percent of the student’s eighth grade class composed of minority students. As was expected, the results of the probit regression indicated
that the choice of public high school was significantly related to the type and location of the student’s grade school.

School Type

The first category, which depicts the type of private grade school attended by the respondent, was divided into three basic classifications, Catholic, private non-Catholic, and NAIS (National Association of Independent Schools). The first classification refers to grade schools associated with the Catholic religion and which are usually attached to and financed by the local parish church. The second classification, private non-Catholic, refers to grade schools usually associated with churches of a non-Catholic denomination. The third classification, NAIS, refers to private grade schools (or grade school and high school combinations) which are not attached to any formal religious denomination. They typically set very high academic standards and charge higher tuition rates than many other private schools since they depend primarily on tuition and fundraising income to support their operating expenses. Of the three grade school categories, this one is considered to be the most prestigious.

The results of earlier probit regressions indicated that all three were significantly negatively related to public school choice. The size of the regression coefficients indicate that students who graduated from private non-religious and NAIS grade schools are less likely to choose the public school option than those who graduated from Catholic grade schools. One plausible explanation for these results relates to the costs of private education and family income. According to the available information from the data set regarding grade school tuition and family income, the average tuition costs of Catholic, private, and NAIS schools are $1,188, $2,510, and $6,140 respectively. The average family incomes of the children who attended these schools are $35,000, $55,000, and $85,000 respectively. Furthermore, the variable representing family socioeconomic status of the
students in the sample is significantly negatively related to the choice of public high school. The probit results seem to reflect the trend that families of higher socioeconomic status, indicated by their choice of private grade schools, tend to opt out of the public school system and into the private sector.

*Eighth Grade Enrollment*

Public school choice was regressed on two more grade school characteristics-total eighth grade enrollment and base year teacher salaries. As was expected, the size of the students’ eighth grade enrollment was not significantly related to the choice of public high school. This probably reflects the fact that public high schools must accept all students who apply, whether they come from schools with small or large class sizes.

*Grade School Teacher Salaries*

The more interesting result is that base salaries for grade school teachers are significantly negatively related to the choice of public high school. This result is not surprising given the fact that teacher salaries are determined in large part by the local property tax revenues, which in turn are a function of family income. Low income families are more likely to live in neighborhoods with relatively low property taxes which constrain local public school budgets and teacher salary increases over the years. Taxpayers who are not willing or able to provide average salaries for their local public grade school teachers might also be unable to choose the private school option.

*Eighth Grade Peer Group Characteristics*

The effect of peer groups on school choice is here explicitly analyzed. Two variables were included in the probit regression in order to test for the significance of student peer groups on the choice of public schools. The first variable represented the proportion of the student’s eighth grade class composed of minorities. The second variable indicated whether the student was enrolled in
grade school classes for the gifted. Both variables were not significantly related to public school choice and were removed from the probit regression. This might reflect the fact that the majority of students who attend grade school in a certain school district remain in that district until graduation from high school. The ethnic mix and relative ability of student peer groups in a particular public school district might change very little during a student’s tenure in the system, thus making peer group effects difficult to measure.

**Family Background Characteristics**

*Single-Adult Households*

In order to test for the significance of parental marital status in the choice of public high school, the public choice variable was regressed on five variables describing the marital status of the students’ parents including whether or not they were divorced, widowed, separated, never married, or living in a marriage-like relationship. Earlier probit regression results indicated that each of these variables yielded coefficients of similar size and significance. Therefore, the variables representing the parents’ marital status as being divorced, separated, widowed, and never married were combined to form a new category representing a single-adult household. Furthermore, the variables describing the parents’ marital status, either as being married or living in a married-like relationship, were also combined to form a new variable representing a two-adult household. The choice of public high school was then regressed on the single-adult household variable and the other student background characteristics.

The probit regression results indicated that the coefficient of the single-adult household variable was significantly negatively related to the choice of public high school. This outcome marks a tendency for single-parent households to choose available alternatives to the public school system. One possible explanation for this phenomena is that private schools have the flexibility to
adjust their tuition rates on a case by case basis so to cover the costs of added student services required by single-parents with limited time constraints. This flexibility is not always present in public schools where policies are formulated based on the availability of funds determined by the taxpayers in that school district. Furthermore, the allocation of available resources is based on the needs of the average student and the special needs of exceptional students, rather than on the needs of an ever growing number of single-adult households. The lack of services designed to meet the special needs of single-adult households might be one factor contributing to their choice of alternatives to the public school system.

*Parents Expectations for their Children*

The future expectations of parents regarding their children’s education is represented by two variables describing the expectations of the student’s father and mother respectively. Earlier probit regression results indicated that neither expectations of the father or mother were significantly related to the choice of public high school. Because of their insignificant relationship to the choice of public schooling, they were not included in the final probit regression.

*Private High School Tuition*

The tuition costs of private high school alternatives to the public school system were also entered as independent variables into the probit regression. The data set provided information on the tuition costs of three types of private high school including Catholic, religious non-Catholic, and private non-religious. Although the range of private high school tuition was consistent across all three types of school, $0 - $16,000 per year, the mean tuition for each type differed markedly according to the section of the country in which it was located. For ease of analysis, the location of each school was categorized by one of six regions of the country including New England, Mideast, Great Lakes, Plains, Southeast, and West/Far West. The main reason for choosing these categories
is that the National Catholic Educational Association provided statistics on the average Catholic high school tuition for each of the six regions. Average tuition costs for other religious (non-Catholic) and non-religious private high schools were calculated for each of these regions based on the information provided in the data set. These average tuition costs, based on private school type and geographical location, were used as proxies for the costs of alternatives to public high schools in each section of the country. The choice of public high school was then regressed on the average costs of each type of private high school which might be available as alternatives to the public school choice and other student background variables.

According to the results of the probit regression, there was no significant relationship between public school choice and the tuition costs of private religious high schools. The coefficients for the variables representing tuition costs of Catholic and non-Catholic high schools were negative but insignificant. The coefficient for the private non-religious high school variable was significant and positively related to the choice of public high school. One interesting result in the probit analysis is the relatively small difference in average tuition costs between the private and non-Catholic religious high schools compared to the difference between Catholic and non-religious private schools. The probit results would seem to indicate the following: although religious non-Catholic high schools have tuition costs similar to those of the non-religious private schools, the financial burden they impose have no direct significant relationship to the choice of school. It is plausible that both types of religious high school offer student services, not available in the public sector schools, which parents are willing to pay for. The freedom to teach Judeo-Christian values might be one factor in the curriculum which motivates parents to choose them over the other non-religious schools.
CHAPTER VI

PUBLIC HIGH SCHOOL STUDENT GAIN SCORE REGRESSION RESULTS

Selectivity Bias

The second stage of this statistical model involves a value-added approach to student academic achievement in the areas of mathematics, reading, science, and history. Ordinary Least Squares (OLS) analysis is applied to estimate the relationship between public school choice and student gain scores in each of the four subjects listed above, while controlling for a set of student personal, family, peer, and school background variables. Calculated student gain scores in mathematics, reading, science, and history for each student between the eighth and twelfth grades will be regressed separately on a list of student background variables.

Educational Production Function For Public High School Students, Mathematics Gain Score.

(6.1) \[ \Delta M_i = \alpha + \beta X_i + \omega P_i^* + \mu_i \]

\( \Delta M_i \) = Calculated student mathematics gain score between the 8th and 12th grades.
\( \alpha, \beta, \omega \) = Regression coefficients.
\( X_i \) = Vector of student background characteristics except for the public school choice.
\( P_i^* \) = The Mills Ratio for public school choice.
\( \mu_i \) = residual

It is well known that unless all relevant factors are included in the regression equations the unique relationship between each of the factors and the student gain scores will be biased. In the case of estimating the relationship between public school choice and student achievement gains, the regression analysis will attribute the effects of any significant background variables related to both public school choice and student achievement to public school attendance. In order to control
CHART 2: SPECIFICATION OF PUBLIC HIGH SCHOOL STUDENT GAIN SCORE REGRESSIONS ESTIMATED

I. Dependent Variables:

Change in Standardized Mathematics Achievement Test Score
Change in Standardized Reading Achievement Test Score
Change in Standardized Science Achievement Test Score
Change in standardized History Achievement Test Score

II. Independent Variables:

AGE = student’s age in years
FEMALE (1 = yes)
BLACK (1 = yes)
HISP = Hispanic (1 = yes)
ASIAN = Asian Pacific Islander (1 = yes)
NA = Native American (1 = yes)
P8GRADES = student grade composite in eighth grade
PNOHSCHL = student does not expect to finish high school
PCOLL = student expects to attend some college classes but not graduate from college
PFINCOLL = student expects to graduate from college
PCATH = student’s religion is Catholic
POTHER = student’s religion is other than Jewish, Protestant, or Catholic.
PH2HMK = hours per week student spends doing homework in high school
P8ALGEBR = student took Algebra I in eighth grade
S8CATH = student attended a Catholic grade school
S8PRIVAT = student attended a non-religious private grade school
P8COMPUT = student had a computer in eighth grade
S8ENROL = eighth grade enrollment
P8GIFTED = student enrolled in eighth grade track for gifted students
S8MINOR = percent minority in student’s eighth grade
SH2ACADM = student enrolled in high school academic track
SH2VOTEC = student enrolled in high school vocational track
SH2OTHER = student enrolled in high school track other than general, vocational, or general
SH2SUBRB = high school located in suburb
S8PROB = index of problems in grade school such as theft, gangs, teacher abuse, etc.
SH2PROB = index of problems in high school such as theft, gangs, teacher abuse, etc.
GRTECH = percentage of high school’s graduates enrolled in college technical school
GRFORCOL = percentage of high school’s graduates enrolled in a four year college
GROTHER = percentage of high school’s graduates enrolled a program other than college
or vocational school.
SHMINOR = percentage of minorities in high school
SHGEN = percent of high school students enrolled in general program
SHPREP = percent of high school students enrolled in college preparatory program
SHSPEC = percent of high school students enrolled in special program
SINGLE = student’s parents are widowed, separated, divorced, or never married
Chart 2 (continued)

PAEXPCT = student father's educational expectations (0 = no high school 6 = post graduate)
MAEXPCT = student mother's educational expectations
FAMSIZ = student's family size (2 - 10)
SESEQ = socioeconomic status of student (quartiles 1-4)
SHRATIO = high school student/teacher ratio
SHSALARY = base year salary for high school teacher in student's high school
S8SALARY = base year salary for grade school teachers in student's grade school
CATCOST = average tuition of Catholic high school alternative to the local public high school
RELCOST = average tuition of religious high school alternative to the local public high school
PRIVCOST = average tuition of private high school alternative to the local public high school
MILLS = Mills ratio to correct for selectivity bias

for the possible omission of significant student characteristics, or other unmeasurable related to school choice and ability, the Mills Ratio (P+) has been constructed using the results from the probit equation in the first part of the model and inserted into the equations in the second part of the model as one of the independent variables. The size and significance of the coefficient of the Mills Ratio will indicate, and correct for, any bias in the regression statistics resulting from the omission of important student attributes. Calculated student gain scores for each of the four subjects were regressed on the Mills Ratio as well as a set of independent variables representing student, family, school, and peer characteristics. Chart 2 summarizes and defines the variables used in the student gain score regressions whereas Table 2 provides the regression results for the student gain scores in Mathematics between the eighth and twelfth grade, their standard errors, and t-ratios.

Large national surveys, similar to the one used for this study, are often plagued with the problem of missing data for several of its items or variables. Due to the missing data in the current survey (NELS88), the same listwise deletion process used in the probit regression was applied to the student achievement regressions for all four subjects, limiting the analysis to those cases with
Table 2: Public High School Student Mathematics Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE **</td>
<td>-133.613</td>
<td>20.160</td>
<td>-6.595</td>
</tr>
<tr>
<td>FEMALE **</td>
<td>-172.729</td>
<td>21.752</td>
<td>-7.941</td>
</tr>
<tr>
<td>BLACK **</td>
<td>-174.475</td>
<td>46.676</td>
<td>-3.738</td>
</tr>
<tr>
<td>NA **</td>
<td>-332.563</td>
<td>134.587</td>
<td>-2.471</td>
</tr>
<tr>
<td>P8GRADES **</td>
<td>11.254</td>
<td>1.881</td>
<td>5.981</td>
</tr>
<tr>
<td>P8MTTRACK</td>
<td>-74.558</td>
<td>24.467</td>
<td>-3.047</td>
</tr>
<tr>
<td>PNOHSCHL</td>
<td>-209.889</td>
<td>154.104</td>
<td>-1.362</td>
</tr>
<tr>
<td>PFINCOLL **</td>
<td>90.657</td>
<td>30.492</td>
<td>2.973</td>
</tr>
<tr>
<td>PCOLPLUS</td>
<td>53.636</td>
<td>40.030</td>
<td>1.340</td>
</tr>
<tr>
<td>S8NAIS</td>
<td>-215.414</td>
<td>159.266</td>
<td>-1.353</td>
</tr>
<tr>
<td>PCATH **</td>
<td>60.163</td>
<td>26.018</td>
<td>2.312</td>
</tr>
<tr>
<td>P8ALGEBR **</td>
<td>66.256</td>
<td>22.487</td>
<td>2.946</td>
</tr>
<tr>
<td>S8ENROL</td>
<td>8.933</td>
<td>10.612</td>
<td>0.842</td>
</tr>
<tr>
<td>SHENROL</td>
<td>0.032</td>
<td>0.024</td>
<td>1.300</td>
</tr>
<tr>
<td>PH2HMWK **</td>
<td>36.463</td>
<td>5.762</td>
<td>6.329</td>
</tr>
<tr>
<td>SH2ACADM **</td>
<td>161.571</td>
<td>25.785</td>
<td>6.266</td>
</tr>
<tr>
<td>SH2VOTEC *</td>
<td>-69.111</td>
<td>35.511</td>
<td>-1.946</td>
</tr>
<tr>
<td>SH2SUBRB **</td>
<td>84.339</td>
<td>35.556</td>
<td>2.372</td>
</tr>
<tr>
<td>SHMINOR</td>
<td>0.624</td>
<td>0.563</td>
<td>1.108</td>
</tr>
<tr>
<td>GRTECH **</td>
<td>38.059</td>
<td>18.402</td>
<td>2.068</td>
</tr>
<tr>
<td>GRFORCOL *</td>
<td>26.846</td>
<td>13.887</td>
<td>1.931</td>
</tr>
<tr>
<td>SHGEN **</td>
<td>2.173</td>
<td>0.597</td>
<td>3.643</td>
</tr>
<tr>
<td>SHPREP **</td>
<td>2.235</td>
<td>0.711</td>
<td>3.140</td>
</tr>
<tr>
<td>SHSPEC **</td>
<td>11.879</td>
<td>2.594</td>
<td>4.579</td>
</tr>
<tr>
<td>PAEXPCT</td>
<td>25.661</td>
<td>18.819</td>
<td>1.364</td>
</tr>
<tr>
<td>MAEXPCT</td>
<td>21.155</td>
<td>19.066</td>
<td>1.110</td>
</tr>
<tr>
<td>FAMSIZ</td>
<td>9.450</td>
<td>8.853</td>
<td>1.067</td>
</tr>
<tr>
<td>CATCOST</td>
<td>18.990</td>
<td>32.759</td>
<td>0.580</td>
</tr>
<tr>
<td>RELCOST</td>
<td>7.074</td>
<td>4.424</td>
<td>1.599</td>
</tr>
<tr>
<td>PRIVCOST</td>
<td>6.362</td>
<td>8.061</td>
<td>0.789</td>
</tr>
<tr>
<td>SHSALARY</td>
<td>-0.001</td>
<td>0.003</td>
<td>-0.378</td>
</tr>
<tr>
<td>SHRATIO *</td>
<td>-15.555</td>
<td>8.614</td>
<td>-1.806</td>
</tr>
<tr>
<td>INTERCEPT **</td>
<td>2262.094</td>
<td>395.971</td>
<td>5.713</td>
</tr>
</tbody>
</table>

(Continued)

Degrees of Freedom: 36
Number of observations used in calculations: 4,698
F Value: 20.595
* significant at the 10% level of significance
** significant at the 5% level of significance
Table 2: (continued)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>SESQ</td>
<td>-3.246</td>
<td>11.713</td>
<td>-0.277</td>
</tr>
<tr>
<td>SINGLE</td>
<td>5.647</td>
<td>34.678</td>
<td>0.163</td>
</tr>
<tr>
<td>MILLS *</td>
<td>108.918</td>
<td>64.120</td>
<td>1.699</td>
</tr>
<tr>
<td>INTERCEPT **</td>
<td>2262.094</td>
<td>395.971</td>
<td>5.713</td>
</tr>
</tbody>
</table>

Degrees of Freedom: 36
Number of observations used in calculations: 4,698
F Value: 20.595
* significant at the 10% level of significance
** significant at the 5% level of significance

complete data on all items. The result is that all statistics for each gain score regression are based on the same body of data.

Change in Public High School Student Mathematics Test Scores

Student Characteristics

Age

As was expected, the age of the student is significantly negatively related with student gain scores in mathematics. As was reported in chapter 5, the average age of the twelfth grade student was 17.4 years, but its range was from 12 to 21 years old. The younger students in the twelfth grade would be there due to their above-average ability, whereas the older students could be there because they were held back earlier in their academic careers. If student age serves as a proxy for ability, the negative relationship between it and student gain scores would not be surprising.

Gender

One of the more interesting results of the first regression equation is the fact that the change in mathematics scores was significantly negatively related to the female gender. Although the size of the coefficient was in the upper range of the regression results, its t-ratio was the largest of all
with a value of -7.941. The statistics imply that female gain scores in mathematics between the eighth and twelfth grades were 7.6 percentage points below the average for the sample. This result is consistent with previous studies concerning students scores in junior high school mathematics and high school mathematics classes.  

Despite extensive research on gender differences in home and classroom interactions, little is known about the specific factors which might account for the differences in mathematics achievement scores. Socialization in the home is one factor often suggested as preventing girls from developing the characteristics associated with scientific or mathematical studies, such characteristics as independence, convergent thinking, logic, and experimentation. Girls are often socialized into characteristics such as dependence, nurturance, and passivity. It has also been suggested that observed gender differences in mathematics achievement are the result of intrinsic student characteristics relating to personal competitiveness and self-confidence nurtured in the home long before students enroll in grade school. Proponents of this position hold that students are a reflection of the values of our society and might enter schools with preset stereotypes already built into their personalities. Therefore, many of the differences observed in male and female classroom behavior leading to differences in participation may be the result of gender-role socialization in the home.

The school as well as the home has also been seen as contributing to the existing gender differences in achievement and course enrollment by means of student teacher interactions within the classroom. In the field of secondary mathematics, several studies have indicated that females

have fewer interactions with teachers than males.\textsuperscript{76} For instance, in geometry classes females were called on less frequently and were asked questions at a lower cognitive level than those directed at males. Other studies found that teacher expectations for success also affected student interactions in mathematics classes.\textsuperscript{77} While males received more feedback than females and were expected to be more successful, teachers provided more feedback to females from whom they expected less than to females they expected to be successful. Different educational experiences based on gender might be another one of the factors responsible for the observed lower gain scores in mathematics for females.

\textit{Race}

Many studies indicate that trends in achievement scores have varied among different ethnic groups. The gap in average scores between black and nonminority students has been and still remains large. While Hispanic students appear to have gained relative to nonminority students, the data pertaining to Hispanic students are less clear cut. The mathematics regression results indicate a similar pattern for the minority students in the present survey. While mathematics gain scores of Hispanic students did not differ significantly from the average of the students in the survey, African and Native American student gain scores were significantly negatively related with their ethnic background. These ethnic backgrounds were associated with mathematics gain scores 7.7 and 14.9 percentage points below the average for the sample, respectively.

Since minority school age children are more than twice as likely to live in a single-parent home, the lower than average mathematics achievement scores of African and Native Americans have often been attributed in part to living in single-parent households. A number of cross sectional

\textsuperscript{76} Brophy, J. (1985). "Interactions of Male and Female Students With Male and Female Teachers." In L. Wilkinson and C. Marrett (Eds.). \textit{Gender Influences in classroom Interactions} (pp. 115-142). Orlando, FL: Academic Press.
studies have found that children from single-parent households have lower than average scores on a number of measures of intellectual development and achievement, including IQ tests, standardized achievement tests, and school grades. However a recent nationally representative study found that while the scores of elementary school children from two-parent households exceeded those from single-parent homes, the corresponding differences among high school students were found to be negligible. The regression results for all four subject matters in the present survey indicate that the coefficient of the variable representing single-parent households is not significantly related to student gain scores between the eighth and twelfth grades. The regression outcomes do not provide sufficient evidence to support the hypothesis that the lower than average mathematics gain scores for African and Native American students in the public school system are the result of living in single-adult homes.

