This is to certify that the

dissertation entitled

Three Essays on the Economics of Education

presented by

Douglas N. Harris

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Economics

Major professor

Date 11/5/2000

THREE ESSAYS ON THE ECONOMICS OF EDUCATION

By

Douglas N. Harris

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

2000

ABSTRACT

THREE ESSAYS ON THE ECONOMICS OF EDUCATION

By

Douglas N. Harris

The three chapters of this dissertation focus on two of the fundamental propositions that are currently driving education policy: (a) that market incentives can be used to improve school efficiency without causing adverse social consequences; and (b) that the level and allocation of school resources is not optimal. The wide range of research results related to these ideas is partially explained by methodological problems. Therefore, the evidence presented here relates both to the propositions and to methodological issues that apply more broadly.

Chapter 1: "Charter School Location." The intention of charter school programs, and school choice programs more generally, is to use market mechanisms to improve school efficiency, innovation, and program variety. This study provides evidence on these issues using data on Michigan and California. The number of charter schools in each school district is regressed on the characteristics of students and public schools in those same districts. The results indicate that more charter schools locate where populations are diverse in terms of race, income, and adult education levels. This is interpreted as demand for horizontal differentiation, reflecting both preferences for homogeneous student populations and preferences for specific education programs. In both states, charter school location is negatively related with public school test scores.

This implies that parents pay attention to test scores (vertical differentiation), even though these scores may be imperfect signals for school performance.

Chapter 2: "New Approaches to Meta-Analysis with Applications to Education Production Functions." Meta–analysis is potentially useful in reconciling differences in results across studies. However, common approaches make untested, and potentially false, assumptions about the role of methodology in explaining these differences. In the case of education production functions (EPF), the observed methodological differences are found to explain 47-57 percent of the variation in parameter estimates, i.e., "methodology matters" when trying to understand whether money matters in education. In addition, the experimental evidence relating to EPF parameters is shown to be quite robust across studies, providing useful information for identifying gold standard econometric techniques.

Chapter 3: "Optimal School Inputs." The level of real resources going to U.S. schools has tripled since 1960, and increasing portions are going toward smaller class sizes. It has been difficult to evaluate these reforms because the effects that school inputs have on student outcomes have been imprecisely estimated. Chapter 2 suggests that school input effects are positive and magnitudes fall within a relatively narrow range. However, there is little, if any, evidence about whether these new estimates are large enough to justify recent policy reforms. Here, new estimates are used in a calibrated partial equilibrium model to estimate the socially optimal level and allocation of resources. The results suggest that increases in human capital from current school inputs are not sufficient to justify the current level of resources. In addition, the portion of resources going toward teacher salaries is too small relative to class size reductions.

Dedicated with love to my father who taught me attention to detail and devotion to public service, to my mother who taught me what is really important in life, and to my brothers who taught me to work hard for the right reasons.

ACKNOWLEDGMENTS

I thank my advisor, Professor Gerhard Glomm, for his effort and guidance in my development as an economist and researcher. He went beyond the call of an advisor, reading dozens of drafts of each paper and providing careful comments about both the details and the big picture. In addition, he allowed me to go beyond the usual student role, working as a co-author and collaborator on several research projects. Through our many debates, he challenged me to ask the right questions and use sound reasoning in every aspect of my work. In the process, we became good friends.

I thank David Plank of the MSU College of Education for helping me incorporate ideas from a different discipline, education, that is the main subject of my dissertation. In addition, he showed me how to bridge the gap in communicating economic ideas to educators and others who are not normally, but certainly need to be, part of the audience.

I thank Ron Fisher for our many conversations early in my graduate training, which provided the ideas that developed into my second and third chapters. I also thank Jeffrey Wooldridge for many contributions to many of the more challenging econometric issues in first two chapters.

The above four faculty comprised my dissertation committee, which proved to be as helpful as it was diverse. In addition to my committee members, Scott Adams and David Neumark provided particularly useful comments and suggestions.

In my four and a half years at MSU, I had the benefit of interacting with almost everyone in the Department of Economics, from our softball and basketball games to my ongoing job search. Therefore, I thank all of the department's faculty, graduate students, and support staff.