The lower than average achievement scores have also been attributed to the low socioeconomic status of many African and Native American students who are more likely to live in low-income neighborhoods. In order to test for the existence of a significant relationship between socioeconomic status and student gain scores, an index variable was created using available data on family income and the level of parental education. When student gain scores for mathematics were separately regressed on the student’s socioeconomic status as well as other background variables, no significant relationship was found. This result implies that family income and parental education

are not in themselves sufficient characteristics to account for the lower gain scores of African and Native American students in mathematics.

There is, however, another explanation which could account for the lower mathematics gain scores of Native and African American students. It involves the possibility of involuntary segregation of minorities into low income ghettos. Two recent studies in particular have tested for the presence of discrimination in the housing market. In the first study the behavior of lenders for home loans is modeled as a function of certain explanatory variables measuring the "riskiness" of the borrower. Once risk is accounted for, if lenders are fair, then the race or ethnicity of the borrower should have no relationship to the granting of a loan. However, these studies tend to find that race is important and that blacks receive fewer loan approvals than whites, even after the riskiness of the borrower and the location of the house are controlled.\(^8^0\)

A second type of study was conducted to test for the presence of housing discrimination against minorities in which black and white actors were given identical characteristics and then sent to real estate agents to inquire about home purchases. Significant discrimination against minorities was observed.\(^8^1\) For instance, real estate agents consistently told blacks about fewer homes in white areas and showed them fewer, thus engaging in a behavior called "steering". Although illegal, this type of behavior on the part of real estate agents could result in the persistence of segregated neighborhoods. If ethnic minorities are regularly segregated into neighborhoods with low housing values, the property tax revenues used to support local public schools may not be sufficient to provide the quality of education available to nonminority students in high income suburbs. This may


be one explanation for the lower than average mathematics gain scores of African and Native American students.

Religion

One of the more surprising results of this regression analysis is that the coefficient of the Catholic religion variable was the only one of the five religious categories significantly positively related to student gain scores in mathematics, science, and social studies. As was mentioned previously, student religious background was divided into five broad categories including Protestant, Catholic, Jewish, Other, and no religious background. The mathematics gain score regression results indicate that Catholic students tend to outperform students of other religious backgrounds in the public school system by an average of 3.1 percentage points. Earlier regressions tested for the relationship between student religious devotion and academic achievement only to find out that the two were negatively related, but that relationship was not significant.

One plausible explanation for the apparent success of Catholic students is that they might be the children of ambitious upwardly mobile minority group families. The variable for socioeconomic status of the students in the sample was not significantly related with student achievement in mathematics. This might indicate that parental education and family income have little if any impact on student gain scores. However, a recent study concerned with the relative earnings of various ethnic groups in the United States indicated the following: if the observed productive characteristics of each group were equalized, men of Austrian, German, Hungarian, Irish, Italian, and Russian ancestry would have earned significantly more than the average for the sample.82 Since a significant portion of these ethnic groups are traditionally Catholic (a variable including the Russian Orthodox religion) the regression results might be providing evidence of important intergen-

---

erational transfers of ethnic human capital manifested in the rates of return to education. Evidence of this has been reported in two recent studies on the productive characteristics of various ethnic American groups.\textsuperscript{83} If the variable representing Catholic religion serves as a proxy for a student's productive ethnic background, then the above average student gain score in mathematics might be a manifestation of intergenerational transfers of ethnic human capital.

\textit{Student Eighth Grade College Aspirations}

Student ability and college aspirations may have a direct impact on the decision to attend college and finish with a four-year diploma. Students' eighth grade college aspirations were significantly related to their gain scores in mathematics. Future academic aspirations of eighth grade students were divided into six categories ranging from the expectation to attend high school for a few years to the hopes of completing post graduate college studies. Of these six categories, only one was significantly related to academic achievement in mathematics, the intention to graduate from a four year college. It was not surprising to find that students who did not intend to graduate high school had mathematics gain scores 13.2 percentage points lower than the average for the sample, whereas students who intended to graduate from a four-year college received gain scores 2.8 percentage points above the average for the sample.

It has often been suggested that since educated and affluent families are more determined that their children should have a college education, they provide them with the physical apparatus which facilitates academic success. However, when the socioeconomic status of the family is controlled, a variable representing both family income and parental education, its coefficient is reduced to insignificance in the mathematics regression results. At the same time, the coefficient for the

variable representing student academic aspirations remains significant. A plausible explanation for these results may be that students who are confident of their ability to graduate from a four-year college have planned to attend college since the eighth grade and make more effective intellectual use of their time in high school. A student’s self-confidence may be the result of parental encouragement in the home combined with consistent positive feedback and positive reinforcement from the school in the form of good grades for academic performance. Both of these variables were included in the regression analysis and both were significantly positively related with academic achievement in mathematics.

*Eighth Grade Score Composites*

Since schooling is one of the ways students receive information regarding their ability for academic achievement, eighth grade test composites would not only serve as a proxy for ability but could also impact the decision to pursue higher academic degrees. It is plausible that students form their future expectations early in their academic careers based on information concerning their natural abilities in the form of grades. Students who receive consistently low grades in elementary school would conclude that, in the long run, formal schooling is not in their best interests. The opposite would be true for those with superior academic achievement. In order to test the validity of this hypothesis, mathematics gain scores were regressed on a variable which combined all subject scores into one composite grade point average for each of the students when they were in the eighth grade. The regression resulted in a significant positive relation between the student’s eighth grade composite and the high school mathematics gain score. This grade school composite would serve as a proxy for feedback information concerning the student’s ability for academic success, as well as encouragement to make the best academic use of their time in high school to help achieve their academic goals.
Algebra I in Eighth Grade

Similar to the regression results for student eighth grade composite scores, it was found that students who took Algebra I in the eighth grade scored significantly higher in their mathematics gain scores than those who did not. The magnitude of the coefficient indicated a 4.0 percent increase in mathematics gain scores above the average for the sample. At first glance, one might conclude that the strong relationship between the two measures indicates the effectiveness of accelerated grade school mathematics programs on student high school achievement levels. However, the data set provides additional information regarding the enrollment of grade school students in special tracks for the gifted, tracks which offered specific courses in enriched mathematics, reading, science, and social studies. Surprisingly, when student gain scores in mathematics were regressed on the variable representing student enrollment in the enriched mathematics program, a significant negative relationship emerged. The size of the coefficient indicated that students enrolled in this type of program realized a gain score 5.8 percentage points lower than the average for the sample.

There are two plausible explanations which could shed light on these seemingly contradictory results. First, the strong positive relationship between student mathematics gain scores and the introduction to Algebra I in the eighth grade indicates that general enriched mathematics courses may be no substitute for a basic class in high school Algebra. Second, eighth grade students who enrolled in accelerated mathematics courses scored well above average in the first round of standardized tests but gained very little during their high school career compared to the other students in the sample. If this is the case, then the regression results would imply that high schools receiving students of above average ability in mathematics may be contributing very little to their performance. The present system of public schools might be adding more to the mathematics skills of the
average student than to the above average student who is not sufficiently challenged to achieve at full potential.

School Characteristics

Most of the studies outlined in the Literature Review focused on the relationship between high school background variables and student achievement at the high school level. Although most of the eighth grade school characteristics included in earlier regressions turned out to be insignificantly related to student gain scores in mathematics, their inclusion here may shed light on future policy issues regarding public grade schools.

Grade School Location, Type, Student/Teacher Ratio

The location of the grade school, whether it be in an urban, suburban, or rural area, was not significantly related to student gain scores in mathematics. Similarly, the type of grade school, whether it be public, Catholic, or private, was not significantly related to later academic achievement gains in mathematics. Finally, the student/teacher ratio of the different public grade schools was not significantly related to student achievement in mathematics.

High School Location

Although most of the grade school characteristics included in the regression analysis were not significantly related to later academic achievement in mathematics, many of the high school characteristics were. Beginning with the location of the public high school, the regression results indicated that schools located in suburban areas were significantly positively related to student gain scores in mathematics. Since urban schools were included in the constant term of the regression, these results indicate that suburban and rural high school students outperform them by an average of 3.6 percentage points in their mathematics gain scores. A potential explanation might be related to the availability of larger school budgets in the suburban areas, due to larger property tax reve-
nues, compared to the corresponding urban areas. Although the data set does not provide information on annual property taxes paid, there is information on annual family income. The parents of students attending suburban public high schools report annual incomes which are approximately 35% higher on average than those whose children attend urban public high schools. Parents who choose to locate their households in high tax suburban communities may demand from their children a greater return on their investments to education and provide incentives to their children to achieve academically.

Higher property tax revenues from more affluent families located in suburban and rural areas also enjoy police protection and less violent neighborhoods compared to the urban areas of our inner cities. The regression results indicate that the presence of problems in the school is significantly negatively related to student gain scores in mathematics. The variable representing school problems is an index of local neighborhood and school characteristics, including the presence of gangs and student violence reported by the students. According to the reports of violence made by the public school students in this sample, the experience of school problems was 10% greater in the urban public high schools than in the suburban public high schools. A recent study reported that the formation of gangs often results from an individual's fear of personal violence or victimization. A young boy would then join a gang in which all members would feel reasonably safe within defined geographical limits. This group of friends would occasionally fight with other groups, and any member who would refuse to join the fight would be expelled. An interesting point of this study is that over 50% of the inner city students surveyed joined gangs whose geographical limits included the streets used to go and come from their neighborhood high schools. Youths who are members of these gangs are less apt to fear the school enterprise, the immediate neighborhood,

---

and every other social setting, whereas the opposite might be true for non-members. The presence of gangs, along with the concomitant possibility of violence, would have a negative effect on student achievement in the urban school setting where such violence is more prevalent than in the suburbs.

*Tracking*

Of all the variables related to academic achievement in mathematics, the variable with the largest positive coefficient was student enrollment in an academic track. The size of the coefficient for academic track indicates that students in this category produced a gain score in mathematics over 6.9 percentage points above the average for those enrolled in the general school track. This came as no surprise since the majority of studies concerning the relationship between student achievement and grouping according to ability or motivation had similar results. The regression results for students enrolled in the vocational high school track indicated that this group scored below the average mathematics gain score for those in the general track by 3.2 percentage points. It is plausible that students enrolled in a vocational track would not be exposed to the higher order mathematical skills studied by students in the general or academic tracks. Perhaps one way to improve the overall quality of education for female and minority students is to encourage them to enroll in the school’s academic track.

*Homework*

The total number of hours students spend, during the week and on weekends, on homework outside of school is significantly positively related to academic gains in mathematics. The variable’s t-ratio was among the largest of all others in the regression analysis. Once again this result came as no surprise since higher order mathematical skills are learned by experience and practice. For example, some of the lower level questions on the standardized mathematics tests included
simple operations with decimals, powers, and roots, whereas higher level questions required a proficiency in solving complex multi-step word problems as well as the ability to demonstrate knowledge of mathematics material found in advanced courses. Lower order skills can often be learned by rote memory at the grade school level whereas those of the higher order require substantial reasoning and reading comprehension introduced in courses offered to sophomores, juniors, and seniors. Student ability certainly plays a significant role in achievement, but the assignment of homework provides the time and practice necessary to master the higher order skills involving reading comprehension and problem solving which cannot be learned by simple rote memory. The regression results indicate that the introduction and mastery of lower level skills, through an introduction to Algebra I in the eighth grade, lays the foundation for the mastery of higher order skills which can be improved through the assignment of homework outside of class.

*High School Problems*

While academic track and hours of homework are among the variables most significantly associated with academic achievement, the school characteristic which is most negatively related to student gain scores in mathematics is the presence of social problems in the school. The variable representing school problems is an index of problems reported and experienced by the students relating to the presence of gangs, theft, drugs, alcohol, racial tensions, and the abuse of teachers, physical or otherwise. Many of these problems related to student and teacher violence are present in our urban public schools neighborhoods and may provide a plausible explanation for the lower student gain scores relative to student achievement levels in the suburban schools. The presence of violence in and around our urban schools is a serious problem, consistently associated with lower student achievement, which needs to be addressed if progress is to made in reducing truancy, drop-
outs, and enhancing students' desire to attend school. It is difficult to learn if one is afraid to go to
school, to be near school, or even to be inside a school.

**Student/Teacher Ratio**

Finally, unlike the student/teacher ratios reported by the grade schools, the regression re-
sults indicate that at the high school level, the student/teacher ratio is significantly negatively re-
lated to academic achievement in mathematics at the 10 percent level. This result is not surprising
given that smaller classes enable teachers and their assistants to provide students with the individ-
ual attention they need to master higher order skills involved in high school mathematics.

**Peer Group Characteristics.**

Up to this point, we have observed the relationship between various student and school
characteristics which might have been related to academic achievement in the public school system.
Now we turn our attention to the relationship between student achievement and various high school
peer groups.

**Minorities**

The first peer characteristic to consider is the percent of the high school student body com-
posed of minorities. While it has often been held that the presence of minorities is detrimental to
academic achievement, the regression results indicate that this is not the case. The coefficient of the
minority variable is not significantly related to student gain scores in mathematics. The point to be
stressed is that there is not sufficient evidence in this analysis to associate low academic achieve-
ment in mathematics with the presence of minorities in our public high schools.

The regression results of the present study indicate that neither socioeconomic status in it-
self nor the proportion of the student body composed of minorities is significantly negatively asso-
ciated with low student mathematics gain scores. What is found, however, is that the students' ex-
perience of violence in and around the vicinity of the school, regardless of their socioeconomic status, is negatively related to low student gain scores in mathematics. Policy makers concerned with efforts to increase the equality of educational opportunity for minorities and the quality of education which they receive need to focus less on policies involving desegregation in itself and more on eliminating the student violence permeating many of our urban schools. Furthermore, since academic track and hours of homework are positively associated with student achievement, our nation's public schools might benefit by increasing the proportion of students enrolled in academic programs as well as the amount of homework assigned.

**High School Graduates**

The second peer group category to consider is the percentage of recent high school graduates (graduated in the spring of 1991) enrolled in technical and four-year colleges. In both cases the coefficients of the variables are significantly positively related to the mathematics gain scores of students enrolled in high school. Despite the limitations of standardized tests, as well as the danger that teachers could "teach to the test", the argument that tests do not measure performance of any kind is not supported by the results of this regression analysis. Rather we find a strong positive relationship between a school's graduates placed in technical and four-year colleges and its current students' gain scores on standardized mathematics tests. It is plausible that the higher order mathematics skills, learned at the high school level and reflected in student gain scores, provide the rigorous training indispensable to the application of logic and the scientific method necessary for success in technical and other four year colleges.

**Peer Group Enrollment in Various Academic Tracks**

The third category of peer group relationships concerns the percentage of the student body enrolled in the general, college preparatory, and special education tracks. As was expected, there
was a positive and significant relation between student gain scores in mathematics and the percent of the student body enrolled in the general and academic tracks. The more interesting result is that of the three categories included in the regression, student gain scores were largest and most significantly related (t-ratio = 4.479) to student enrollment in special education tracks. Previous studies show that many disabled or non-English-speaking students receive needed special programs, which have dramatically improved their test scores in recent years. At the same time, the regression results exhibit no evidence that help for these students has had a negative effect on the academic progress of their more advantaged peers. Unfortunately, it is a well known fact that special education programs are very costly, but the mathematics gain scores for students in them surpasses even those enrolled in academic tracks. The positive effects of special education programs reported in the regression results should be given serious consideration by policy makers who might, in their efforts to control the escalating costs of secondary education, be tempted to remove them.

**Family Background Characteristics.**

**Single-Adult Homes**

The proportion of children living in single-adult households has grown markedly over the past thirty years from 9% in 1960 to approximately 25% in 1990. A number of cross-sectional studies have found that children from single-adult households have lower average scores on a number of measures of intellectual and academic achievement, including IQ tests, standardized achievement tests, and school grades. In contrast, the corresponding differences among secondary school students from single and two-adult homes were found to be negligible in a parallel study for

---


that age group.\textsuperscript{88} Therefore, it was not surprising to find no significant relationship between the two when student gain scores in mathematics were regressed on the variable representing single-adult households. In fact there was no significant relation between single-adult households and student gain scores in any of the four subjects included in this study. Although previous studies imply that the grade school children of single-adult homes experience lower achievement levels than those from two-adult households, that relationship is reduced to insignificance at the high school level.

\textit{Socioeconomic Status}

It has been a common procedure, in studies concerning student achievement, to proxy student ability, often considered to be an unmeasurable characteristic, using the student’s socioeconomic status. In the present regression analysis, the student’s socioeconomic status (measured in quartiles, i.e. 1 through 4) is not significantly related to student gain scores in mathematics and reading. This outcome is not surprising when one considers that the reason for its use in the past was primarily due to the lack of available information which could explain student achievement. The fact that this variable is reduced to insignificance may reflect the presence of variables in the regression which do account for student achievement in mathematics. For example, variables which measure the number of hours students spend doing homework each week, eighth grade college aspirations, academic track, exposure to Algebra I in the eighth grade, and peer group effects seem to be the characteristics which significantly impact student academic achievement in mathematics at the high school level. If this regression has targeted the variables which really do count, then it should not be surprising to find that socioeconomic status is no longer significantly associated with student achievement in mathematics.

Parents' Aspirations for their Children

Parents’ future academic expectations for their children have often been positively associated with student achievement. In order to test for the possibility of a significant relationship between parental expectations and academic achievement, two variables, each representing the future academic aspirations of the student’s father and mother respectively, were added as independent variables to each of the four regression equations. It was found that neither expectations were significant. The regression results indicate that the future academic expectations of the students themselves are more strongly associated with academic gain scores than are those of the parents’.

Mills Ratio

The first part of this model included a probit regression for choice of public school verses all other school types, i.e. Catholic, private, and NAIS. If there is a statistically significant self-selection into the public high school sector based on unmeasured factors, this model postulates a relation between the residual of the school sector choice equation and the residual of the gain score equation. Such a correlation seems to exist since the Mills Ratio, which corrects for self-selection, is significantly positively related to the student mathematics gain scores at the 10 percent level. The significance of the Mills Ratio is indicative of the presence of unmeasured factors which account for the selection of the public school sector and student gains in mathematics.

The presence of such unmeasured factors may be explained by the stepwise deletion process used to isolate the most significant variables in the gain score regression equation. The earliest mathematics gain score regression included 62 independent variables, most of which were not significantly related to student mathematics achievement in public high schools. At the earliest stages of the model, the Mills Ratio had a coefficient equal to 118.689 and a t-ratio of only .64 indicating that it was not a significant variable in the regression at this point. As variables were deleted from
the mathematics gain score regression, the Mills Ratio grew in significance until it reached the value found in the regression results. A plausible explanation is that when each of the 38 variables was deleted from the achievement regression, its effect or relationship to student gain scores was picked up by coefficient of the Mills Ratio variable. Although the individual effects of each independent variable was not significant, the sum total of all the effects of the 38 deleted variables resulted in a significantly negative bias to the model, reflected in and corrected by the Mills Ratio.

The fact that there were so many variables to deal with in both equations made it difficult to isolate the particular variables which might be significant in the final model. The purpose of the Mills Ratio is to correct for the presence of unmeasured characteristics deleted from the model and insure that the coefficients of the included variables remain dependable.

**Change in Public High School Student Reading Test Scores**

We now turn our attention to the student reading gain score regression results. The change in standardized reading scores, calculated from the eighth and twelfth grade standardized test results, are regressed on the relevant student background variables. A stepwise regression technique is used to delete the independent variables which proved to be insignificant in earlier reading gain score regression results. In order to control for the possible omission of significant student characteristics or other unmeasureables related to both school choice and ability, the same calculated Mills Ratio, using the results of the probit equation in the first part of the model, is inserted into the third regression equation as one of the independent variables. The size and significance of the Mills Ratio will indicate and correct for any bias in the regression statistics resulting from the omission of important student attributes. Furthermore, individual factors which were removed from the regression in the stepwise deletion process, because of their insignificant relationship to student reading gain scores, resulted once again in a significant Mills Ratio when taken cumulatively. It is
therefore included in the reading gain score regression as one of the independent variables in order to correct for both unmeasured factors relating to both choice of public school and student reading gain scores.

**Educational Production Function For Public High School Students, Reading Gain Score.**

\[(6.2) \quad \Delta R_i = \alpha + \beta X_i + \omega P_i^{+} + \mu_i\]

- \(\Delta R_i\) = Calculated student reading gain score between the 8th and 12th grades.
- \(\alpha, \beta, \omega\) = Regression coefficients.
- \(X_i\) = Vector of student background characteristics except for the public school choice.
- \(P_i^{+}\) = The Mills Ratio for public school choice.
- \(\mu_i\) = residual

The standardized reading test administered to students in the eighth and twelfth grades marked three basic levels of proficiency. The first level tested for student ability in the basic skills of reading comprehension, including the reproduction of detail and/or the author's main thought. The second level concerned the student's ability to make relatively simple inferences beyond the author's main thought as well as to understand and evaluate relatively abstract concepts. The third level tested the student's ability to make complex inferences or evaluate judgments that require piecing together multiple sources of information from the passage. As with the mathematics gain scores, reading gain scores were calculated by subtracting a student's reading score on the eighth grade test from the same student's score on the twelfth grade test. Calculated student gain scores in reading were regressed on the Mills Ratio and a set of independent variables representing student, family, school, and peer characteristics. The results of the student reading gain score regression are reported in Table 3.

99
Table 3: Public High School Student Reading Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>-63.130</td>
<td>20.022</td>
<td>-3.153</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-1.565</td>
<td>20.989</td>
<td>-0.075</td>
</tr>
<tr>
<td>BLACK</td>
<td>-134.555</td>
<td>48.194</td>
<td>-2.792</td>
</tr>
<tr>
<td>ASIAN</td>
<td>54.357</td>
<td>49.255</td>
<td>1.104</td>
</tr>
<tr>
<td>HISP</td>
<td>-74.462</td>
<td>44.323</td>
<td>-1.680</td>
</tr>
<tr>
<td>NA</td>
<td>-290.694</td>
<td>131.896</td>
<td>-2.204</td>
</tr>
<tr>
<td>P8GRADES</td>
<td>-1.271</td>
<td>-1.791</td>
<td>-0.709</td>
</tr>
<tr>
<td>P8RTRACK</td>
<td>-11.171</td>
<td>22.814</td>
<td>-0.490</td>
</tr>
<tr>
<td>PCOLL</td>
<td>-89.186</td>
<td>43.198</td>
<td>-2.065</td>
</tr>
<tr>
<td>PFINCOLL</td>
<td>-28.377</td>
<td>37.803</td>
<td>-0.751</td>
</tr>
<tr>
<td>PCOLPLUS</td>
<td>-36.692</td>
<td>46.951</td>
<td>-0.781</td>
</tr>
<tr>
<td>PCATH</td>
<td>30.222</td>
<td>25.806</td>
<td>1.171</td>
</tr>
<tr>
<td>POTHER</td>
<td>-45.095</td>
<td>43.742</td>
<td>-1.031</td>
</tr>
<tr>
<td>S8MINOR</td>
<td>0.269</td>
<td>0.566</td>
<td>0.476</td>
</tr>
<tr>
<td>S8PROB</td>
<td>2.963</td>
<td>1.167</td>
<td>2.539</td>
</tr>
<tr>
<td>S8ENROL</td>
<td>-1.633</td>
<td>10.393</td>
<td>-0.157</td>
</tr>
<tr>
<td>SHENROL</td>
<td>0.037</td>
<td>0.023</td>
<td>1.609</td>
</tr>
<tr>
<td>PH2HMWK</td>
<td>25.856</td>
<td>5.558</td>
<td>4.652</td>
</tr>
<tr>
<td>PH2TV</td>
<td>-1.554</td>
<td>4.514</td>
<td>-0.344</td>
</tr>
<tr>
<td>P8COMPUT</td>
<td>38.291</td>
<td>24.431</td>
<td>1.567</td>
</tr>
<tr>
<td>SH2ACADM</td>
<td>23.768</td>
<td>24.550</td>
<td>0.968</td>
</tr>
<tr>
<td>SH2VOTEC</td>
<td>-63.311</td>
<td>34.735</td>
<td>-1.823</td>
</tr>
<tr>
<td>SH2PROB</td>
<td>-18.041</td>
<td>3.329</td>
<td>-5.419</td>
</tr>
<tr>
<td>GRTech</td>
<td>11.581</td>
<td>17.490</td>
<td>0.662</td>
</tr>
<tr>
<td>GRFORCOL</td>
<td>3.406</td>
<td>13.344</td>
<td>0.255</td>
</tr>
<tr>
<td>SHPREP</td>
<td>0.360</td>
<td>0.540</td>
<td>0.666</td>
</tr>
<tr>
<td>SHSPEC</td>
<td>2.226</td>
<td>2.341</td>
<td>0.951</td>
</tr>
<tr>
<td>PAEXPCT</td>
<td>28.194</td>
<td>18.264</td>
<td>1.544</td>
</tr>
<tr>
<td>MAEXPCT</td>
<td>-8.123</td>
<td>18.608</td>
<td>-0.437</td>
</tr>
<tr>
<td>FAMSIZ</td>
<td>14.736</td>
<td>8.614</td>
<td>1.711</td>
</tr>
<tr>
<td>CATCOST</td>
<td>26.343</td>
<td>31.194</td>
<td>0.844</td>
</tr>
<tr>
<td>RELCOST</td>
<td>-5.007</td>
<td>4.244</td>
<td>-1.180</td>
</tr>
<tr>
<td>PRIVCOST</td>
<td>-1.876</td>
<td>7.662</td>
<td>-0.245</td>
</tr>
<tr>
<td>INTERCEPT</td>
<td>1470.627</td>
<td>384.500</td>
<td>3.825</td>
</tr>
</tbody>
</table>

(Continued)

Degrees of Freedom: 39
Number of observations used in calculations: 4,360
F Value: 4.235
* significant at the 10% level of significance
** significant at the 5% level of significance
Table 3: (Continued)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHSALARY</td>
<td>0.004</td>
<td>0.003</td>
<td>-1.232</td>
</tr>
<tr>
<td>S8SALARY **</td>
<td>40.915</td>
<td>9.988</td>
<td>4.097</td>
</tr>
<tr>
<td>SHRATIO **</td>
<td>-17.953</td>
<td>8.523</td>
<td>-2.106</td>
</tr>
<tr>
<td>SESQ</td>
<td>-16.462</td>
<td>11.518</td>
<td>-1.429</td>
</tr>
<tr>
<td>SINGLE</td>
<td>-13.964</td>
<td>33.453</td>
<td>-0.417</td>
</tr>
<tr>
<td>MILLS **</td>
<td>106.226</td>
<td>37.341</td>
<td>2.845</td>
</tr>
<tr>
<td>INTERCEPT **</td>
<td>1470.627</td>
<td>384.500</td>
<td>3.825</td>
</tr>
</tbody>
</table>

Degrees of Freedom: 39
Number of observations used in calculations: 4,360
F Value: 4.235
* significant at the 10% level of significance
** significant at the 5% level of significance

Student Characteristics

Age

As we saw previously, the student’s age is significantly negatively related to reading improvement scores between the eighth and twelfth grades. This result is similar to the findings of the National Assessment of Literacy, conducted in 1986, which revealed striking deficiencies in the ability of young adults (ages 21-25) to use written text in a variety of ways.\(^9\) Less than 40 percent of those in the sample could synthesize the main argument of a lengthy newspaper article while only 60 percent could extract information from a bar graph. Given their low performance in mathematics gain scores, it is not unexpected that some items in the literacy assessment, which entailed the use of arithmetic, also revealed serious deficiencies. One plausible explanation for this result is that student mathematics and reading gain scores are proxies for student ability. An alternative view is that early years of schooling are important determinants of achievement in later

---

grades. If certain critical reading skills are not learned during a student’s early grade school years, that person might never develop the higher order reading skills needed to read well. It is possible that both natural ability and the lack of proper instruction may be responsible for the lower gain scores in mathematics and reading associated with the older students in the sample.

Race

The gap in average reading gain scores between black and nonminority students is significant. According to the regression results, African Americans scored a full 9.1 percentage points below the nonminority average gain score in reading. Although past studies have associated lower student achievement scores with single-adult homes and low socioeconomic status, the present regression results indicate that past studies might have overstated the impact of these variables on student achievement in reading.

There are a number of other related factors which could account in part for the lower reading gain scores associated African American students. First, their mathematics gain scores were 5.9 percentage points lower than the average nonminority gain scores. The lower mathematics gain scores might be indicative of deficiencies which could have affected their understanding of certain items in the literacy tests involving the use of basic arithmetic skills, such skills as the interpretation of percentages or bar graphs.

Second, the African American students in this sample attended public grade schools and high schools with average minority enrollments of 56%, compared to the white students who attended public schools with average minority enrollments of approximately 14%. Furthermore, the majority of black students in the sample (approximately 40%) attended urban and inner city schools, often plagued with violence and constrained by low budgets due to local tax systems. The higher level of violence in our urban public schools was previously reported in the student gain

102
score results for mathematics. Furthermore, the average reported family income of black students was approximately half the reported income of white students attending public schools. If family income serves as a proxy for property taxes paid, then black students segregated into low income urban neighborhoods might find themselves attending public high schools with inadequate facilities and limited police protection. Limited funding, rather than ethnic mix, might be a factor contributing to the lower reading gain scores for black students in the public schools.

Finally, an issue which has received little attention in the Economics of Education literature is the use of “Black English”, an idiom spoken by many inner city urban blacks but not by blacks brought up in middle or upper class families except as an affectation. The extensive or even exclusive use of Black English is detrimental to student reading gain scores insofar as it impairs a person’s ability to communicate with the larger world of business, science, and government, or to master the national language, the principal subjects of interest on standardized reading tests. The confluence of lower mathematics improvement scores, urban violence, and local idioms could substantially contribute to the lower reading gain scores associated with African American students in the sample.

The regression results indicate that the coefficient for the variable representing Native American ethnic background was significantly negatively related to student reading gain scores. The size of the coefficient implies that Native American reading gain scores were 16.6 percentage points below the average for the sample. There is one characteristic common to older students as well as African, and Native Americans which might partially account for their lower than average reading gain scores, and that is their lower than average achievement in mathematics. The Native Americans scored 22.6 percentage points lower than the average for the sample and this result was

---

significant at the 5% level. Once again, this is indicative of serious deficiencies in basic mathematics skills which might have had a negative effect on Native American students' ability to interpret the meaning of items on the reading test requiring some proficiency in lower order mathematics skills. Plausible explanations for the lower mathematics gain scores of Native American students lie beyond the scope of this study, but policy makers concerned with equality in education issues might be well advised to study the problem in order to identify and attempt to remedy its causes.

An interesting result of the regression analysis concerns the variable representing the Asian American ethnic background, which was significantly positively associated with reading gain scores between the eighth and twelfth grades. The size of the coefficient shows a 5.2 percentage point gain over the average gain score for the sample. This result was surprising, given the outcomes of a recent national study describing the superior performance of Asian American students on the mathematics section of the SAT test as well as their below average performance on the corresponding verbal section. Their higher performance on the mathematics section is most often attributed to the social pressure to succeed academically, whereas their lower performance on the verbal section is attributed to their limited English-speaking abilities, which often impede success.

The regression results for Asian students attending public high schools in this sample do not support the outcomes of the national survey mentioned above. First, the Asian students' reading gain scores as well as their eighth and twelfth grade reading scores were higher than those of all other ethnic groups, including whites. Second, their eighth grade score composites were higher than any other ethnic group in the sample. Third, they attended public schools with student minority enrollments of thirty-five to forty percent on average. One student characteristic which might account for their superior achievement levels in mathematics and reading is that they reported spending more time doing homework than any other ethnic group. On average they reported spending
25% more time during the week and on weekends doing homework than did students of other ethnic backgrounds. In order to test for other significant student characteristics which might account for their superior achievement in mathematics and reading, the variable representing the Asian ethnic background was regressed on a set of forty-two student background characteristics. Due to the discrete nature of the choice process, a probit regression was used. The results of the probit regression indicated that the Asian variable was significantly positively related to eighth grade score composites, hours spent doing homework, eighth grade enrollment in Algebra I, and strong paternal expectations. They were the only ethnic group in the sample of public school students with this particular set of characteristics. It is plausible that these are the important factors, combined with the social pressure (from the family or community) exerted on them to succeed academically, which account for the above average gains in reading skills and mathematics achieved between the eighth and twelfth grades.

The reading gain score regression results indicate that the variables representing mainstream American religious groups including the Protestant, Catholic, and Jewish sects were not significantly related to the student reading gain scores in this sample. On the other hand, the variable representing religious affiliations outside the American mainstream such as Muslim, Eastern, and other non-Christian religions was significantly negatively related with the student reading gain scores. The size of the coefficient indicates that students within this group scored 4.7 percentage points below the average for the sample. The regression results show no significant difference between this religious characteristic and student gain scores for mathematics, science, or social studies. Because this result is found only in the reading scores regression, it is probable that cultural or linguistic factors may be responsible for their lower than average reading gain scores.

School Characteristics

*Grade School Teacher Salaries*

Of all the eighth grade school characteristic variables initially included in the reading gain score regression, one of the more interesting results concerns grade school teacher salaries. The regression results indicate that the base year salary for grade school teachers with a bachelor's degree is significantly positively related to student reading gain scores. The logic behind this result is that schools which offer relatively high base salaries attract a large pool of applicants from which it could hire the most experienced and qualified teachers to instruct students in basic reading skills. Higher base salaries also serve as incentives to retain the high quality productive teachers which they attract. The strong relation between grade school salaries and high school student reading gain scores may be indicative of the higher quality of education offered by highly paid teachers at the grade school level.

The issue of teacher salaries has been very controversial in recent decades given the steady decline in Scholastic Aptitude Test (SAT) scores and the corresponding drop in the number of high achieving students from the early 1960s to the mid 1980s. In 1985 for example, the U.S. Navy reported that one-quarter of its recruits could not read at the ninth-grade level, the minimum needed to understand written safety instructions. Approximately 13% of all 17-year-old students and 40% of minority youths were functionally illiterate.92

Concomitant with this drop in student achievement levels was the sharp rise in annual expenditures per pupil, corrected for inflation, from an equivalent of $478 in 1929 to $2,670 in 1982, and $5,208 in 1991.93 Increases in teacher salaries were often blamed for the rising costs of education. However, the statistic not reported was the proliferation of public school bureaucracies. The

---

93 Ibid., p. 347.
number of pupils per administrator has dropped from 523 in 1950 to 295 in 1980, a 44% decrease in the number of pupils per administrator. When administrative costs are included in the annual expenditures per pupil, we find that over a thirty year period the teachers’ portion of per pupil costs dropped from 52.4% in 1960 to only 38.8% in 1980, a decrease of 26%. Policy makers interested in reducing the per pupil cost of education need to concern themselves with using the available funds more efficiently by directing them away from burgeoning bureaucracies to the classrooms where they are needed. These regression results concerning the positive relationship between grade school teacher salaries and high school student reading score gains is one example of the positive impact that school funding can have on student achievement.

*Grade School Location & Type*

Similar to the results of the mathematics gain score regression, the location of the grade school in urban, suburban, or rural areas was not significantly related to student reading gain score. Neither the coefficients of the type of school, public, Catholic or private nor the student teacher ratios in the public grade schools were significantly related to student achievement in reading scores.

*High School Tracking*

Although most of the grade school characteristics included in the regression analysis were not significantly related to student reading gain scores, three high school characteristics were. First, student enrollment in a school’s vocational track is significantly negatively related to reading gain scores. Unlike the regression results for mathematics gain scores, we find that the size of the reading gain score coefficient indicates that students enrolled in a vocational track score 4.3 percentage points below the average for the sample, and this result is significant at the 10 percent level. The
regression results imply that academic track has less of an impact on reading gain scores than it does on mathematics gain scores.

One plausible explanation for the different student gain scores in mathematics and reading may be related to the fact that the critical division in reading scores between the skills-proficient and skills-deficient students occurs in the fourth grade. During the fourth grade, the demands made upon students in reading change, moving quickly beyond the decoding of sentence structure to critical thinking skills involving the assimilation, evaluation, and ordering of information. Students who do not make the critical transition to the higher level reading skills in the fourth grade are often drilled extensively in rudimentary skills to the point of frustration and still cannot read well. The available evidence tends to show that students learn to read in different ways, at different rates, and at different times in their lives. Although the reasons for these differences remain unclear, enrollment in high school academic tracks seem to have little impact on the improvement of student reading skills. Since the fundamental skills required for higher order reading levels are acquired in grade school, it is plausible that students self-select into vocational or academic reading tracks based on information about their natural abilities reflected in their grades. High school reading scores may proxy for natural ability leading students of higher ability to self-select into academic reading tracks while others would sort themselves into vocational or other tracks.

*Homework*

As in the case of the mathematics gain scores, the regression results indicate that hours of homework is significantly positively related to reading gain scores. The size of the coefficient’s t-ratio is 4.625, the largest of the regression’s positive coefficients. This does not contradict the regression results for academic track, but suggests that the lower level reading skills learned in the early years of grade school enable students to master the higher order skills which can be improved
through the assignment of homework outside of class. Although student ability plays an important role in reading skills, the assignment of homework provides the time and practice necessary to improve and master the higher order skills required for reading comprehension, such as the assimilation, evaluation, and ordering of information.

**High School Problems**

Finally, the third high school characteristic most negatively related to student reading score gains was the student's report of violence in and around the school vicinity. The t-ratio corresponding to the variable's coefficient was -5.419, the largest in absolute value for any of the other variables. These regression results re-affirm that the presence of violence in our urban and suburban high schools is a serious problem, is consistently related with lower student achievement, and needs to be addressed if progress is to be made in reducing truancy, dropouts, and enhancing students' desire to attend school.

**Peer Group and Family Characteristics.**

Peer group effects, family background variables, socioeconomic status, high school location, and parental aspirations have no significant relationship to student reading gain scores. Because of this, these variables were deleted from the final regression through the stepwise deletion process.

**Change in Public High School Student Science Test Scores**

Having reviewed the regression results for the relationship between various student, school, peer, and family background characteristics and gains in mathematics and reading scores, we now turn to the third part of the model concerning the student science gain score regression. The change in standardized science test scores for each student is measured between the eighth and

---

twelfth grades. A stepwise regression technique was used to delete variables which proved to be insignificant in earlier science gain score regression results. The removal of individual factors from the science gain score regression did not affect the significance of the remaining variables, a result which indicates the robust nature of the final regression results. The Mills Ratio, calculated from the results of the public school probit regression, was included as an independent variable in the science gain score regression. The regression results indicated no significant relationship between it and student science gain scores.