TABLE OF CONTENTS

LIST OF TABLES viii

LIST OF FIGURES xi

CHAPTER 1: CHARTER SCHOOL LOCATION 1

- I. Introduction 1
- II. An Informal Theory 5
- III. Data and Methodology 8
- IV. Results 16
- V. Conclusion 33
- Appendix A: Summary Statistics (Michigan) 35
- Appendix B: Summary Statistics (California) 36
- Appendix C: The Educational Production Frontier Function (Michigan) 37
- Bibliography 39

CHAPTER 2: NEW APPROACHES TO META-ANALYSIS WITH APPLICATIONS TO EDUCATION PRODUCTION FUNCTIONS 43

- I. Introduction 45
- II. Introduction to Meta-Regression 48
- III. Application: Education Production Functions 57
- IV. Results: Meta-Regression 63
- V. Identification of an EPF Econometric Gold Standard 71
- VI. Conclusion 87
- Appendix A: Statistical Properties of MRE 90
- Appendix B: Profiles of Gold Standard Estimates 93
- Bibliography 95

CHAPTER 3: OPTIMAL SCHOOL INPUTS 99

- I. Introduction 99
- II. Models 103
- III. Parameters 113
- IV. Results 124
- V. Conclusion 128
- Appendix A: Sensitivity Analysis 132
- Bibliography 134

LIST OF TABLES

CHAPTER 1: CHARTER SCHOOL LOCATION

Table 1: Charter School Growth in Michigan 2
Table 2a: Mean of Independent Variables Categorized by Number of Charter Schools in Michigan School Districts 11
Table 2b: Mean of Independent Variables Categorized by Number of Charter Schools in California School Districts 12
Table 2b (cont'd) 13
Table 3a: Regression Results for Michigan 19
Table 3a (cont'd) 20
Table 3b: Regression Results for California 21
Table 3b (cont'd) 22
Table 4: Regression Results With Sample Restrictions (Michigan) 30
Table 5: Regression Results with Contiguous Districts (Michigan) 32
Table 6a: Summary Statistics (Michigan) 35
Table 6b: Summary Statistics (California) 36
Table 7: The Educational Production Frontier Function (Michigan) 37

CHAPTER 2: NEW APPROACHES TO META-ANALYSIS WITH APPLICATIONS TO EDUCATION PRODUCTION FUNCTIONS

Table 1: Signs and Significance Levels of Coefficients from Three EPF Samples 45
Table 2: Methods in the New Samples of EPF Studies 62
Table 3: Meta-Regression with Class Size-Test Score Coefficients 65
Table 3: (con't) 66
Table 4: Meta-Regression with Class Size-Wage Coefficients 68
Table 4: (con't) 69
Table 5: Effect Sizes by Study and Sample Characteritics 73
Table 6: Methods of Gold Standard Versus Other Estimates 79
Table 7: Statistics from Gold Standrd Versus Other Estimates 82

Table 7: (con't) ... 83
Table 8: Comparing Class Size Point Estimates from Different Meta-Analyeses ... 85
Table 9: Profiles of Gold Standard Estimates ... 93

CHAPTER 3: OPTIMAL SCHOOL INPUTS

Table 1: School Resource Allocation in U.S. States, 1960 and 1990 ... 100
Table 2: Summary of EPF Estimates from Chapter 2 ... 114
Table 3: Transformed EPF Estimates ... 116
Table 4: Salaries in Teaching and Various Professions, 1998 ... 118
Table 5: Summary of Teacher Supply Elasticities ... 120
Table 6: Summary of Parameter Values ... 123
Table 7: Optimal Inputs and Required Parameters (Unconstrained Maximization) ... 125
Table 8: Optimal Input Results (Maximization with Resource Constraints) ... 126
Table 9: Inputs Levels and Costs for Test Score Gains ... 127
Table 7a: Sensitivity Analysis for Optimal Input Levels ... 132

LIST OF FIGURES

Figure 1: Horizontal and Vertical Differentiation of Schools 7

CHAPTER 1

Charter School Location [1]

I. Introduction

Before 1990, choice in the American education system was largely limited to assigned public schools and private schools. Since then, there has been a significant restructuring. Many states have allowed public school choice, allowing students to attend the public schools outside the districts in which they reside. Voucher plans in Milwaukee, Cleveland, San Antonio, and Florida have further expanded these choices to include private schools. These small-scale programs appear to be precursors to widespread choice programs, such as those proposed in Michigan, California, and other states.