**Educational Production Function For Public High School Students, Science Gain Score.**

\[
(6.3) \quad \Delta S_i = \alpha + \beta X_i + \omega P_i + \mu_i
\]

- \(\Delta S_i\) = Calculated student science gain score between the 8th and 12th grades.
- \(\alpha, \beta, \omega\) = Regression coefficients.
- \(X_i\) = Vector of student background characteristics except for the public school choice.
- \(P_i\) = The Mills Ratio for public school choice. \(\mu_i\) = residual

The standardized science tests administered to students in the eighth and twelfth grades marked three levels of proficiency. The first level concerned the understanding of basic science concepts which can be acquired in everyday life. The second level tested for an understanding of fundamental science concepts upon which more complex scientific knowledge can be built. The third and final level concerned an understanding of relatively complex scientific concepts, typically requiring an additional problem solving step. Science gain scores were calculated by subtracting a student’s science score on the eighth grade test from the same student’s score on the twelfth grade test. Calculated science gain scores were then regressed on a set of independent variables representing student, family, school, and peer characteristics. The regression results for student science gain scores are reported in Table 4.
Table 4: Public High School Student Science Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE **</td>
<td>-43.577</td>
<td>12.317</td>
<td>-3.538</td>
</tr>
<tr>
<td>FEMALE **</td>
<td>-120.613</td>
<td>12.998</td>
<td>-9.279</td>
</tr>
<tr>
<td>BLACK **</td>
<td>-158.302</td>
<td>27.115</td>
<td>-5.838</td>
</tr>
<tr>
<td>HISP **</td>
<td>-75.815</td>
<td>25.076</td>
<td>-3.023</td>
</tr>
<tr>
<td>P8GRADES **</td>
<td>3.598</td>
<td>1.092</td>
<td>3.294</td>
</tr>
<tr>
<td>P8TRACK **</td>
<td>-30.801</td>
<td>14.836</td>
<td>-2.076</td>
</tr>
<tr>
<td>PNOHSCHL</td>
<td>33.455</td>
<td>106.230</td>
<td>0.315</td>
</tr>
<tr>
<td>POTHER</td>
<td>-13.949</td>
<td>26.636</td>
<td>-0.524</td>
</tr>
<tr>
<td>PCATH **</td>
<td>43.329</td>
<td>16.012</td>
<td>2.706</td>
</tr>
<tr>
<td>S8ENROL</td>
<td>7.501</td>
<td>6.509</td>
<td>1.152</td>
</tr>
<tr>
<td>S8PROB *</td>
<td>1.292</td>
<td>0.721</td>
<td>1.793</td>
</tr>
<tr>
<td>P8COMPUT **</td>
<td>31.234</td>
<td>15.114</td>
<td>2.067</td>
</tr>
<tr>
<td>P8ALGEBR</td>
<td>-8.709</td>
<td>12.624</td>
<td>-0.690</td>
</tr>
<tr>
<td>SHENROL</td>
<td>0.016</td>
<td>0.014</td>
<td>1.132</td>
</tr>
<tr>
<td>SH2ACADM **</td>
<td>67.313</td>
<td>14.700</td>
<td>4.579</td>
</tr>
<tr>
<td>SH2SUBRB</td>
<td>17.571</td>
<td>20.617</td>
<td>0.852</td>
</tr>
<tr>
<td>SH2PROB **</td>
<td>-6.626</td>
<td>2.034</td>
<td>-3.258</td>
</tr>
<tr>
<td>PH2HMWK **</td>
<td>10.804</td>
<td>3.409</td>
<td>3.169</td>
</tr>
<tr>
<td>GRTECH</td>
<td>9.221</td>
<td>10.841</td>
<td>0.851</td>
</tr>
<tr>
<td>SHGEN **</td>
<td>0.821</td>
<td>0.353</td>
<td>2.330</td>
</tr>
<tr>
<td>SHPREP **</td>
<td>1.473</td>
<td>0.386</td>
<td>3.821</td>
</tr>
<tr>
<td>SHSPEC **</td>
<td>4.329</td>
<td>1.532</td>
<td>2.827</td>
</tr>
<tr>
<td>SHSALARY</td>
<td>-0.0003</td>
<td>0.002</td>
<td>-0.176</td>
</tr>
<tr>
<td>S8SALARY</td>
<td>5.568</td>
<td>6.135</td>
<td>0.908</td>
</tr>
<tr>
<td>MAEXPCT</td>
<td>2.586</td>
<td>11.164</td>
<td>0.232</td>
</tr>
<tr>
<td>PAEXPCT</td>
<td>0.695</td>
<td>11.038</td>
<td>0.063</td>
</tr>
<tr>
<td>SESQ</td>
<td>7.470</td>
<td>7.021</td>
<td>1.064</td>
</tr>
<tr>
<td>FAMSIZ</td>
<td>0.260</td>
<td>5.284</td>
<td>0.049</td>
</tr>
<tr>
<td>SINGLE</td>
<td>-4.691</td>
<td>20.636</td>
<td>-0.227</td>
</tr>
<tr>
<td>CATCOST</td>
<td>-6.942</td>
<td>19.289</td>
<td>-0.360</td>
</tr>
<tr>
<td>RELCOST</td>
<td>-0.418</td>
<td>2.653</td>
<td>-0.157</td>
</tr>
<tr>
<td>PRIVCOST **</td>
<td>10.125</td>
<td>4.723</td>
<td>2.144</td>
</tr>
<tr>
<td>MILLS</td>
<td>5.489</td>
<td>36.402</td>
<td>0.151</td>
</tr>
<tr>
<td>INTERCEPT **</td>
<td>854.981</td>
<td>236.376</td>
<td>3.617</td>
</tr>
</tbody>
</table>

Degrees of Freedom: 33
Number of observations used in calculations: 4,456
F Value: 11.798
* significant at the 10% level of significance
** significant at the 5% level of significance
Student Characteristics

Age

As in the case of the students' mathematics and reading gain scores, the age of the student was significantly negatively related to the gain in science scores between the eighth and twelfth grades. This result came as no surprise given that older students in the sample might have lacked the higher order skills in reading and mathematics needed to comprehend and analyze some of the items on the science test. This problem is further exacerbated if students who performed poorly in earlier mathematics courses opted out of taking higher level science courses in high school. Although the evidence is not sufficient to establish firm conclusions regarding causality, the cumulative effect of deficiencies in reading and mathematics exhibited by the older students may well have exerted downward pressures on their science gain scores.

Gender

The regression results for the gain in science scores between the eighth and twelfth grades indicate that females scored significantly below the average for the sample. The size of the coefficient was in the upper range of the regression results, and its t-ratio was -9.279, the largest of all the others in absolute value. Disturbing as this result might be, it was not surprising given the lower than average gain scores female students received in mathematics -7.6 percentage points below the average for the sample. The results of previous studies concerning female scores in high school science tests are consistent with the regression outcomes and may offer plausible explanations for their lower gain scores.

A cursory examination of the differences in science gain scores might lead to the conclusion that males have more innate ability than females and thus score higher on standardized tests. For example, the National Science Foundation (1986) reports that male students scored higher on the
verbal and mathematics portions of the SAT tests in 1974, 1981, and 1984, and that they also scored higher on the chemistry, biology, and physics achievement tests administered in 1984. The same trend is reported for Verbal, Analytical, and Quantitative Graduate Record Exam (GRE) scores (NSF, 1986). The situation, however, is far more complicated and requires that course background as well as socialization factors be taken into consideration.

First, enrollment in higher level mathematics courses directly influences science course enrollment insofar as they provide the higher order critical skills essential to the scientific method of inquiry. These higher level mathematics courses have traditionally enrolled a larger number of male students than female students. Lack of sufficient mathematics background has been described as a "critical filter" which might deny female students entry into upper level science classes. Second, gender differences extend beyond differences in achievement. The fact that few or no gender differences are apparent until puberty suggests that socialization may contribute part of the explanation for these differences. One study indicated that at age nine females want science experiences, but were not getting them. By the time they were between the ages of 13 and 17, they no longer felt a desire to enroll in science since they were perceived as boring and male oriented. Finally, previous studies point out that male and female students in high school science classes had different experiences, despite being in the same classrooms together. In one crucial area, academic interactions, female students received less attention then males in the lecture format as well as in the laboratory classroom. In the lecture format, male students are more often chosen by name to respond to questions leaving the females with fewer opportunities to participate in class. In the self-paced labora-

tory classroom, teachers often favor males by asking them more procedural questions than females, even though females initiated more academic questions than the males. The lack of sufficient mathematics background, the difference in early science experiences, and the subsequent change in attitude toward science by females may be based in cultural and social factors engendered in the school by way of student/teacher interactions or by the lack of them.

Race

The science gain score regression results show that both African and Hispanic American students scored below the average for the sample by 18.5 and 8.8 percentage points respectively. The large t-ratios, associated with each of the two ethnic variables, indicate that the results are robust. The lower gain scores for the African American students could be related to the cumulative effects of their lower gain scores in mathematics and reading, 7.7 and 9.1 percentage points below the average for the sample respectively. Given the complex nature of the interaction of factors affecting student achievement in science, there may be several plausible explanations for this achievement gap. First, as is the case with female students, the lack of sufficient mathematics background may be a filter which either denies blacks access to upper level science classes, or severely impedes their ability to perform well. Second, school factors such as the presence of violence and low budgets in our inner city and urban schools, many of which have a large proportion of black and minority students, might exert downward pressures on minority student science gain scores. Third, their lower than average reading gain scores, related to both school and cultural factors, might affect their ability to understand and analyze word problems common in the study of the physical sciences.

The lower than average science gain scores for Hispanic students is more difficult to explain. The regression results indicate a negative relationship between reading gain scores and Hispanic ethnic background, but the relation was barely significant at the 10 percent level. Hispanic social and cultural characteristics might provide insights into the causes of their below average science gain scores.

It has often been held that American Hispanics form a minority group which shares a common cultural, historical, and religious heritage as well as an experience of low educational and economic attainment. These experiences do not produce a homogeneous group of students. First, the claim that they experience low educational attainment may be true in the long run, but these regression results indicate that, of the four subjects tested, only their science gain scores are significantly lower than the average, at the 5 percent level. Second, Hispanic linguistic abilities run the gamut of total adaptation to the English language, from residents of the Southwest who have inhabited the area for four hundred years, to Chicanos born and brought up in the "barrios" as monolingual Spanish speakers until they enter grade school. This variety of English linguistic ability might be reflected in the negative, and barely significant, relation between the Hispanic variable and reading gain scores. Furthermore, Hispanics have a mixed Indian/European/African ancestry. Their skin and culture reflect this mix, which has engendered attitudes and behaviors that differ in subtle and significant ways from those of the dominant culture of the United States. Some culturally determined behaviors include ways of dealing with authority figures, expressing disagreement, and communication styles which could create difficulties for Hispanic students in the classroom and provide possible explanations for their lower than average gain scores in science.

Teachers and administrators are authority figures of high status in Hispanic culture. Hispanics often keep a respectful distance between themselves and their teachers. This could lead to difficulties if teachers interpret their silence as a lack of interest rather than as a sign of politeness. Disagreeing with or questioning the opinions of others, most especially their teachers, would also be a sign of disrespect. In such cases their silence could be interpreted as a lack of interest or concern. For Hispanics, eye-to-eye contact is a sign of challenge or, if between members of the opposite sex, seduction. American teachers may get the impression that the Hispanic student who does not maintain eye contact, actively participate in class discussions, or question the teacher is not paying attention or is uninterested in the class. Cultural differences in the modes of social communication can easily lead to a breakdown in student/teacher interactions within the classroom contributing to ethnic differences in science achievement scores. The lower science gain scores of Hispanic students might very well be the result of insufficient student/teacher interactions, a factor also attributed as a cause for the lower science gain scores of female students in the public school system.

Religion

Of the five religious backgrounds described earlier in this chapter, all of them except the Catholic variable, were not significantly related to student gain scores in science. The size of the coefficient of the Catholic religion variable indicated that Catholic students achieved a gain score 5.1 percentage points higher than the average for the sample. The magnitude of the regression result is small, but significant at the 5% level. This result is consistent with the findings concerning the student gain scores in mathematics, which were 2.6 percentage points higher than the average for the sample. The reasons for the difference are beyond the scope of the regression results, but
one could postulate the presence of a work ethic peculiar to Catholics which might explain why Catholic students tend to outperform the average student in mathematics, science, and reading.

**Student Eighth Grade Score Composites**

Student academic aspirations in the eighth grade were not significantly related to science gain scores, so those variables were deleted from the present regression by way of the stepwise deletion process described earlier in the chapter. The variable representing exposure to Algebra I in the eighth grade was not significantly related to student science score gains in high school and was also deleted from the present regression analysis. The background academic variable which was significantly related to science gain scores was the student’s eighth grade score composite. Since this composite variable includes all of a student’s elementary school grades, it might serve as a proxy for student academic ability. The coefficient of this variable was not significantly related to student gain scores in reading and history (as we shall see below). However, the fact that it is positively associated with high school gain scores in both mathematics and science might indicate that students who master the skills taught in grade school are better prepared for high school mathematics and science courses.

**School Characteristics**

**Grade School Location, Type, Student/Teacher Ratio**

As we saw in the mathematics and reading gain score regression results, the location of the grade school in urban, suburban, or rural areas was not significantly related to student science gain scores. Neither the coefficient of the type of grade school-public, Catholic, or private-nor the student/teacher ratios in the grade schools were significantly related to student achievement in science gain scores. The regression results would seem to indicate that these grade school characteristics have little impact on student science achievement levels in high school.
High School Tracking

Of the school variables which remained significantly related to science gain scores, academic track had the largest positive coefficient and t-ratio. The size of the coefficient indicated that students enrolled in a public school’s academic track tended to achieve gain scores in science more than 7.9% higher than the average gain score for the sample. Academic tracking in science was associated with largest gain scores for all of the subjects tested. A possible explanation for this result is that science is a field which combines the higher order skills of inquiry and reasoning learned in other subjects, such as reading and mathematics, and applies them to the study of the natural laws and processes of the physical and biological sciences. The regression results indicate a strong positive relationship between academic tracking and superior academic achievement in each of the three other subjects where these higher order skills are learned. It is not unreasonable to conclude that high school students who choose to enroll in an academic track are better prepared to succeed in science because of natural ability or strong motivation.

Unlike the regression results for mathematics, enrollment in a non-academic track, such as the vocational track, was not significantly related to student science gain scores. A possible reason for this result is that students enrolled in the vocational track might take the same basic science courses required of all students including those in the general high school track. Therefore, it was not surprising to find a lack of any significant difference in science gain scores between those enrolled in the vocational and other high school tracks.

Homework

Similar to the regression results for mathematics and reading, the total number of hours students reported doing homework outside of school is significantly positively related to science gain scores. The variable’s positive t-ratio of 3.169 is second only to academic track, which has a
value of 4.579. This result came as no surprise given the fact that the higher order skills applied in the scientific method often cannot be learned by rote memory but rather by experience and practice. Although student ability and motivation play an important role in academic achievement, the assignment of homework provides the time and practice required to master the higher order skills implicit to the scientific method. Even the most able and motivated students need time to practice these skills if they are to master them. The regression results indicating the strong positive relationship between the assignment of homework and science gain scores during the high school years should provide educators with an incentive to improve student science skills through the assignment of extra homework.

School Problems

The presence of social problems on our high school campuses has been consistently negatively and significantly related to academic gain scores in mathematics and reading. The regression results for student gain scores in science are no different and show a significant negative relation to the presence of social problems on the high school campus. The size of the coefficient is not large, i.e. -6.626, but is significant with a t-ratio of -3.258. These results should provide educators and policy makers with sufficient evidence that the presence of violence in and around the high school campus is consistently associated with low academic achievement and needs to be seriously addressed if progress is to be made in reducing truancy, dropouts, and enhancing students’ desire to attend school.

Peer Group Characteristics

Ethnic Groups

Another consistent result of this model has been that the ethnic mix of a high school is not significantly related to student academic achievement. Studies of academic achievement in high
schools where a significant portion of the student body is made up of minorities have often associated low academic achievement with the presence of minority groups. However, when the students' experience of violence is added to the regression, the school's ethnic mix and its relationship to academic achievement reduces to insignificance. Ethnic mix in itself is not associated with low academic achievement: rather it is the presence of low school budgets, poverty, and violence which are.

Other Peer Groups

While the percent of high school graduates enrolled in four-year or technical colleges is not significantly related to student science score gains, the percent of the student body enrolled in the general, preparatory, and special educational tracks is significantly positively related to science gain scores. This result was common to both the science and mathematics gain score results. The size of the coefficients are small relative to the coefficient of the constant term, but are significant with t-ratios of 2.33, 3.82, and 2.83 respectively. The presence of productive peer groups, regardless of their ethnic origin, seems to have a positive impact on student science gain score.

Family Characteristics

Similar to the regression results for the reading gain scores, the family background characteristics representing parental marital status, family size, socioeconomic status, and the like were not significantly related to science gain scores. Therefore, all of these variables have been removed from the present regression analysis by way of the stepwise deletion process described earlier in the chapter. The lack of any significant relationship between student family characteristics and science gain scores may be an important result in so far as it underscores the significance of the school environment in promoting student productivity. The implications of this result is that policy makers need not be as concerned with policies aimed at changing family characteristics, such as socioeco-
nomic status, parental marital status, or even family size. Rather, they could be engaged in the formation of policies concerning school characteristics, over which they do have more control, such characteristics as school size, the assignment of homework, the reduction of violence in and around the high school campus, the promotion of student/teacher interactions regardless of gender or race, and the encouragement of student enrollment in academic tracks.

**Change in Public High School Student History Test Scores**

This equation of the model regresses the change in student standardized test scores in social studies (history) between the eighth and twelfth grades. Student gain scores were calculated by subtracting a student’s eighth grade score from the same student’s twelfth grade score in social studies. The calculated gain scores were then regressed on a set of independent variables representing the student’s personal, family, peer, teacher and school characteristics. As in the previous parts, a stepwise deletion process was used to remove variables which proved to be insignificant in earlier history gain score results. In order to control for the possible omission of significant student characteristics, or other unmeasurable-related to both school choice and student ability, the Mills Ratio, constructed using the results from the probit equation, was inserted into the regression equation as one of the independent variables. The size and significance of the Mills Ratio will indicate and correct for any bias in the regression statistics resulting from the omission of important student attributes. It was found that individual factors removed from the regression equation in the stepwise deletion process, due to their insignificant relationship to student history gain scores, when taken cumulatively once again resulted in a significant Mills Ratio coefficient. Therefore it was included in the history gain score regression, as one of the independent variable, to correct for unmeasured factors relating to both choice of public high school and history achievement scores. The results of the student history gain score regression analysis are presented in Table 5.
Educational Production Function For Public High School Students, History Gain Score.

\[ \Delta H_i = \alpha + \beta X_i + \omega P_{i}^{*} + \mu_i \]

\[ \Delta H_i = \text{Calculated student history gain score between the 8th and 12th grades.} \]
\[ \alpha, \beta, \omega = \text{Regression coefficients.} \]
\[ X_i = \text{Vector of student background characteristics except for the public school choice.} \]
\[ P_{i}^{*} = \text{The Mills Ratio for public school choice} \]
\[ \mu_i = \text{residual} \]

Student Characteristics

Age

Since the regression results for student standardized gain scores in history are consistent with the previous results in mathematics, reading, and science, summaries of the probable causes rather than detailed discussions relating to the more significant results will be provided.

Gender

The regression results indicate that being female is significantly negatively related to student gain scores in social studies. The size of the coefficient indicates that female gain scores were 6.1 percentage points below the average for the sample. The cause for the lower achievement of female students in social studies is puzzling given the fact that their gain scores in reading were not significantly different from the rest of the sample. This implies that there were no significant gender differences regarding the higher order cognitive skills required to understand and synthesize the ideas which have shaped the larger social, political, and economic structures of our society. It may well be that gender-role socialization in the home and student-teacher interactions in the school have led to differences in interest and participation in the high school social studies curriculum.

Race

The regression results for the African and Hispanic Americans indicate that there is no
Table 5: Public High School Student History Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE **</td>
<td>-13.585</td>
<td>10.712</td>
<td>-1.268</td>
</tr>
<tr>
<td>FEMALE **</td>
<td>-40.906</td>
<td>11.582</td>
<td>-3.532</td>
</tr>
<tr>
<td>BLACK</td>
<td>14.206</td>
<td>24.404</td>
<td>0.582</td>
</tr>
<tr>
<td>HISP</td>
<td>10.072</td>
<td>21.964</td>
<td>0.459</td>
</tr>
<tr>
<td>ASIAN **</td>
<td>58.912</td>
<td>26.765</td>
<td>2.201</td>
</tr>
<tr>
<td>PCOLL **</td>
<td>-43.792</td>
<td>18.139</td>
<td>-2.414</td>
</tr>
<tr>
<td>PCATH **</td>
<td>37.428</td>
<td>14.769</td>
<td>2.534</td>
</tr>
<tr>
<td>POTHER</td>
<td>15.751</td>
<td>23.707</td>
<td>0.664</td>
</tr>
<tr>
<td>PNONE</td>
<td>14.606</td>
<td>23.569</td>
<td>0.620</td>
</tr>
<tr>
<td>S8PRIVAT **</td>
<td>174.912</td>
<td>52.859</td>
<td>3.309</td>
</tr>
<tr>
<td>P8HTTRACK *</td>
<td>-25.986</td>
<td>13.970</td>
<td>-1.860</td>
</tr>
<tr>
<td>PH2HMWK **</td>
<td>9.592</td>
<td>2.838</td>
<td>3.380</td>
</tr>
<tr>
<td>S8PROB **</td>
<td>2.440</td>
<td>0.642</td>
<td>3.798</td>
</tr>
<tr>
<td>S8SALARY **</td>
<td>17.342</td>
<td>4.720</td>
<td>3.674</td>
</tr>
<tr>
<td>S8ENROL</td>
<td>6.569</td>
<td>5.694</td>
<td>1.154</td>
</tr>
<tr>
<td>SHENROL</td>
<td>-0.016</td>
<td>0.012</td>
<td>-1.301</td>
</tr>
<tr>
<td>PH2HMWK</td>
<td>8.832</td>
<td>3.036</td>
<td>2.909</td>
</tr>
<tr>
<td>SH2SUBRB **</td>
<td>-38.477</td>
<td>18.439</td>
<td>-2.087</td>
</tr>
<tr>
<td>SH2ACADM **</td>
<td>36.417</td>
<td>13.250</td>
<td>2.748</td>
</tr>
<tr>
<td>SH2VOTE **</td>
<td>-70.763</td>
<td>19.471</td>
<td>-3.634</td>
</tr>
<tr>
<td>SH2OTHER **</td>
<td>45.270</td>
<td>25.698</td>
<td>-1.762</td>
</tr>
<tr>
<td>SH2PROB **</td>
<td>-7.115</td>
<td>1.809</td>
<td>-3.934</td>
</tr>
<tr>
<td>GROTHER</td>
<td>7.211</td>
<td>5.519</td>
<td>1.307</td>
</tr>
<tr>
<td>SHSPEC **</td>
<td>4.689</td>
<td>1.431</td>
<td>3.277</td>
</tr>
<tr>
<td>FAMSIZ *</td>
<td>8.493</td>
<td>4.626</td>
<td>1.836</td>
</tr>
<tr>
<td>SHSALARY</td>
<td>0.001</td>
<td>0.002</td>
<td>0.684</td>
</tr>
<tr>
<td>CATCOST</td>
<td>-16.846</td>
<td>17.735</td>
<td>-0.950</td>
</tr>
<tr>
<td>RELCOST</td>
<td>-1.710</td>
<td>2.426</td>
<td>-0.705</td>
</tr>
<tr>
<td>PRIVCOST</td>
<td>5.257</td>
<td>4.257</td>
<td>1.235</td>
</tr>
<tr>
<td>SESQ *</td>
<td>10.572</td>
<td>5.939</td>
<td>1.780</td>
</tr>
<tr>
<td>SINGLE</td>
<td>-10.589</td>
<td>18.832</td>
<td>-0.562</td>
</tr>
<tr>
<td>MILLS **</td>
<td>24.656</td>
<td>33.860</td>
<td>0.728</td>
</tr>
<tr>
<td>CONSTANT **</td>
<td>666.906</td>
<td>199.396</td>
<td>3.345</td>
</tr>
</tbody>
</table>

Degrees of Freedom: 30
Number of observations used in calculations: 4,169
F Value: 6.744
* significant at the 10% level of significance
** significant at the 5% level of significance
significant difference between their gain scores and those in the rest of the sample. This implies that they performed better in social studies than they did in mathematics and science. One of the more interesting regression results is that Asians enjoyed a gain score in social studies more than 8.8 percentage points above the average for the sample. The larger than average history gain scores could be attributed in part to the fact that Asian students reported spending 25% more time doing homework than any other ethnic group in the sample. It is well known that academic success is highly valued among Asian immigrants. Social pressures from the family (especially the father) or community exerted on Asian students could account for the above average gains in social studies between the eighth and twelfth grades.

**Religion**

Similar to the gain score regression results for mathematics, reading, and science, the Catholic religious background variable is the only one significantly related to student history gain scores. The size of the coefficient indicated that Catholic students gained 5.6 percentage points more than the average student in the sample. The t-ratio for this coefficient is 2.53 indicating that this difference is significant at the 5% level. The reason for this religious background advantage is still not clear, but it is plausible that this variable serves as a proxy for a student’s productive ethnic background which manifests itself in above average student gain scores in social studies, mathematics and science.

**Student Eighth Grade College Aspirations**

Previous regression results for mathematics indicated that eighth grade students who intended to finish college enjoyed larger gain scores than the average students in the sample. In the regression results for social studies and reading, the student’s intention in the eighth grade to finish college was not significantly related to student gain scores. However, the intention to attend college
for at least a couple of years resulted in a coefficient significantly negatively related to student gain scores. The size of the coefficient indicated that these students scored 6.4 percentage points below the average for the sample. It is plausible that grade school students base their future academic expectations on feedback they receive in the form of grades. If at an early age a student believes that he or she is not academically inclined, then that student’s college expectations may be somewhat limited in scope. A student’s personal assessment of lower academic ability may have a negative effect on motivation and partially account for lower achievement in reading and social studies.

School Characteristics

Grade School Characteristics

Earlier gain score regression results for mathematics, reading, and science indicated that the type of grade school, public, Catholic, or private, was not significantly related to these student gain scores. This result is dramatically different in the area of social studies, where private non-religious grade school students achieved history gain scores 26 percentage points above the average for the sample. Upon re-examining the data for the public high school student standardized test scores, it was found that students who graduated from private non-religious grade schools had higher gains scores as well as higher standardized scores in all subjects tested compared to students who graduated from public or other private grade schools. In order to test for other significant student characteristics which might account for their superior achievement in social studies, the variable representing student graduation from an eighth grade private non-religious school was regressed on a set of forty-two student background characteristics. Due to the discrete nature of the choice process, a probit regression was used. The results of the regression indicated that the coefficient of this variable was significantly related to low grade school enrollments, few grade school
problems, small family size, high grade school student/teacher ratios, high socioeconomic status, enrollment in eighth grade Algebra I, and involvement in religious services and activities. It is difficult to determine the exact cause for the large increase in history gain scores for the public high school students who attended private non-religious grade schools. It is plausible that these are the important factors, combined with the social pressure (from the family or community) exerted on them to succeed academically, which account for their above average gains in social studies.

*Grade School Teacher Salaries*

The regression results indicate that grade school teacher salaries are significantly positively related to student gain scores in social studies. The result implies that higher grade school teacher salaries are associated with large social studies gain scores for students in the twelfth grade. It is plausible that schools which offer high base salaries attract and retain highly productive teachers whose students benefit from the quality instruction received at the grade school level.

*High School Tracking*

Student enrollment in a school's academic track is significantly positively related to student gain scores in social studies. The size of the coefficient indicates that students enrolled in the academic track realize a gain score 5.5 percentage points higher than the average for the sample. Students enrolled in the vocational track experienced on average a gain score in social studies 10.6 percentage points lower than the average for the sample. The coefficient of the variable representing a school track other than academic or vocational was also significantly negatively related to student gain score in history. The size of the coefficient indicated that students enrolled in tracks other than the general, vocational, or academic scored 6.8 percentage points below the average for the sample. These results should not be surprising since the emphasis in non-academic tracks
would be to train students in a different set of skills not always related to formal academic achievement.

*Homework*

The hours a student spends in doing homework outside of school, during the week and on weekends, is also significantly positively related to history gain scores. This result is similar to those for mathematics, reading, and science. Although one may not be able to attribute causation to the hours of homework a student is engaged in per week, there certainly is a significant relationship between it and student gain scores in all four subject areas.

*School Problems*

The presence of problems in the high school was also significantly negatively related to student gain scores in social studies. This result is consistent with every other gain score in the study, and indicates that the presence of problems in the high school is always associated with lower student achievement.

*Peer Group Characteristics*

Once again we find the regression results indicate that the ethnic mix of a high school is not significantly related to student gain scores. The percent of the student body enrolled in special education classes is significantly positively associated with student gain scores in social studies. This result is consistent with those for mathematics and science. The regression results also indicate that the presence of these programs do not negatively affect the education of those enrolled in the general, academic, or vocational programs. Although such programs are expensive, their effects on academic achievement seem to be well worth the costs.

*Family Characteristics*

*Family Size*
The only family characteristic significantly related to social studies gain scores was the size of the student's family. This characteristic was positively related to student gain scores and is significant at the 10 percent level. There are two factors which make this particular result interesting. First, the significant positive relationship between family size and student gain scores is peculiar only to the social studies gain score results. The student gain scores regressions for the other three subjects resulted in positive coefficients for the variable representing family size, but the t-ratios indicated that they were not significant, even at the 10 percent level. The second factor concerns the results of past studies, which often indicate a negative relationship between family size and student achievement scores. These results, however, do not represent a consensus in the research literature.

Most cross-sectional research shows that children from larger families tend on average to leave school earlier and to score lower on intelligence and achievement tests than their own peers from smaller families. However, many researchers have reached fundamentally different conclusions about the causes of this association between family size and achievement. The primary root of the disagreement concerns the problem of confounding. In research pertaining to the relationship between student background variables and academic achievement, confounding involves the omission or inadequate treatment of certain variables associated with both student achievement and student background variables of interest which might be responsible for the associations between them. For example, family size (the student characteristic of interest) is usually related to other characteristics such as ethnicity and socioeconomic status, that are in turn significantly related to student academic achievement. In the United States, the average number of children per family was

---

approximately 1.8 among whites, 2.2 among Hispanics, and 1.9 among blacks in 1984.\textsuperscript{102} Similarly, families with a greater number of children are headed by parents who have on average lower educational attainment and lower occupational prestige. Unless past studies control for both ethnicity and socioeconomic status, the effects of these characteristics would be mistakenly attributed to family size.

The student gain score regression results for all four subjects controlled for both socioeconomic status as well as ethnicity. The first result is that socioeconomic status was not significantly associated with student gain scores in two of the four regression results. Second, blacks, Hispanics, and Native Americans received lower gain scores on at least two of the four subjects tested. Third, when controlling for both ethnicity and socioeconomic status, family size was significantly positively related to student gain scores in the area of social studies, but not in any of the three other subject matters. According to these regression results, there is not sufficient evidence to support the claim that students of larger families tend to gain less from schooling than their peers in smaller families.

\textsuperscript{102} Congress of the United States, "Educational Achievement," p. 61.
CHAPTER VII

PROBIT REGRESSION RESULTS FOR CATHOLIC HIGH SCHOOL CHOICE

Selectivity Bias

Chapter 6 was concerned with the study of the academic performance of public high school students measured in terms of their gain scores in mathematics, reading, science, and social studies. This section is concerned with the achievement outcomes of students attending Catholic high schools. It is well noted in the education literature that Catholic school students academically outperform their peers in the public schools. Initially, the superior performance of Catholic high school students led many to conclude that Catholic schools were more effective than the public schools in promoting academic achievement measured by standardized test scores. However, the effectiveness of Catholic schools in promoting student achievement has been seriously questioned due to the possibility of selectivity bias. Selectivity bias is potentially the most important problem in the literature on the effectiveness of private schools. Unlike public schools, which must accept virtually all students who live within their attendance boundaries, Catholic high schools are free to select their students and to expel students because of poor behavior or academic performance.

Thus, part of the Catholic school effect could be due to the way they choose their students.

Furthermore, it has been noted in the education literature that Catholic high school students are more likely to come from two-adult, high income, well educated families, i.e. they have better observed characteristics. Children of such families generally would be expected to do well in

---

school for two reasons. First, the parents would see that their children attend a better than average school. Second, these children will succeed in part because of factors that cannot be observed but are under the parents' control. They would spend more time reading to their children, they would stress the importance of good grades, and they would see that the children do their homework. Therefore, if students with more ability or students from families that place a higher value on education are more likely to attend Catholic high schools, then models which do not control for the possibility of selectivity bias would overstate the effects of Catholic high schools on student achievement.

A two-equation model, similar to the one used for the public school regressions, was constructed for this analysis to control for the potential correlation between unmeasured determinants of Catholic high school choice and student achievement outcomes. In the first equation of this model, the choice of Catholic or non-Catholic school is explicitly analyzed. Essentially the choice of high school (Catholic = 1 or non-Catholic = 0) is regressed on a set of personal, school, peer, teacher, and family characteristics. Due to the discrete nature of the choice process, a probit regression is required. Since probit regressions operate on individual cases, a listwise deletion process confines analysis to those cases with complete data on all variables in the regression. This process insures that all relevant statistics are based on the same body of data. The results of the Catholic school probit equation are used to construct a new Mills Ratio for Catholic high school students which corrects for potential self-selection bias based on unmeasured factors. The previously calculated student gain scores, based on the results of standardized tests administered to the students attending Catholic high schools, are separately regressed on the Mills Ratio and other student background characteristics.
Probit Regression For Catholic School Students

(7.1) \[ C_i = \alpha + \beta X_i^* + \nu_i \]

- \( C_i \) = dummy variable indicating the choice of Catholic school = 1, otherwise = 0.
- \( \alpha, \beta \) = regression coefficients
- \( X_i^* \) = vector of student and family, teacher, school, and peer characteristics, plus Catholic grade attendance which is significantly related to the choice of Catholic high school but not related to student achievement gains, and not included in the educational production function.
- \( \nu_i \) = residual

In order to insure that the two equation model is properly identified, at least one variable needs to be specified which is related to the choice of Catholic school but not to student achievement. In the present analysis, a dummy variable indicating whether or not a student graduated from a Catholic grade school (1 = attended Catholic grade school, otherwise = 0) is used to identify the system of equations. This particular background variable was chosen because it is significantly positively related to the choice of a Catholic high school, but not significantly related to academic achievement in any of the four gain score regressions.

Earlier versions of the model used a dummy variable which described the student’s religious affiliation as being Catholic (1 = Catholic religion, otherwise = 0) to identify the system of equations. The logic here is that it would be reasonable to expect that being a Catholic would be positively related to the choice of attending a Catholic high school but not significantly related to academic achievement. However, regression results for the Catholic school students did show a significant relationship between the coefficient of the Catholic religion variable and student gain scores. This outcome was not surprising given the results of the public school regressions which also indicated a significant positive relationship between the Catholic religion variable and student achievement in three of the four student gain score regressions. Thus the variable representing a student’s religious background was included in the probit and each of the four gain score
### Table 6. Probit Regression Results of Catholic School Choice on Background Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Pr&gt;Chi</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE **</td>
<td>-0.204</td>
<td>0.084</td>
<td>0.0149</td>
</tr>
<tr>
<td>FEMALE **</td>
<td>-0.247</td>
<td>0.079</td>
<td>0.0017</td>
</tr>
<tr>
<td>BLACK **</td>
<td>0.420</td>
<td>0.150</td>
<td>0.0053</td>
</tr>
<tr>
<td>ASIAN</td>
<td>-0.033</td>
<td>0.168</td>
<td>0.8459</td>
</tr>
<tr>
<td>HISP **</td>
<td>-0.302</td>
<td>0.145</td>
<td>0.0367</td>
</tr>
<tr>
<td>NA</td>
<td>0.317</td>
<td>0.615</td>
<td>0.6062</td>
</tr>
<tr>
<td>P8GRADES</td>
<td>-0.006</td>
<td>0.006</td>
<td>0.3486</td>
</tr>
<tr>
<td>PCOLL *</td>
<td>-0.290</td>
<td>0.152</td>
<td>0.0557</td>
</tr>
<tr>
<td>PFINCOLL **</td>
<td>0.184</td>
<td>0.083</td>
<td>0.0262</td>
</tr>
<tr>
<td>PCATH **</td>
<td>0.688</td>
<td>0.091</td>
<td>0.0001</td>
</tr>
<tr>
<td>S8CATH **</td>
<td>3.298</td>
<td>0.160</td>
<td>0.0001</td>
</tr>
<tr>
<td>S8PRIVAT **</td>
<td>1.340</td>
<td>0.198</td>
<td>0.0001</td>
</tr>
<tr>
<td>S8NAIS **</td>
<td>0.593</td>
<td>0.146</td>
<td>0.0001</td>
</tr>
<tr>
<td>S8RURAL **</td>
<td>-0.545</td>
<td>0.197</td>
<td>0.0056</td>
</tr>
<tr>
<td>S8SUBURB</td>
<td>0.135</td>
<td>0.092</td>
<td>0.1437</td>
</tr>
<tr>
<td>S8SALARY **</td>
<td>0.156</td>
<td>0.036</td>
<td>0.0001</td>
</tr>
<tr>
<td>NH2SUBRB **</td>
<td>-1.153</td>
<td>0.101</td>
<td>0.0001</td>
</tr>
<tr>
<td>MAEXPCT **</td>
<td>0.167</td>
<td>0.082</td>
<td>0.0414</td>
</tr>
<tr>
<td>PAEXPCT</td>
<td>-0.011</td>
<td>0.079</td>
<td>0.8854</td>
</tr>
<tr>
<td>SINGLE</td>
<td>-0.102</td>
<td>0.137</td>
<td>0.4588</td>
</tr>
<tr>
<td>FAMSIZ</td>
<td>-0.041</td>
<td>0.032</td>
<td>0.1944</td>
</tr>
<tr>
<td>SESQ **</td>
<td>0.254</td>
<td>0.048</td>
<td>0.0001</td>
</tr>
<tr>
<td>CATCOST **</td>
<td>0.097</td>
<td>0.038</td>
<td>0.0112</td>
</tr>
<tr>
<td>RELCOST</td>
<td>-0.007</td>
<td>0.014</td>
<td>0.6241</td>
</tr>
<tr>
<td>PRIINCOST</td>
<td>0.007</td>
<td>0.023</td>
<td>0.7633</td>
</tr>
<tr>
<td>INTERCEPT</td>
<td>1.382</td>
<td>1.538</td>
<td>0.3690</td>
</tr>
</tbody>
</table>

Log of the likelihood function: -668.434  
Degrees of Freedom: 25  
Number of observations used in calculations: 5,124
CHART 3: SPECIFICATION OF CATHOLIC SCHOOL PROBIT MODEL ESTIMATED

I. Dependent Variable:

Choice of Catholic high school (Catholic school = 1, otherwise = 0)

II. Independent Variables:

AGE = student’s age
FEMALE = female
BLACK = African American
ASIAN = Asian Pacific Islander
HISP = Hispanic
P8GRADES = student grade composite in eighth grade
PCOLL = student intends to attend college
PFINCOLL = student intends to graduate college
PCATH = student is a member of the Catholic church
S8CATH = student graduated from Catholic grade school
S8PRIVAT = student graduated from non-Catholic grade school
S8NAIS = student graduated from a National Association of Independent Schools grade school
S8SUBURB = student graduated from suburban grade school
S8SALARY = grade school teacher salary
S8RURAL = student graduated from a rural grade school
SH2SUBRB = Student chose to attend suburban high school
MAEXPCT = mother’s future academic expectations for her children
PAEXPCT = father’s future academic expectations for his children
SINGLE = student’s parents are divorced, separated, widowed, never married
FAMSIZ = size of student’s family
SESEQ = student’s socioeconomic status measured in quartiles
CATCOST = Catholic high school tuition
RELCOST = religious high school tuition
PRIVCOST = non-religious private high school tuition
regressions for Catholic high school students. The results of the probit regression are included in Table 6, while descriptions of the included variables are listed in Chart 3.

**Student Characteristics**

*Age*

According to the results of the probit equation, the choice of Catholic high school is significantly negatively associated with the age of the student. The negative relation between a student’s age and the choice of Catholic high school might simply be an indication that the majority of students who chose the Catholic school option are in the grade proper to their age, i.e. an eighth grade student would be thirteen years old. Some students who choose this option would be older because they might have been held back earlier in their academic career or started later than the average student. On the other hand, it is also possible that older students tend to opt out of the Catholic system because they are in need of special education programs only available in the public school system. In this case a student’s age could proxy for the presence of a disability better handled in the public schools.

*Gender*

The probit results indicated that the choice of Catholic high school was significantly positively related to being male. This particular result can be somewhat deceiving at first glance since the actual percentage of the male students attending Catholic high schools does not differ dramatically from that of the public schools. The public school sample was 52.4% male while the Catholic school sample was 53.3% male, a difference of less than one percentage point between the two groups. Given these statistics, there seems to be no significant gender related differences in enrollment between the public and Catholic school sectors.
Race

Similar to the regression results for public school choice, Hispanic and Asian Americans tend to choose the public school option, although the variable for Asian background was not significant. Lack of information concerning private school alternatives, the presence of geographical or ethnic boundaries, and the low socioeconomic status generally characterizing American Hispanics could partially account for their choice of public schools.

The regression results also indicate that the African American variable is significantly positively related to the choice of Catholic schools. This result is not surprising given the increasing enrollment of black Americans in Catholic secondary schools over the past ten years. These schools have a reputation for being college-preparatory schools, and almost certainly one of the reasons parents would choose such schools for their children is that the educational experience provided would better prepare them for college admission and college degrees. Evidence for this is found in the probit regression results concerning parental college aspirations for their children. The variable representing the mother's college expectations for her children was significantly positively related to the choice of Catholic high school. The fact that parental educational expectations were not significantly related to the choice of public high school underlies one difference between Catholic and public school students which might account in part for the differences in academic achievement. Parents of black students who place a high priority on education would be willing to pay the tuition costs of private schooling, especially if the local public high schools are dangerous or of poor quality. The college preparatory reputation of Catholic secondary schools, combined with parental college expectations of African American students, would provide strong incentives for them to choose the Catholic school option.

Student Ability. College Aspirations

Student ability and eighth grade college aspirations may have some impact on the decision to attend college and graduate with a four year diploma. In this analysis, Catholic school choice is regressed on the student’s eighth grade score composite as well as the student’s intention to attend college. The student’s eighth grade score composite was not significantly related to the choice of Catholic high school. This result is important insofar as it does not provide any evidence to support the argument that students in Catholic high schools outperform their public school peers because the Catholic schools themselves screen out the less productive or less able students. What is significant in the choice regression is the student’s intention, in the eighth grade, to graduate college with a four year degree. Future academic expectations can serve as strong motivation for students to make good use of their intellectual time in order to succeed in high school. It might be motivation rather than ability which accounts for the higher academic outcomes of Catholic high school students.

Student Religion

As was expected, the student’s affiliation with the Catholic religion was the only religious characteristic significantly related to the choice of Catholic high school. Approximately 74% of the Catholic school students in this sample were members of the Catholic Church, 14% were of a Protestant denomination, and the rest were affiliated with a variety of other religions. This variable was initially chosen to identify the system of equations since it was assumed to be related to school choice but not academic achievement. However, when the student’s affiliation with the Catholic religion was found to be significantly related to Catholic high school choice and student achievement, a new variable was needed to identify the two-equation system. The variable representing a student’s graduation from a Catholic grade school was significantly related to the choice of at-
tending a Catholic high school, but not significantly related to student achievement. This variable, used to identify the model, was included in the probit regression but not in the four gain score regressions.