A third instrument of providing greater school choice is the charter school system. Charter schools are publicly financed and often subject to less regulation than traditional public schools. Some oversight is usually administered by third parties, including universities or state government agencies, rather than local school boards. Charters receive a fixed amount per student enrolled. However, in contrast to public schools, they do not receive a separate allotment for capital expenditures.[2] In addition, operating

---

[1] The author thanks participants in association meetings of the AEA, Econometric Society, and Public Choice, as well as seminars at the Federal Reserve Bank of Chicago, Michigan State University, the University of Michigan, and the University of Kentucky. We especially thank Bih-Shiow Chen, Julie Cullen, Tom Downes, David Figlio, Larry Kenny, Bob Rasche, Peter Schmidt, and John Strauss for helpful comments. Financial support for this project from the Business College at MSU is gratefully acknowledged.

[2] Most public school districts can raise capital funds through local property tax levies. Arizona is a slight exception to the rule for charter policies, allowing charter start-up grants up to $100,000. However, this is quite small compared with total required capital costs for most schools.

revenues in charter schools are sometimes as low as 50 percent of neighborhood public schools with an average of about also 80 percent (Finn, Manno, and Vanourek, 2000). Therefore, total spending per student is usually much lower in charter schools compared with nearby public schools.

The charter school movement is the most rapidly developing part of the U.S. school choice movement (Paris, 1998). By the end of 1994, eleven states had adopted charter programs. By 1999, thirty-three states had charter policies, yielding over 1,700 charter schools and 350,000 enrolled students (Finn et al, 2000). In Michigan, the charter school growth was similarly rapid, as indicated in table 1.

Table 1: Charter School Growth in Michigan

Number of Charter Schools in District	Year				
	94-95	95-96	96-97	97-98	98-99
0	545	527	508	495	487
1	10	21	35	43	47
2	0	3	6	9	12
3	0	2	3	4	4
4	0	1	1	1	2
5	0	0	1	2	2
7	0	1	0	0	0
13	0	0	1	0	0
21	0	0	0	1	0
36	0	0	0	0	1
Total charter schools in MI	10	44	78	108	137

Perhaps the most compelling argument for school choice in general, and charter schools in particular, is that it will improve schooling by increased reliance on market mechanisms (Friedman, 1962). One argument here is that students will "vote with their feet" for a better alternative school if charter school productivity is not sufficiently high. The market thus imposes discipline on the charter schools. Only good schools will survive. Peterson (1999) finds some support for this view in his review of research on recent, small-scale voucher programs. However, Bettinger (2000) finds that charter schools are not any more productive than public schools.

A second argument is that charter schools may improve productivity in the traditional public schools since they stand to lose enrollment, and hence state funding, to the charter schools. If this efficiency argument is valid, we expect to find a negative relationship between the number of charter schools and public school performance as measured by productive efficiency within each school district. Hoxby (1994, 1997) finds some evidence for this hypothesis in her studies of competition among private schools and public school districts.

A third aspect of efficiency relates to the Tiebout hypothesis, which roughly states that people will receive their optimal bundles of government-provided goods if there are many options to choose from and free mobility across jurisdictions. These bundles include the level of local taxation, private school tuition, school performance (vertical differentiation), the type of education being provided (horizontal differentiation), and other amenities. In the case of education, the less perfect is the sorting, the more likely it is that parents will seek other bundles when given alternatives.

There is anecdotal evidence that school choice produces schools with specialized curricula and homogeneous student bodies within schools. For instance, the El-Hajj Malik EL-Shabazz Academy in Lansing, Michigan describes its mission as to "serve students using a holistic, Afrocentric curriculum." More generally, 70 percent of all charter school students in Michigan are minority compared with 22 percent of public school enrollment.[3] In California, Grutzik, et al (1995) show that "the communities surrounding charter schools are primarily white and have income levels at or above the city and county averages." Wells et. al. (1996) comes to a similar conclusion, as do other survey-based studies. These outcomes are apparently similar in other countries that have implemented expansive school choice programs.[4]

In this paper, we address the question whether charter schools do indeed locate in those districts where the lack of Tiebout sorting has provided for inefficient outcomes. What are the criteria that charter schools use to make their location decisions? Why might the demand for these schools be greater in some districts than others? For a given demand, what factors influence the supply of charter schools? Do charter schools indeed enter in districts where there are few existing alternatives and imperfect Tiebout sorting?