**Grade School & Teacher Background Characteristics**

In order to test for the significance of grade school characteristics on the choice of high school, the Catholic school choice variable was regressed on six variables describing the type, location, and base year teacher salaries of the student's grade school. As was expected, the results of the probit regression indicated that the choice of Catholic high school was significantly related to type, location, and grade school teacher salaries.

**Grade School Type**

The first category, depicting the type of private grade school attended by the respondent, was divided into three classifications: Catholic, private religious non-Catholic, and private non-religious. All three were significantly positively related to Catholic high school choice. Catholic high school students tend to be graduates of a variety of private grade schools, although the majority were graduates of Catholic grade schools.

**Grade School Location**

The results of the probit regression indicate that students who have attended a rural grade school are less likely to opt for the Catholic high school choice. Of all the students attending Catholic high schools in the sample, only 1.5% graduated from rural grade schools. The most plausible explanation for this result is that 81.3% of the Catholic high schools in the sample were located in urban areas while the other 18.7% were located in the suburbs. Since students living in rural areas would have less access to these schools, it is likely that the probability of choosing one is far less than would be for those living in the urban and suburban neighborhoods. This turned out
to be the case since over 96% of the students attending the Catholic high schools in the sample reported graduating from urban and suburban grade schools.

**Grade School Teacher Salaries**

An interesting result of the probit regression is that the base year salaries for the students' grade school teachers were significantly positively related to the choice of Catholic high school. Teacher salaries in the Catholic sector are a function of the school's tuition, which varies by location. For example, Catholic grade schools which charge the highest tuition on average are the urban and suburban schools. Over 94 % of the students in this sample who chose the Catholic high school option graduated from urban and suburban grade schools, most of which were private and Catholic. The starting salaries of these teachers are above the national average for the Catholic sector and might account in part for the positive relationship between grade school teacher salaries and the choice of attending a Catholic high school.

**Eighth Grade Peer Group Characteristics**

The relationship between student peer groups and Catholic high school choice is here explicitly analyzed. Two variables which reflect characteristics of the students' grade school peer groups were included in earlier estimates of the probit regression. One describes the percentage of the student's eighth grade class composed of minorities, and the second represents whether the student was enrolled in classes for the gifted. Since neither of these variables were significantly related to Catholic school choice, it would be reasonable to conclude that ethnic mix and relative ability of a student or her peers were not important factors in the decision to attend a Catholic high school.

**Family Background Characteristics**

In order to test for the significance of various family background characteristics, the choice

---

of Catholic high school was regressed on five variables describing the number of parents (adult guardians) in the home, the size of the family, parental expectations, and socioeconomic status.

**Single-Adult Household**

The choice of Catholic high school was regressed on the single-adult household variable along with other student background characteristics. The probit results indicated no significant relationship between the coefficient of the single-adult household variable and the choice of attending a Catholic high school. Although the probit regression results for the public school sector indicated that the choice of public high school was negatively related to coefficient of the single-adult household variable, there is not sufficient evidence to support the argument that single-adult families are opting to send their children to Catholic high schools.

**Parental Expectations for their Children**

The future academic expectations of parents regarding their children’s education is represented by two variables describing the expectations of the student’s father and mother respectively. The regression results indicate that while the mother’s expectations are significantly related to the choice of Catholic high school, the father’s are not. This result highlights one family characteristic which is significantly related to the choice of Catholic high school but not related to the choice of public high school—the college aspirations of the student’s mother for her children. Given the reputation that Catholic high schools have for being college-preparatory schools, it is reasonable to expect parents (or mothers), with strong college aspirations for their children, to choose them as one means of preparation for college. This may be one more factor which accounts for the academic success of Catholic high school students.

**Socioeconomic Status**
The variable representing student socioeconomic status was constructed using available information from the base-year parent questionnaire collected during the 1987-1988 school year. The following data regarding the student’s parents were used: father’s education level, mother’s education level, father’s occupation, mother’s occupation, and family income. Occupational data were recoded using the Duncan SEI continuous index and used in the calculation of the constructed variable. The calculated results were divided into four quartiles, quartile 1 being the lowest and quartile 4 being the highest. The choice of Catholic high school was regressed on this and other family background characteristics. The results of the probit regression indicated that the choice of Catholic high school was found to be significantly positively related to socioeconomic status. Students who chose the Catholic high school option tended to be the children of well educated, high income families.

**Tuition Costs of Alternative Private High Schools**

The tuition costs of private alternatives to the Catholic high school system were entered into the probit regression as independent variables. The same procedure used to proxy the costs of alternatives to the public school option in the public high school probit regression was applied here. The probit regression results indicated no significant relationship between Catholic school choice and the average estimated tuition costs of the two private alternatives. Although the Catholic school option is the least expensive of the three types of private school, there is not sufficient evidence to support the theory that high prices of other private school alternatives drive parents to choose the relatively inexpensive Catholic option. According to the probit results, whatever school and family characteristics contributed to the choice of Catholic high school, tuition costs do not seem to be one of them.

---

Summary

The probit results of the Catholic choice model provide a set of characteristics and experiences often associated with the choice of Catholic high schools. Therefore, students who possess these characteristics are more likely to choose this option over the others available to them. The following is a brief summary of the student background characteristics significantly associated with the choice of Catholic high school:

1.) Catholic high school students tend to be male and in the grade proper to their age group.

2.) The students tend to have strong college aspirations early in their academic careers. These future aspirations are shared (and possibly instilled) by their maternal parents.

3.) The students are most likely to be members of the Catholic Church and to have graduated from a private grade school. Catholic grade schools being the most common.

4.) Finally, they tend to come from families of slightly higher socioeconomic status.
CHAPTER VIII

CATHOLIC HIGH SCHOOL STUDENT GAIN SCORE REGRESSION RESULTS

Selectivity Bias

The second stage of this statistical model involves a value-added approach to student academic achievement in the areas of mathematics, reading, science, and history. Ordinary Least Squares (OLS) analysis is applied to estimate the relationship between Catholic school choice and student gain scores in each of the subjects listed above, while controlling for student personal, family, peer, and school background characteristics. Calculated gain scores in mathematics, reading, science, and history for each student between the eighth and twelfth grades will be regressed separately on a set of student background variables.

Unless all of the relevant factors are included in the regression equations, the unique relationship between each of the factors and student gain scores will be biased. In order to control for the possible omission of significant student characteristics, or other unmeasurableables related to school choice and ability, the Mills Ratio is constructed using the results of the probit equation in the first part of the model and inserted into the regression equations in the second part of the model as one of the independent variables. The size and significance of the Mills Ratio will indicate and correct for any bias in the regression statistics resulting from the omission of important student attributes. Chart 4 summarizes and defines the variables used in the four student gain score regressions. Table 7 reports the regression results for the student gain scores in mathematics, their standard errors, and t-ratios.
CHART 4: SPECIFICATION OF CATHOLIC SCHOOL STUDENT GAIN SCORE REGRESSIONS ESTIMATED

I. Dependent Variables:

- Change in Standardized Mathematics Achievement Test Score
- Change in Standardized Reading Achievement Test Score
- Change in Standardized Science Achievement Test Score
- Change in standardized History Achievement Test Score

II. Independent Variables:

- AGE = student’s age in years
- FEMALE (1 = yes)
- BLACK (1 = yes)
- HISP = Hispanic (1 = yes)
- ASIAN = Asian Pacific Islander (1 = yes)
- NA = Native American (1 = yes)
- P8GRADES = student grade composite in eighth grade
- PNOHSCHL = student does not expect to finish high school
- PCOLL = student expects to attend some college classes but not graduate from college
- PFINCOLL = student expects to graduate from college
- PCATH = student’s religion is Catholic
- POTHER = student’s religion is other than Jewish, Protestant, or Catholic.
- PH2HMWK = hours per week student spends doing homework in high school
- P8ALGEBR = student took Algebra I in eighth grade
- S8CATH = student attended a Catholic grade school
- S8PRIVAT = student attended a non-religious private grade school
- P8COMPUT = student had a computer in eighth grade
- S8ENROL = eighth grade enrollment
- P8GIFTED = student enrolled in eighth grade track for gifted students
- S8MINOR = percent minority in student’s eighth grade
- SH2ACADM = student enrolled in high school academic track
- SH2VOTEC = student enrolled in high school vocational track
- SH2OTHER = student enrolled in high school track other than general, vocational, or general
- SH2SUBRB = high school located in suburb
- S8PROB = index of problems in grade school such as theft, gangs, teacher abuse, etc.
- SH2PROB = index of problems in high school such as theft, gangs, teacher abuse, etc.
- GRTETECH = percentage of high school’s graduates enrolled in college technical school
- GRFORCOL = percentage of high school’s graduates enrolled in a four year college
- GROTHER = percentage of high school’s graduates enrolled a program other than college or vocational school.
- SHMINOR = percentage of minorities in high school
- SHGEN = percent of high school students enrolled in general program
- SHPREP = percent of high school students enrolled in college preparatory program
- SHSPEC = percent of high school students enrolled in special program
- SINGLE = student’s parents are widowed, separated, divorced, or never married
Chart 4 (continued)

PAEXPCT = student father’s educational expectations (0 = no high school 6 = post graduate)
MAEXPCT = student mother’s educational expectations
FAMSIZE = student’s family size (2 - 10)
SESEQ = socioeconomic status of student (quartiles 1-4)
SHRATIO = high school student/teacher ratio
SHSALARY = base year salary for high school teacher in student’s high school
S8SALARY = base year salary for grade school teachers in student’s grade school
CATCOST = average tuition of Catholic high school alternative to the local public high school
RELCOST = average tuition of religious high school alternative to the local public high school
PRIVCOST = average tuition of private high school alternative to the local public high school
MILLS = Mills ratio to correct for selectivity bias

Large national surveys, similar to the one used for this study, are often plagued with the problem of missing data. Due to the missing data in the current survey (NELS88), the same listwise deletion process used in the probit analysis is applied to the student gain score regressions for all four subjects. This process limits the analysis to only those cases with complete data on all items. The current survey has information on 928 of the students attending Catholic high schools in the sample. The results of the first regression, concerned with student gain scores in mathematics, are based on a subsample of 377 students about whom there is complete data on all the variables included in the regression equation.

Educational Production Function For Catholic School Students, Mathematics Gain Score.

\( \Delta M_i = \alpha + \beta X_i + \omega C_i^* + \mu_i \)

\( \Delta M_i \) = Calculated student mathematics gain score between the 8th and 12th grades.
\( \alpha, \beta, \omega \) = Regression coefficients.
\( X_i \) = Vector of student background characteristics except for Catholic school choice.
\( C_i^* \) = The Mills Ratio for Catholic school choice.
\( \mu_i \) = residual
Table 7: Catholic School Student Mathematics Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>-3.639</td>
<td>96.783</td>
<td>-0.038</td>
</tr>
<tr>
<td>FEMALE **</td>
<td>-224.401</td>
<td>86.774</td>
<td>-2.586</td>
</tr>
<tr>
<td>BLACK **</td>
<td>463.600</td>
<td>218.369</td>
<td>2.123</td>
</tr>
<tr>
<td>NA</td>
<td>-394.784</td>
<td>704.072</td>
<td>-0.561</td>
</tr>
<tr>
<td>P8GRADES **</td>
<td>18.305</td>
<td>7.623</td>
<td>2.401</td>
</tr>
<tr>
<td>PMTRACK</td>
<td>-105.290</td>
<td>88.514</td>
<td>-1.190</td>
</tr>
<tr>
<td>P8ALGEBR</td>
<td>8.747</td>
<td>76.359</td>
<td>0.115</td>
</tr>
<tr>
<td>PNOHSCHL</td>
<td>1116.828</td>
<td>511.764</td>
<td>2.182</td>
</tr>
<tr>
<td>PFINCOLL *</td>
<td>-239.049</td>
<td>132.861</td>
<td>-1.799</td>
</tr>
<tr>
<td>PCOLPLUS **</td>
<td>-409.639</td>
<td>156.612</td>
<td>-2.616</td>
</tr>
<tr>
<td>PCATH **</td>
<td>327.663</td>
<td>112.379</td>
<td>2.916</td>
</tr>
<tr>
<td>S8NAIS</td>
<td>-238.026</td>
<td>378.569</td>
<td>-0.629</td>
</tr>
<tr>
<td>S8ENROL</td>
<td>-45.951</td>
<td>52.789</td>
<td>-0.870</td>
</tr>
<tr>
<td>SHENROL</td>
<td>0.092</td>
<td>0.176</td>
<td>0.524</td>
</tr>
<tr>
<td>PH2HMKW **</td>
<td>-41.648</td>
<td>20.601</td>
<td>-2.022</td>
</tr>
<tr>
<td>SH2ACADM **</td>
<td>188.341</td>
<td>95.441</td>
<td>1.973</td>
</tr>
<tr>
<td>SH2VOTEC</td>
<td>-199.237</td>
<td>414.190</td>
<td>-0.481</td>
</tr>
<tr>
<td>SH2SUBRB **</td>
<td>-481.322</td>
<td>121.325</td>
<td>-3.967</td>
</tr>
<tr>
<td>SH2PROB **</td>
<td>-23.205</td>
<td>11.240</td>
<td>-2.064</td>
</tr>
<tr>
<td>SHMINOR **</td>
<td>6.151</td>
<td>2.665</td>
<td>2.308</td>
</tr>
<tr>
<td>GRTECH</td>
<td>-15.887</td>
<td>101.686</td>
<td>-0.156</td>
</tr>
<tr>
<td>GRFORCOL **</td>
<td>-180.003</td>
<td>85.067</td>
<td>-2.116</td>
</tr>
<tr>
<td>SHGEN</td>
<td>-4.463</td>
<td>4.533</td>
<td>-0.985</td>
</tr>
<tr>
<td>SHREP</td>
<td>2.894</td>
<td>2.445</td>
<td>1.184</td>
</tr>
<tr>
<td>SHSPEC</td>
<td>-21.052</td>
<td>15.702</td>
<td>-1.341</td>
</tr>
<tr>
<td>PAEXPCT **</td>
<td>184.675</td>
<td>80.629</td>
<td>2.290</td>
</tr>
<tr>
<td>MAEXPCT</td>
<td>33.127</td>
<td>81.164</td>
<td>0.408</td>
</tr>
<tr>
<td>FAMSZ</td>
<td>6.187</td>
<td>31.417</td>
<td>0.197</td>
</tr>
<tr>
<td>CATCOST</td>
<td>46.537</td>
<td>36.751</td>
<td>1.266</td>
</tr>
<tr>
<td>RELCOST</td>
<td>-2.473</td>
<td>15.463</td>
<td>-0.160</td>
</tr>
<tr>
<td>PRIVCOST **</td>
<td>-58.876</td>
<td>26.874</td>
<td>-2.191</td>
</tr>
<tr>
<td>SHSALARY **</td>
<td>-0.044</td>
<td>0.017</td>
<td>-2.606</td>
</tr>
<tr>
<td>SHRATIO</td>
<td>-27.135</td>
<td>91.312</td>
<td>-0.297</td>
</tr>
<tr>
<td>SESQ **</td>
<td>112.225</td>
<td>50.451</td>
<td>2.224</td>
</tr>
<tr>
<td>SINGLE *</td>
<td>-238.404</td>
<td>137.818</td>
<td>-1.730</td>
</tr>
<tr>
<td>MILLS</td>
<td>173.635</td>
<td>152.094</td>
<td>1.142</td>
</tr>
<tr>
<td>INTERCEPT</td>
<td>1548.572</td>
<td>1904.37</td>
<td>0.813</td>
</tr>
</tbody>
</table>

Degrees of Freedom: 36
Number of observations used in calculations: 358
F Value: 2.837
* significant at the 10% level of significance
** significant at the 5% level of significance
Change in Catholic High School Student Mathematics Test Scores

Student Characteristics.

Gender

One of the more interesting results of the first regression equation is that the change in mathematics scores was significantly negatively related to being female. The negative sign of the coefficient came as no surprise, given the similar results of the public school gain score regression. While female students in the public schools scored approximately 7.6 percentage points below the average for the sample, those in the Catholic high schools tended to score approximately 14.5 percentage points below the sample’s average gain score. This result would normally imply a serious deficiency within the Catholic schools regarding the education of female students in the study of mathematics. A comparison between the average mathematics scores of students in the Catholic and public school sectors will shed light on the nature of these differences.

First, the dependent variable in the regression is a measure of the change in standardized test scores in mathematics between the eighth and twelfth grades. For students in the public high school sector, males increased their scores by 34% while females increased their scores by 31.6% on average. For students in the Catholic sector, males increased their mathematics scores by 37.6% while females increased their scores by 34.7% on average. The significant point here is that the female gain scores in the Catholic sector were larger than those of the female and male students in the public sector.

Second, although the female students in the Catholic sector did not perform as well as their male counterparts, their standardized mathematics test scores were higher than those of the female and male students in the public high schools. When the students were tested in the twelfth grade, the female students in the Catholic schools outperformed the public high school male and female
students them by margins of 11.4% and 8.6% respectively. Female students in the Catholic sector outperformed the public school students of both genders in standardized mathematics test scores as well as gain scores between the eighth and twelfth grades. The reasons for the higher achievement levels of male and female Catholic high school students are still unclear but warrant further research if educators wish to raise the mathematics test scores of public high school students.

Despite the superior performance of Catholic high school female students relative to students in the public sector, there still remains a significant difference in achievement outcomes related to males. One plausible explanation is that in both school types, males outnumber females in the enrollment of more advanced mathematics courses. Taking more advanced mathematics courses may partially account for the superior performance of male high school students in the study of mathematics. Socialization in the home is another possible factor often suggested as preventing girls from developing the characteristics associated with scientific or mathematical studies, characteristics such as independence, convergent thinking, logic, and experimentation. It has also been suggested that observed gender differences in mathematics achievement are the result of intrinsic student characteristics relating to personal competitiveness and self-confidence nurtured in the home long before students enroll in school.

It is noteworthy that students of both genders received higher standardized scores and higher gain scores in mathematics than their counterparts in the public schools. Past studies indicate that mathematics proficiency scores for those who have taken more advanced mathematics courses are higher than those who have not. For example, the proportions of Catholic school students who complete algebra I, geometry, and algebra II are higher than the proportions of students who complete these same courses in the national average. While 14% more catholic school students
complete algebra I than do students in the public schools on average, these figures rise to 28% more students in Catholic schools who complete geometry, and 27% more who complete algebra II. Given the strong relationship between advanced course-taking and proficiency, these differences are very probably the explanation for the higher mathematics proficiency of Catholic school students. These differences in advanced course taking in mathematics might be related to Catholic school specific attitudes which encourage students to enroll in academic tracks and take more advanced courses, which in turn lead to higher levels of student achievement. Future research into these school specific attitudes might provide educators and administrators with valuable insight on how to raise the mathematics achievement levels of public high school students.

**Race**

Many studies indicate that trends in achievement scores have varied among different ethnic groups. The gap in average scores between black and nonminority students still remains large in most cases. The mathematics regression results indicate no similar pattern for any of the minority students enrolled in the Catholic high schools of this study. Mathematics gain scores for Hispanic, Asian, and Native Americans did not differ significantly from the gain scores of the nonminority students in the sample. However, the mathematics gain scores for African American students in Catholic high schools were significantly greater than the average for the sample. Furthermore, all ethnic groups in the Catholic sector received gain scores in mathematics which were greater than those of the minority as well as nonminority students in the public sector. Future research concerning Catholic school and teacher specific characteristics related to student achievement might provide valuable insight to educators and administrators in the public high schools on how to continue to raise achievement levels in mathematics for minority and female students.

---

Religion

The religious background of students attending Catholic high schools was divided into four broad categories including Protestant, Catholic, Other, and no religious background. Similar to the public school regression results, the coefficient of the Catholic religion dummy variable was the only one of four religious categories significantly positively related to student gain scores in mathematics. However, it should be noted that, unlike the public school regression results, mathematics and reading were the only subjects where Catholic high school student gain scores were significantly related to the Catholic religion variable.