To help answer these questions, we assemble data for all charter schools in Michigan and California, and match each charter school with the public school districts in which they locate. We then regress the number of charter schools in a district on the student and public school characteristics of those same districts. The econometric

[3] These results come from a study commissioned by the State of Michigan: Public Sector Consultants and MAXIMUS (1999). The results are similar in Arizona, where the numbers are 6 percent and 4.3 percent respectively (Gifford, Ogle and Solmon, 1998).
[4] England, Chile, China, Sweden, New Zealand, South Africa, and the Czech Republic are a few examples. See Harris, Oliver, and Plank (2000) for discussion of these

methodology we use is similar to the one used by Downes and Greenstein (1996), using a statistical model of count data advanced by Hausman, Hall and Griliches (1984) and Cameron and Trivedi (1986).

Downes and Greenstein study entry (location) by privately funded schools in California. Bettinger (2000) and Filer and Munich (2000) estimate regressions similar in form to those estimated here. The Bettinger paper also relates to charter schools in Michigan, but his main purpose is to obtain instruments for other regressions that study the impact of charter schools on public school performance. Most of the variables we find to be important are excluded in Bettinger's regressions. Filer and Munich's specification is more similar to those used here, but they study new choice-based schools in the Czech Republic. These authors also exclude a large number of variables that we find to be important, but they do find that charter school location is negatively related with a measure of public school efficiency.

In section II, we provide an informal theory of charter school location based on product differentiation. We describe our data in section III. Empirical tests of our hypotheses are reported and discussed in section IV.

II. An Informal Theory

The evidence above suggests that charter school entry is closely related to product differentiation, including both horizontal (h) and vertical (v) dimensions. Common examples of vertical differentiation (quality) in education include graduation rates, value-added, and the proportion of kids going on to college.

---

countries. For a description of differences across U.S. states, see Wohlstetter et. al (1995), Nathan (1996), and Mintrom (1998).

In most markets, inputs would not be considered an aspect of quality because more productive inputs are reflected in better outputs and higher prices. However, there are many reasons to believe that the price-quality relationship is rather weak in education. First, dependence on the tax system means that non-consumers are paying most of the cost. Second, the movement toward state level funding may be weakening the connection between local taxation and local school funding. Third, information problems may prevent parents from being able to observe actual school quality. These three facts imply that it may be reasonable to interpret input levels as signals of public school quality. We consider student-teacher ratios, expenditure per student, teacher salaries, and special education expenditures.

There are also various dimensions of horizontal product differentiation. These are dimensions along which preference heterogeneity generates disagreement among consumers over what is best. Some consumers may prefer an academic curriculum in high school, while others favor more vocational training. Other consumers may favor schools with racial and ethnic diversity, while others may favor homogeneity. Some favor authoritarian schooling by Catholic nuns, while others favor schooling in which children are free to determine their own rules of conduct.

Ideally, the location of schools ought to be considered as the outcome of a location game such as in Hotelling (1929) or Prescott and Visscher (1977). In these papers, location is an outcome of a game played by profit-maximizing firms. Unfortunately, the objective functions for public and private schools are controversial.[5] Instead of fully specifying such a location game and characterizing the equilibrium location, we only illustrate potential outcomes of such games.

In figure 1, product differentiation is shown in *h-v* product space. Household preferences are distributed over the *h-v* space. The small open circle represents the most preferred bundle of an individual, assuming that person has to pay for the full cost of education. Small closed circles represent schools of which there are three types: public schools U_i, private schools P_i, and charter schools C_i.