The significant positive relation between student affiliation with the Catholic Church and mathematics gain scores raises questions concerning the presence of certain student characteristics associated with the Catholic religion variable, which might account for the superior achievement of Catholic students. In order to identify which student background characteristics related to student religious affiliation might be contributing to student productivity, each of five religious background variables (Catholic, Protestant, Jewish, Other, no religious background) was regressed separately on a list of forty-five student characteristics. Since each of the religious categories were represented by a dummy variable, (if a student is Catholic, then PCATH = 1, otherwise = 0, if a student is Protestant, then PPROT = 1, otherwise = 0, etc.) the discrete nature of the choice process requires a probit regression. A probit regression was constructed and run for each of the five religious background variables using available data from every student in the entire sample. A brief summary of the more interesting results is provided here.

First, graduation from a Catholic grade school was significantly positively related to the Catholic religion variable. Second, Catholic students tend to graduate from grade schools with low
student/teacher ratios. They tend to come from larger families than the average for the entire student sample. In fact, the Catholic religion variable was the only one of the five religious background variables significantly positively related to family size. The larger than average family size of Catholic students does not seem to have a negative impact on student achievement. Finally, a far more interesting result is that the Catholic religion variable is significantly positively associated with the student's intention in the eighth grade to attend and graduate from a four-year college. The results of all five probit regressions indicate that this combination of background characteristics is unique to the Catholic students in the sample. The amount of time students spent watching TV or doing homework were both not significantly related to Catholic religious background. Parental expectations and marital status were not significantly related to the Catholic religious background variable. These regression results would seem to indicate that student educational aspirations at an early age might serve as a powerful motivation for success at the high school level. These college aspirations may be the result of intergenerational transfers of human capital often characteristic of Catholic families.

*Eighth Grade Score Composites*

Student mathematics gain scores were regressed on a variable which combined all of a students' eighth grade scores into one composite grade point average. As expected, the regression results indicated a significant positive relation between the Catholic high school student's eighth grade composite and mathematics gain score. This grade school composite would serve the purpose of providing students with feedback information concerning their ability for academic success as well as encouragement to make the best use of their academic time in high school.

*School Characteristics*
Most of the grade school characteristics included in the mathematics gain score regression turned out to be insignificantly related to student achievement. For example, the location and size of the school, the percent of the student body made up of minorities, student-teacher ratio, and special programs for the gifted were not significantly related to mathematics gain scores.

**High School Location**

According to the regression results for the public high schools, those located in the suburbs were associated with relatively high student gain scores in mathematics. On the other hand, Catholic high schools located in the suburbs were significantly negatively associated with student gain scores in mathematics. Similar results were found for the gain score regression results for science and history. Of all the students in the sample attending Catholic high schools, 18.3% of them were attending suburban or rural high schools and 81.3% were attending high schools in urban areas. These results indicate that students attending Catholic high schools located in urban areas are outperforming students enrolled in suburban Catholic high schools.

This particular result might reflect the general trend in Catholic secondary education where students enrolled in urban Catholic high schools score higher in science and mathematics standardized tests than their peers in the suburbs. One plausible explanation is that urban Catholic high schools are older and more established than their suburban counterparts. Although the urban sprawl of the past thirty years has given rise to the construction of new suburban high schools, 53% of all Catholic high schools are located in urban and inner-city areas. The primary reason for this is financial. Given the limited financial resources of the Catholic Church and the recent shortage of religious sisters and brothers required to staff new high schools, it is often more economically feasible to bus students from the suburbs to the urban schools. The older Catholic high schools have large, well established, and often wealthy alumni groups which help provide the re-
sources, both human and financial, required to operate the schools. Catholic high school principals are well aware of the demands for academic excellence made by those individuals and organizations who support them. Principals and administrators are also aware that if they do not deliver on their promises to provide a high quality product in the form of superior student achievement, important sources of revenue will vanish. The need to preserve the school's reputation is often a very powerful force motivating administrators and teachers to provide high quality education at lower costs.

**Tracking**

Similar to the public school regression results, we find that academic track is significantly positively related to student gain scores in mathematics. This came as no surprise since the majority of studies concerning the relationship between student achievement and grouping according to ability had similar results. There was no significant relationship between student gain scores and other educational tracks offered by the Catholic schools in this study. The lack of any significant differences between the various tracks offered indicates either the presence of a homogeneous group of students in the Catholic high schools or the lack of alternative vocational tracks from which to choose.

**Homework**

The total number of hours students reported doing homework outside of school was significantly negatively related to their gain scores in mathematics. However, it should be noted that on average Catholic school students report spending approximately 15% more time on their homework than their public school peers. Furthermore, students in the urban Catholic high schools report spending about 23% more time doing homework than their peers in the Catholic suburban high schools. The added time spent doing homework might account in part for the superior performance.

---

of the urban Catholic high school students in their mathematics gain scores relative to their peers in the suburban Catholic and public high schools.

**School Problems**

The public school regression results indicated that the presence of school problems was the school characteristic most negatively associated with student gain scores in all four subjects tested. Similar results were found for students attending Catholic high schools. The presence of problems in and around Catholic high schools is associated with lower student gain scores in mathematics. Comparative statistics indicate that the Catholic high school students reported experiencing approximately 17% fewer problems on average than the students in the public schools. For students attending urban high schools the difference was even greater. Catholic school students reported 23% fewer problems than did the students attending public urban high schools. The lower incidence of school violence may be one factor contributing to the superior academic achievement of students in the Catholic high schools.

**Family Characteristics**

**Single-Adult Homes**

The proportion of children living in single-adult homes has grown markedly over the past thirty years, from 9% in 1960 to approximately 25% in 1990. A number of cross-sectional studies have found that children from single-adult households have lower average scores on a number of measures of intellectual and academic achievement. Unlike the results of the public school regressions, which indicated no significant relationship between student gain scores and living in a single-adult home, the Catholic high school regression results indicate a significant negative relationship between the two. Catholic school students from single-adult homes tend to have lower gain scores in mathematics and reading than students from two-adult homes.
Parental Aspirations for their Children

The parents' future academic expectations for their children has often been positively associated with student achievement. In order to test for the possibility of a significant relation between parental expectations and academic achievement, two variables representing the future academic aspirations of the student's father and mother respectively, were added as independent variables in each of the four gain score regression equations. It was found that the father's expectations were significantly positively related to student gain scores in mathematics while the mother's expectations were not. The regression results seem to indicate that the future academic expectations of the students themselves are more strongly associated with academic gain scores than those of the parents.

Mills Ratio

The first part of this model included a probit regression for the choice of Catholic school versus all other types, i.e. public, private, and NAIS. If there is a statistically significant self-selection into the Catholic high school sector based on unmeasured factors, this model postulates a relation between the residual of the school sector choice equation and the residual of the achievement equation. Such a correlation does not exist for the student mathematics gain scores since the Mills Ratio, which is designed to correct for self-selection into the Catholic school sector, is not significantly related to student mathematics gain scores. Therefore, there is not sufficient evidence to indicate the presence of any selectivity bias in these regression results.

Change in Catholic High School Student Reading Test Scores

Now we turn to the reading gain score regression results for students attending Catholic high schools. The change in standardized reading test scores are calculated using the eighth and twelfth grade test results for each student in the sample. Calculated reading gain scores are then
regressed on a list of student background variables. A stepwise regression technique is used to delete variables which proved to be insignificant in earlier reading gain score regression results. In order to control for the possible omission of significant student characteristics or other unmeasurable related to both school choice and ability, the calculated Mills Ratio is inserted into this regression equation as one of the independent variables. The size and significance of the Mills Ratio will indicate and correct for any bias in the regression statistics resulting from the omission of important student attributes. Furthermore, individual factors removed from the regression in the stepwise deletion process, due to their insignificant relationship to student reading gain scores when taken cumulatively resulted once again in a significant Mills Ratio. The Mills Ratio is therefore included in the reading gain score regression to correct for unmeasured factors relating to choice of Catholic school and student reading achievement.

Educational Production Function For Catholic School Students Reading Gain Score.

\[ \Delta R_i = \alpha + \beta X_i + \omega C_i^+ + \mu_i \]

\( \Delta R_i \) = Calculated student reading gain score between the 8th and 12th grades.
\( \alpha, \beta, \omega = \) Regression coefficients.
\( X_i = \) Vector of student background characteristics except for Catholic school choice.
\( C_i^+ = \) The Mills Ratio for Catholic school choice.
\( \mu_i = \) residual

As with the mathematics gain scores, reading gain scores were calculated by subtracting a student's reading score on the eighth grade test from the same student's score on the twelfth grade test. Calculated student reading gain scores were regressed on the Mills Ratio and a set of independent variables representing student, family, school, and peer characteristics. The results of the student reading gain score regression are reported in Table 8.
Table 8: Catholic High School Student Reading Gaia Score Regression Coefficients, Standard Errors, and t-ratios by Specification:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>-65.458</td>
<td>97.002</td>
<td>-0.675</td>
</tr>
<tr>
<td>FEMALE *</td>
<td>-148.922</td>
<td>83.532</td>
<td>-1.783</td>
</tr>
<tr>
<td>BLACK **</td>
<td>-797.143</td>
<td>248.277</td>
<td>-3.211</td>
</tr>
<tr>
<td>ASAIN</td>
<td>184.883</td>
<td>171.603</td>
<td>1.077</td>
</tr>
<tr>
<td>HISP **</td>
<td>-333.332</td>
<td>145.813</td>
<td>-2.286</td>
</tr>
<tr>
<td>NA</td>
<td>-170.050</td>
<td>682.487</td>
<td>-0.249</td>
</tr>
<tr>
<td>P8GRADES</td>
<td>-10.028</td>
<td>7.415</td>
<td>-1.352</td>
</tr>
<tr>
<td>P8TRACK</td>
<td>-85.566</td>
<td>90.912</td>
<td>-0.941</td>
</tr>
<tr>
<td>PCOLL</td>
<td>-112.429</td>
<td>244.812</td>
<td>-0.459</td>
</tr>
<tr>
<td>PFINCOLL</td>
<td>-145.669</td>
<td>217.353</td>
<td>-0.670</td>
</tr>
<tr>
<td>PCOLPLUS</td>
<td>-192.984</td>
<td>239.690</td>
<td>-0.805</td>
</tr>
<tr>
<td>PCATH **</td>
<td>279.335</td>
<td>111.284</td>
<td>2.510</td>
</tr>
<tr>
<td>POOTHER</td>
<td>-198.081</td>
<td>246.326</td>
<td>-0.804</td>
</tr>
<tr>
<td>S8MINOR **</td>
<td>6.871</td>
<td>2.151</td>
<td>3.194</td>
</tr>
<tr>
<td>S8PROB</td>
<td>-0.018</td>
<td>4.357</td>
<td>-0.004</td>
</tr>
<tr>
<td>S8ENROL</td>
<td>19.120</td>
<td>56.954</td>
<td>0.336</td>
</tr>
<tr>
<td>SHENROL</td>
<td>-0.039</td>
<td>0.174</td>
<td>-0.224</td>
</tr>
<tr>
<td>PH2TV **</td>
<td>38.402</td>
<td>17.272</td>
<td>2.223</td>
</tr>
<tr>
<td>P8COMPUT</td>
<td>20.518</td>
<td>78.421</td>
<td>0.262</td>
</tr>
<tr>
<td>SH2ACADM</td>
<td>-47.654</td>
<td>93.327</td>
<td>-0.511</td>
</tr>
<tr>
<td>SH2VOTEC **</td>
<td>-1108.31</td>
<td>406.610</td>
<td>-2.726</td>
</tr>
<tr>
<td>SH2PROB **</td>
<td>-50.936</td>
<td>11.476</td>
<td>-4.438</td>
</tr>
<tr>
<td>GRTech **</td>
<td>147.390</td>
<td>84.768</td>
<td>2.918</td>
</tr>
<tr>
<td>GRFORCOL *</td>
<td>143.300</td>
<td>75.427</td>
<td>1.900</td>
</tr>
<tr>
<td>SHPREP *</td>
<td>-4.462</td>
<td>2.022</td>
<td>-2.206</td>
</tr>
<tr>
<td>SHSPEC **</td>
<td>-26.876</td>
<td>13.281</td>
<td>-2.024</td>
</tr>
<tr>
<td>PAEXPCT</td>
<td>26.696</td>
<td>79.371</td>
<td>0.336</td>
</tr>
<tr>
<td>MAEXPCT</td>
<td>8.581</td>
<td>78.103</td>
<td>0.110</td>
</tr>
<tr>
<td>FAMSIZ *</td>
<td>-53.042</td>
<td>30.870</td>
<td>-1.718</td>
</tr>
<tr>
<td>CATCOST *</td>
<td>49.160</td>
<td>29.158</td>
<td>1.686</td>
</tr>
<tr>
<td>RELCOST</td>
<td>9.582</td>
<td>14.164</td>
<td>0.677</td>
</tr>
<tr>
<td>PRIVCOST</td>
<td>-54.607</td>
<td>26.976</td>
<td>-2.024</td>
</tr>
<tr>
<td>INTERCEPT</td>
<td>1275.438</td>
<td>1809.609</td>
<td>0.705</td>
</tr>
</tbody>
</table>

(Continued)

Degrees of Freedom: 39
Number of observations used in calculations: 346
F Value: 2.498
* significant at the 10% level
** significant at the 5% level
Table 8: (continued)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHSALARY</td>
<td>-0.010</td>
<td>0.015</td>
<td>-0.668</td>
</tr>
<tr>
<td>S8SALARY</td>
<td>-11.161</td>
<td>34.642</td>
<td>-0.322</td>
</tr>
<tr>
<td>SHRATIO</td>
<td>47.405</td>
<td>87.968</td>
<td>0.539</td>
</tr>
<tr>
<td>SESQ</td>
<td>58.394</td>
<td>48.092</td>
<td>1.214</td>
</tr>
<tr>
<td>SINGLE **</td>
<td>-297.338</td>
<td>141.828</td>
<td>-2.096</td>
</tr>
<tr>
<td>MILLS **</td>
<td>301.155</td>
<td>117.836</td>
<td>2.556</td>
</tr>
</tbody>
</table>

Degrees of Freedom: 39
Number of observations used in calculations: 346
F Value: 2.498
* significant at the 10% level
** significant at the 5% level

Student Characteristics

Race

The reading gain score regression results for both Catholic and public high schools indicate that non-minority students achieve higher reading levels than do the Hispanic and African American students. Among both ethnic groups, the average performance of students in Catholic high schools, as well as their average reading gain scores, exceed that of their peers in the public schools. There are two trends in the data which might help explain the lower reading scores of the Hispanic and African American students relative to the average student attending Catholic high schools. First, the socioeconomic status of Hispanic and African American students was lower than that of majority students in both the Catholic and public high schools. This result is consistent with other research, which has shown repeatedly that children’s achievement levels are associated with their parents’ educational attainment, a factor included in the variable representing the family's
socioeconomic status. A second and more dramatic trend lies in the differences in the proportion of minorities enrolled in the grade schools from which the students graduated. For example, white students attending Catholic high schools graduated from grade schools where the average minority enrollment made up about 9.2% of the total student enrollment. The African and Hispanic American students attended grade schools with average minority enrollments making up 70% and 50% of the total student enrollment respectively. Although the regression results do not allow one to draw firm conclusions regarding causality, the lower socioeconomic status of the Hispanic and African American students relative to their nonminority counterparts, combined with the higher minority enrollments in the grade schools they attended, might account for part of the differences in reading achievement outcomes.

School Characteristics

High School Tracking

Although most of the grade school characteristics included in the regression analysis were not significantly related to student gain scores, two Catholic high school characteristics were. The first characteristic of interest, student enrollment in a vocational track was significantly negatively related to reading gain scores. Since fundamental skills required for higher order reading levels are acquired in grade school, it is plausible that students self-select into vocational or academic tracks based on information received from grade school teachers, in the form of grades, concerning their natural abilities. Grade school reading scores may proxy for natural ability and lead students of higher ability to self-select into academic high school tracks. Others of lesser ability might sort themselves into vocational or other tracks. Evidence for this is found in the different grade school composites for students enrolled in the vocational and academic tracks. Catholic high school stu-

---

udents in the vocational track had eighth grade score composites 22% lower than those enrolled in the academic track. Furthermore, students enrolled in the vocational track reported spending less time doing homework, 23% less than the average for those in the academic track. Lower student ability, proxied by lower eighth grade composites, and less time spent in doing homework might account in part for the lower student reading gain scores for those enrolled in Catholic high school vocational tracks.

Peer Group Characteristics

*Special Education Programs*

The regression results of the present study indicate that the proportion of Catholic high school students enrolled in special education programs is significantly negatively related to student gain scores in reading. Similar negative relationships are found in the gain score regressions for science and history as well. This particular result is somewhat surprising given the relative effectiveness of the public schools in promoting overall student achievement through student enrollment in special education programs. According to the public school gain score regressions, the proportion of students enrolled in special education programs was significantly positively related with student gain scores in mathematics, science, and history. The ineffectiveness of Catholic high school special education programs relative to those found in the public school system might be related to the type of student each system was designed to serve.

Historically, public schools have been forced to accept virtually all students residing in their geographical boundaries. Issues concerning the demand for educational equity for the academically challenged students required the formation of programs and the hiring of properly trained teachers to meet the needs of these special students. The student gain score regression results would seem to indicate that the public schools have been relatively successful in this regard. On the other hand,
Catholic high schools have traditionally exercised the right to either exclude students who did not meet certain academic standards or send them to the public schools. This normally occurred for two reasons. First, the financial constraints of the Catholic secondary schools did not always allow for the establishment of special education programs or the full-time faculty needed to staff them. Second, the purpose of Catholic secondary education was to educate students in the Catholic faith and to prepare them for college. Students with special needs were often, though not always, encouraged to attend the local public high schools designed to meet their special needs.

Over the last twenty years, the system of Catholic high schools within the United States has experienced a severe shortage of priests, religious sisters, and brothers who worked for a fraction of the salaries paid to lay (non-religious) teachers. The average twelfth grade tuition rates for Catholic high schools has nearly doubled over the last ten years, and tuition income only covers approximately 73% of total student expenditures. In order to meet their growing operating expenses, increasing numbers of students with special needs were admitted and programs designed to meet those needs were created. There is evidence, however, that when special education programs are not properly designed or implemented, they can breed low expectations of students by teachers, principals, parents, and the students themselves. The reading gain score regression may be picking up the effects of increasing enrollments of students in need of special education programs in the Catholic high schools.

**High School Graduates**

The second peer group category to consider is the percentage of recent Catholic high school graduates (1990-1991 school year) enrolled in technical and four year colleges. In both cases the coefficients of the variables are significantly positively related to the mathematics gain

---

scores of students enrolled in high school. Despite the limitations of standardized tests as well as the danger that teachers could “teach to the test”, the argument that tests do not measure performance of any kind is not supported by the results of this regression analysis. Rather we find a strong positive relationship between a school’s graduates placed in technical and four-year colleges and its current students’ gain scores on standardized reading tests. It is plausible that the higher order reading skills, learned at the high school level and reflected in student gain scores, provide the rigorous training indispensable to the application of logic and the scientific method necessary for success in technical and other four year colleges. This relationship may be more indicative of what the school contributes to a student’s human capital or training than to what the older students contribute to their younger peers.

Mills Ratio

The first part of this model included a probit regression for choice of Catholic school versus all other school types, i.e. public, private, and NAIS. If there is a statistically significant self-selection into the Catholic high school sector based on unmeasured factors, this model postulates a relation between the residual of the school sector choice equation and the residual of the achievement equation. Such a correlation seems to exist since the Mills Ratio, which corrects for self-selection, is significantly positively related to the student reading gain scores at the 5 percent level. The Mills Ratio is also significantly related to the student science and history gain scores and will be included as an independent variable in those regression equations.

Change in Catholic High School Student Science Test Scores

Having reviewed the gain score regression results for mathematics and reading, we now turn to the third part of the model concerning the student science gain score regression. The change

---

in standardized science scores are calculated by subtracting student eighth grade scores from the same student’s twelfth grade score. Calculated student science gain scores were then regressed on the Mills Ratio and other student, school, teacher, peer, and family background characteristics. A stepwise regression technique was used to delete variables which proved to be insignificant in earlier science gain score regression results. The removal of individual student characteristics from the science score regression did not affect the significance of the remaining variables. Since the Mills ratio was significantly positively related to student science gain scores, it was included in the regression to correct for selectivity bias due to unmeasured factors related to school choice and student achievement.

**Educational Production Function For Catholic School Students, Science Gain Score.**

\[
\Delta S_i = \alpha + \beta X_i + \omega C_i^i + \mu_i
\]

\(\Delta S_i\) = Calculated student science gain score between the 8th and 12th grades.
\(\alpha, \beta, \omega\) = Regression coefficients.
\(X_i\) = Vector of student background characteristics except for Catholic school choice.
\(C_i^i\) = The Mills Ratio for Catholic school choice.
\(\mu_i\) = residual

**Student Characteristics**

**Gender**

Similar to the results of the mathematics gain score regression, the science gain scores were significantly negatively related to being female. The lack of sufficient mathematics background, the difference in early science experiences, and the subsequent change in attitude toward science by females may be based in cultural and social factors engendered in the school by way of student/teacher interactions or the lack of them.
Table 9: Catholic High School Student Science Gain Score Regression Coefficients, Standard Errors, and t-ratios by Specification.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>-14.429</td>
<td>61.469</td>
<td>-0.235</td>
</tr>
<tr>
<td>FEMALE **</td>
<td>-128.296</td>
<td>53.657</td>
<td>-2.391</td>
</tr>
<tr>
<td>BLACK</td>
<td>-80.717</td>
<td>136.974</td>
<td>-0.589</td>
</tr>
<tr>
<td>HISP</td>
<td>50.590</td>
<td>86.310</td>
<td>0.586</td>
</tr>
<tr>
<td>P8GRADES</td>
<td>1.055</td>
<td>4.649</td>
<td>0.227</td>
</tr>
<tr>
<td>P8STRACK</td>
<td>-20.361</td>
<td>70.096</td>
<td>-0.290</td>
</tr>
<tr>
<td>P8ALGEBR</td>
<td>24.167</td>
<td>45.550</td>
<td>0.531</td>
</tr>
<tr>
<td>PNOHISCHL</td>
<td>356.192</td>
<td>438.511</td>
<td>0.812</td>
</tr>
<tr>
<td>POTHER</td>
<td>-162.430</td>
<td>149.601</td>
<td>-1.086</td>
</tr>
<tr>
<td>PCATH</td>
<td>78.362</td>
<td>73.147</td>
<td>1.071</td>
</tr>
<tr>
<td>S8PROB</td>
<td>-1.704</td>
<td>2.813</td>
<td>-0.606</td>
</tr>
<tr>
<td>S8ENROL</td>
<td>-20.449</td>
<td>35.307</td>
<td>-0.579</td>
</tr>
<tr>
<td>SHENROL</td>
<td>-0.010</td>
<td>0.087</td>
<td>-0.114</td>
</tr>
<tr>
<td>SH2ACADM **</td>
<td>133.592</td>
<td>60.524</td>
<td>2.207</td>
</tr>
<tr>
<td>SH2SUBRB</td>
<td>-85.879</td>
<td>71.981</td>
<td>-1.193</td>
</tr>
<tr>
<td>SH2PROB **</td>
<td>-14.590</td>
<td>7.098</td>
<td>-2.055</td>
</tr>
<tr>
<td>PH2HMWK</td>
<td>-13.118</td>
<td>13.049</td>
<td>-1.006</td>
</tr>
<tr>
<td>GRTECH *</td>
<td>110.348</td>
<td>60.806</td>
<td>1.815</td>
</tr>
<tr>
<td>SHGEN</td>
<td>2.050</td>
<td>2.832</td>
<td>0.724</td>
</tr>
<tr>
<td>SHPREP</td>
<td>1.298</td>
<td>1.633</td>
<td>0.795</td>
</tr>
<tr>
<td>SHSPEC **</td>
<td>-22.226</td>
<td>8.587</td>
<td>-2.588</td>
</tr>
<tr>
<td>SHSALARY</td>
<td>-0.002</td>
<td>0.011</td>
<td>-0.225</td>
</tr>
<tr>
<td>S8SALARY **</td>
<td>44.396</td>
<td>22.202</td>
<td>2.000</td>
</tr>
<tr>
<td>PAEXPCT</td>
<td>34.959</td>
<td>49.227</td>
<td>0.710</td>
</tr>
<tr>
<td>MAEXPCT</td>
<td>-4.646</td>
<td>48.245</td>
<td>0.096</td>
</tr>
<tr>
<td>SESQ *</td>
<td>52.525</td>
<td>31.004</td>
<td>1.694</td>
</tr>
<tr>
<td>FAMSIZ</td>
<td>1.443</td>
<td>19.796</td>
<td>0.073</td>
</tr>
<tr>
<td>SINGLE</td>
<td>-24.864</td>
<td>90.553</td>
<td>-0.275</td>
</tr>
<tr>
<td>CATCOST</td>
<td>14.934</td>
<td>18.627</td>
<td>-0.802</td>
</tr>
<tr>
<td>RELCOST</td>
<td>-6.078</td>
<td>9.083</td>
<td>-0.669</td>
</tr>
<tr>
<td>PRIVCOST</td>
<td>12.106</td>
<td>15.423</td>
<td>0.785</td>
</tr>
<tr>
<td>MILLS *</td>
<td>155.238</td>
<td>83.174</td>
<td>1.866</td>
</tr>
<tr>
<td>INTERCEPT</td>
<td>14.452</td>
<td>1153.84</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Degrees of Freedom: 33
Number of observations used in calculations: 347
F Value: 1.578
* significant at the 10% level
** significant at the 5% level
Race, Eighth Grade Aspirations, Religion

Student characteristics regarding eighth grade college aspirations, ethnic or religious background, and previous academic achievement, and religious background were not significantly related to student science gain scores.

School Characteristics

Grade School Characteristics

Most of the grade school characteristics included in the science gain score regression turned out to be insignificantly related to student achievement. For example, the location, type, and size of the grade school, the percent of the student body made up of minorities, student-teacher ratio, base year teacher salaries, and special programs for the gifted were all insignificantly related to student science gain scores.

Tracking

Similar to the Catholic and high school regression results for student mathematics gain scores, academic track is significantly positively related to student gain scores in science. There was no significant relationship between student science gain scores and enrollment in other educational tracks offered by the Catholic schools. The lack of any significant differences between the various tracks offered indicates the presence of a more homogeneous group of students in the Catholic high schools.

Peer Group Characteristics

Similar to the student reading gain score regression results, the proportion of Catholic high school students enrolled in special education programs is significantly negatively related to student gain scores in science. This result is also contrary to the public school regression results which indicate a significant positive relation between the proportion of students enrolled in special educa-
tion programs and overall student gain scores in science. Given the relatively large gain scores of students who were traditionally low achievers academically, this particular result is difficult to explain. It is plausible that students with special academic needs are better served by the public school system which has the programs and personnel required to maintain effective special education programs.

**High School Graduates**

The second peer group category to consider is the percentage of recent Catholic high school graduates (1990-1991 school year) enrolled in technical colleges. The coefficient of this variable is significantly positively related to student science gain scores. It is plausible that the higher order reading skills, learned at the high school level and reflected in the Catholic school student gain scores, provide the rigorous training indispensable to the application of logic and the scientific method necessary for success in technical and other four year colleges.

**Change in Catholic High School Student History Test Scores**

This part of the model regresses the change in student standardized test scores in social studies (history) between the eighth and twelfth grades. Student gain scores were calculated by subtracting a student’s eighth grade score from the same student’s twelfth grade score in social studies. The calculated gain scores were then regressed on a set of independent variables representing the student’s personal, family, peer, teacher and school characteristics. The results of the regression analysis are presented in Table 10.

**Educational Production Function For Catholic School Students, History Gain Score.**

(8.4) \[ \Delta H_i = \alpha + \beta X_i + \omega C_i^t + \mu_i \]

\[ \Delta H_i \] = Calculated student history gain score between the 8th and 12th grades.
\[ \alpha, \beta, \omega \] = Regression coefficients.
\[ X_i \] = Vector of student background characteristics except for Catholic school choice.
\[ C_i^t \] = The Mills Ratio for Catholic school choice. \[ \mu_i \] = residual
Table 10: Catholic High School Student History Gain Score Regression Coefficient, Standard Errors, and t-ratios by Specification.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>-45.601</td>
<td>50.792</td>
<td>-0.898</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-14.882</td>
<td>43.765</td>
<td>-0.340</td>
</tr>
<tr>
<td>BLACK</td>
<td>30.190</td>
<td>99.330</td>
<td>0.304</td>
</tr>
<tr>
<td>HISP</td>
<td>67.117</td>
<td>67.020</td>
<td>1.001</td>
</tr>
<tr>
<td>ASIAN *</td>
<td>157.803</td>
<td>94.371</td>
<td>1.672</td>
</tr>
<tr>
<td>PCATH</td>
<td>37.745</td>
<td>60.840</td>
<td>0.620</td>
</tr>
<tr>
<td>POTHER *</td>
<td>-252.064</td>
<td>142.003</td>
<td>-1.775</td>
</tr>
<tr>
<td>PNONE **</td>
<td>-293.312</td>
<td>132.650</td>
<td>-2.111</td>
</tr>
<tr>
<td>PCOLL</td>
<td>36.839</td>
<td>80.491</td>
<td>0.458</td>
</tr>
<tr>
<td>S8PRIVAT</td>
<td>-292.296</td>
<td>192.438</td>
<td>-1.519</td>
</tr>
<tr>
<td>P8HTRACK</td>
<td>-21.761</td>
<td>55.792</td>
<td>-0.390</td>
</tr>
<tr>
<td>S8PROB</td>
<td>-0.719</td>
<td>2.218</td>
<td>-0.324</td>
</tr>
<tr>
<td>S8ENROL **</td>
<td>82.303</td>
<td>31.749</td>
<td>2.592</td>
</tr>
<tr>
<td>SHENROL</td>
<td>0.060</td>
<td>0.064</td>
<td>0.939</td>
</tr>
<tr>
<td>PH2HMWK</td>
<td>4.360</td>
<td>10.424</td>
<td>0.418</td>
</tr>
<tr>
<td>SH2SUBRB *</td>
<td>-116.000</td>
<td>60.444</td>
<td>-1.919</td>
</tr>
<tr>
<td>SH2ACADM</td>
<td>45.356</td>
<td>48.911</td>
<td>0.927</td>
</tr>
<tr>
<td>SH2VOTEC</td>
<td>38.955</td>
<td>185.480</td>
<td>0.210</td>
</tr>
<tr>
<td>SH2OTHER</td>
<td>11.823</td>
<td>116.762</td>
<td>0.101</td>
</tr>
<tr>
<td>SH2PROB **</td>
<td>-16.520</td>
<td>6.007</td>
<td>-2.750</td>
</tr>
<tr>
<td>GROTHER</td>
<td>11.422</td>
<td>42.669</td>
<td>0.268</td>
</tr>
<tr>
<td>SHSPEC **</td>
<td>-15.422</td>
<td>6.345</td>
<td>-2.431</td>
</tr>
<tr>
<td>FAMSIZ</td>
<td>-0.741</td>
<td>15.670</td>
<td>-0.047</td>
</tr>
<tr>
<td>SHSALARY</td>
<td>-0.006</td>
<td>0.007</td>
<td>-0.788</td>
</tr>
<tr>
<td>CATCOST</td>
<td>6.663</td>
<td>12.913</td>
<td>0.516</td>
</tr>
<tr>
<td>RELCOST *</td>
<td>13.596</td>
<td>7.459</td>
<td>1.823</td>
</tr>
<tr>
<td>PRIVCOST *</td>
<td>-21.439</td>
<td>12.293</td>
<td>-1.744</td>
</tr>
<tr>
<td>SHSALARY</td>
<td>30.506</td>
<td>27.697</td>
<td>1.101</td>
</tr>
<tr>
<td>SEQ</td>
<td>17.489</td>
<td>24.330</td>
<td>0.719</td>
</tr>
<tr>
<td>SINGLE</td>
<td>-32.877</td>
<td>77.761</td>
<td>-0.423</td>
</tr>
<tr>
<td>MILLS **</td>
<td>137.004</td>
<td>64.365</td>
<td>2.129</td>
</tr>
<tr>
<td>INTERCEPT</td>
<td>1273.57</td>
<td>928.094</td>
<td>1.372</td>
</tr>
</tbody>
</table>

Degrees of Freedom: 30  
Number of observations used in calculations: 366  
F Value: 2.295  
* Significant at the 10% level  
** Significant at the 5% level
Student Characteristics

Age, Gender, Race, Religion, College Aspirations.

Since the regression results for student standardized gain scores in history are consistent with the previous results in mathematics, reading, and science, summaries of the probable causes rather than detailed discussions, relating to the more significant results will be provided. The variables representing the student’s age, gender, and race (except for the Asian students) were not significantly related to student gain scores in history. The variable representing a student with no particular religious background was negatively associated with history gain scores, and was significant at the 5 percent level.

School Characteristics

Grade School Variables

Earlier regression results for mathematics, reading, and science indicated that the type of grade school, public, Catholic, or NAIS, was not significantly related to student gain scores. However, the regression results for the history gain scores revealed a negative relationship to private non-religious grade school attendance and was significant at the 10% level. This negative relation reflects the fact that students who graduated from private non-religious grade schools received the highest eighth grade and twelfth grade scores in the history standardized tests, but gained the least in history standardized scores during their four years in Catholic high school. The reasons for the lower gain scores in history are unclear but would warrant further research.
CHAPTER IX

CONCLUDING REMARKS

Public and Catholic Probit Regression Results Compared.

In this study a two-equation model was constructed to study the relationship between student achievement in the public and Catholic school sectors. The first part of the model uses a probit equation to study how the choice of Catholic or public high school is related to various grade school and student background characteristics. The probit regression first run using the choice of public high school as the dependent variable and then run a second time using the choice of Catholic high school as the dependent variable. The results of each probit regression were used to construct two school choice variables designed to correct for the possibility of selectivity bias in the second part of the model. The second part of the model includes four equations which regress separately student gain scores in mathematics, science, reading, and history on a set of student variables and the Mills Ratio. This was done separately for students in the public and Catholic high schools. The student gain score regression results free of selectivity bias are then used to gain insight into the student, family, school, and other background characteristics which are significantly related to student achievement in each type of high school.

Probit Regression Results.

Single-adult Households

The probit regression results indicated a significant negative relationship between the choice of public high school and the variable representing students from single-adult households. One
would expect that the financial constraints of single-adult families could lead them to choose the less expensive public school option for their children. However, the findings of this study indicate that there might be other factors which account for their choice of the private school option. Future studies might include variables related to parental time constraints, the availability of private school student services, and the presence of alternative family income which could account in part for the choice of private sector high schools.

*Prior Academic Achievement*

The data on student achievement provided by the results of this survey indicate that Catholic high school students scored significantly higher on their standardized tests than did students in public high schools. It has often been contended that high achieving students choose to attend Catholic high schools. In this study, when choice of high school is regressed on student eighth grade composite scores (a measure of previous achievement), there is no significant relationship between the choice of Catholic high school and prior grade school achievement. There is, however, a significant positive relationship between prior student achievement and the choice of attending a public high school. These results suggest that high achievers select the public high school option. Factors other than prior achievement would seem to be responsible for the academic performance of Catholic high school students.

*Race*

The probit regression results indicate that Hispanic and Asian students tend to choose the public school option more frequently than African American students who tend to choose the Catholic high school option.

*Socioeconomic Status*
Socioeconomic status is significantly positively related to choice of Catholic high school and significantly negatively related to the choice of public high school. This variable is constructed using the parents’ education, income, and job type. This student family characteristic might be related to the differences in academic performance in each type of high school.

*Grade School Type*

The probit regression results indicate that students who choose to attend a public high school are most likely graduates of public grade schools. Similarly, students who choose the Catholic high school option are most likely graduates of private religious and non-religious grade schools.

*Future Academic Expectations*

The results of this model indicate no significant relationship between public school choice and the future academic expectations of the student’s mother or father. There is, however, a significant positive relationship between the choice of Catholic high school and the student’s mother’s expectations. This outcome would seem to indicate that the expectations of the student’s mother might play an important role in the choice of Catholic high school.

*High School Tuition*

The public high school probit regression results indicate no significant relationship between the choice of public high school and the tuition costs of Catholic or other private religious high schools. However, there is a significant negative relationship between the choice of public high school and the tuition costs of alternative non-religious private high schools. This reasons for this outcome are not certain and would warrant further research.

*Student Gain Score Regression Results*

*Gender*
The regression results indicate that females in public and Catholic high schools scored significantly lower in mathematics and science than their male counterparts in each sector. This result is consistent with earlier studies concerning gender related issues in student achievement. An interesting fact of this study, however, is that female Catholic high school students had higher standardized scores and larger gain scores in mathematics than the male students attending public high schools. Public school male students earned higher gain scores and standardized scores in science than their female counterparts in both sectors. It is plausible that specific school type characteristics might be responsible for the superior achievement of Catholic high school female students in mathematics and science. Further research is needed to identify which school and teacher characteristics are significantly related to their academic achievement in mathematics and science.

Race

African American students in the public high schools had significantly lower gain scores in mathematics, science, and reading than their non-minority peers. Hispanic students earned significantly lower gain scores in reading and science than the average for non-minority students. Catholic high school African and Hispanic Americans scored significantly lower than the average for the sample in reading only. Furthermore, both minorities attending Catholic high schools, while scoring below the average for the sample, scored closer to the sample mean than did their public school peers.

Student Eighth Grade Score Composites

The regression results for public high school student eighth grade score composites were significantly positively related to student gain scores in mathematics and science. If these score composites proxy for student ability and/or motivation, then it is not surprising to find that the more able students outperform their peers in science and mathematics. The students attending
Catholic high schools had similar results for mathematics. However, their reading gain scores were significantly negatively related to their eighth grade score composites. This indicates that students who were initially low achievers in grade school gained significantly more than the average for the sample in reading skills.

*Advanced Mathematics in Grade School*

The regression results for public high school students indicated that eighth grade students who took an introductory course in Algebra I had larger gain scores in mathematics than the average student in the sample. On the other hand, eighth grade students who were enrolled in other types of accelerated mathematics courses scored well above the average in the first round of standardized mathematics tests but gained very little during their high school careers. These results imply that an introduction to Algebra I in the eighth grade has a significant impact on the improvement of student mathematics skills during high school. However, there is also evidence which indicates that public high schools add more to the mathematics skills of the average student than to the students of above average ability in mathematics. Teachers and administrators could explore the creation of new programs designed to help students with advanced skills in mathematics to achieve at greater potential.

*Academic Track, Homework*

The regression results for students in public high schools indicate that that enrollment in an academic track and hours spent doing homework are both significantly positively related to student gain scores in mathematics, science, and history. Parents, teachers, and school administrators might improve student academic skills by encouraging students to enroll in high school academic tracks and teachers to assign more homework.

*High School Problems*
The only public and Catholic high school characteristic significantly negatively related to student gain scores in all four subjects is the presence of violence in and around the school. This is a problem consistently associated with low student achievement and needs to be addressed if progress is to be made in reducing truancy, dropouts, and enhancing students’ desire to attend school.

Minorities

It has often been held that the presence of minorities is detrimental to the academic achievement of the non-minority student. No significant relationship between minority enrollment and academic gain scores for students in public or Catholic high schools was found in this study. What is found however, is that the students’ experience (or perception) of violence in and around the school vicinity is significantly related to low student gain scores. Policy makers concerned with efforts to increase the equality of educational opportunity for minorities and the quality of education that they receive need to focus less on policies involving desegregation in itself and more on eliminating the violence permeating many of our high school campuses.

Special Education Programs

One very interesting result of this study concerns the significant positive relationship between student gain scores and enrollment in public high school special education programs. Student gain scores in mathematics, science, and history were larger for those enrolled in special education tracks than those enrolled in general or academic tracks. This result is consistent with those of previous studies which show that the test scores of many disabled or non-English-speaking students enrolled in special education programs have dramatically improved in recent years. Furthermore, the regression results for public high school students show no evidence that help for these students has a negative effect on the academic progress of the average student. In fact, the results of this study indicate a significant positive relationship between the number of students enrolled in a high
school’s special education program and overall student achievement in math, science, and history. The positive results of this analysis should be given serious consideration when educational pro-
grams are cut in efforts to control the escalating costs of high school education.

Concluding Remarks

Academics and policy makers have been involved in a decade long debate over the relative effectiveness of public and Catholic high schools. Much of this debate has been waged over a single outcome measure: standardized test scores. In this paper, student gain scores in mathematics, reading, science, and history were regressed separately on a list of background variables for stu-
dents in the Catholic and public school sectors. One important result is that after controlling for selectivity bias and other significant student, school, and family characteristics, Catholic high school students still outperform their public high school peers.

This study leaves open a number of questions. First, it is possible that important variables were omitted from the analysis which could account for the difference in achievement outcomes. Second, if Catholic schools are more effective than public schools, more research regarding the source of their effectiveness is needed. Coleman attributed this success to Catholic school’s empha-
sis on discipline, attendance, and homework. When these variables were controlled in the gain score regression equations, the differences persisted. Finally, if there is something to be learned from Catholic high schools, we need to know if it will ever be possible to apply these lessons to public high schools. In some ways, Catholic high schools are like other private institutions; they must meet the test of the market. In other ways they are obviously fundamentally different, and it is still not clear that they succeed because of the importance of religion or the discipline of competi-
tion.
LIST OF REFERENCES