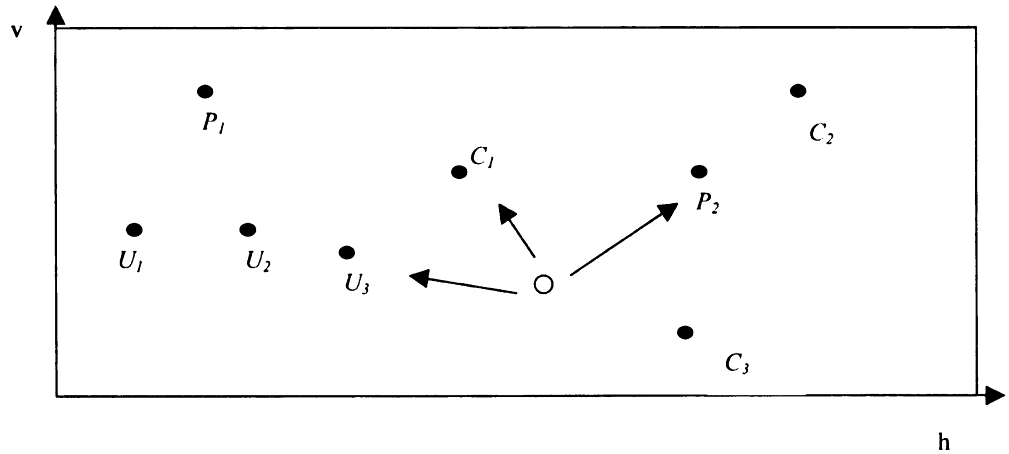

Figure 1: Horizontal and Vertical Differentiation of Schools

The parent illustrated in figure 1 can choose any school in the space, but C_1, P_2, and U_3 most closely match the parent's preferences. If all schools charged the same price, then we would expect the person to locate at the nearest school in the *h-v* space. However, the actual funding system includes tuition only at private schools. Therefore, if P_2 charged high tuition, this option would be eliminated from consideration, leaving C_1 and U_3, which would provide similar satisfaction at a much lower price.

While the above discussion is framed in the language of horizontal product differentiation, we will carry out our econometric work using various measures of taste

[5] See, for example, Manski (1992) and Nechyba (1996).

heterogeneity. Actual measures of product differentiation, such as the amount of time spent teaching foreign languages, are not available. Therefore, we assume that variation in tastes will give rise to variation in education programs.

Public school education programs are chosen by school boards and superintendents for entire districts, therefore, we expect relatively minor variations in product differentiation across public schools within a district. Other models in which public and private schools co-exist, such as the one studied by Epple and Romano (1996), predict that private school quality is higher than public school quality. In addition, horizontal differentiation among the private schools is much greater than among the public schools, providing for the possibility of elite schools with high high-tuition and religious schools with low tuition.

As in Prescott and Visscher (1977), we assume that a charter school needs to attract a sufficient number of customers in order to cover fixed costs. It appears in figure 1 that this is best accomplished by locating away from public and private schools in the *h-v* product space. However, if only private schools charge tuition, then charter schools may seek to provide education similar to private schools, but free of charge to consumers. Therefore, charter schools may instead try to locate very close to private schools in the product space. In any case, the location of charter schools in the *h-v* space is dependent upon the characteristics of both public and private schools in the district.

III. Data and Methodology

In this section we describe our data and the hypothesized relationships between the independent variables and charter school location. All of our data describing public school districts comes from the Departments of Education in Michigan and California. The data on charter schools in Michigan come from the Michigan Association of Public School Academies (MAPSA). The data on California charter schools comes from the California Department of Education. The data on demographic variables comes from the School District Data Book (SDDB), which includes U.S. Census Bureau data organized by school district.

The Michigan sample includes more than 500 observations in each regression specification, out of a total of 555 districts in the state. None of the missing district observations contains a charter school. However, there are several weaknesses of the California data. First, many school districts in California are separated into primary and secondary school districts, which overlap one another, yet charter schools are assigned to only one district. This means that the same school is located in multiple districts, yet we can only match each charter school to the district that granted the charter.

A second weakness is that charters schools may be located outside of the district that granted the charter. A large majority of the districts appear to be located in their chartering districts, therefore, this may not significantly impact the results.[6] A third weakness is that some charter schools serve home schooling populations that may or may not reside near the school. These last two weaknesses imply that the independent variables we choose may not closely accurately describe the market conditions existing near the charter schools.[7]

---

[6] This statement is based on informal discussions with researchers in the California Department of Education.

[7] Our samples in California are in the range of 288-350 district observations, which is substantially lower than the total number of districts. Approximately 60 of the 142 districts that contain charter schools are among the list of missing observations, though

The unit of observation is the school district. This approach is appropriate because school policies are set at the school district level. In addition, it is difficult to obtain data at the school level. We use the following variables: 1) wealth/income as measured by median household income and the poverty rate; 2) ethnic composition of the population; 3) adult (parent) educational attainment; 4) geographic characteristics, such as area and the enrollment density of the district; 5) performance of the public schools measured by outputs, such as test scores, graduation rates, and productive efficiency; 6) public school inputs, such as student-teacher ratios; 7) charter school revenue (state grants) and costs (teacher salaries); 8) the degree of competition from private schools.[8] Summary statistics are provided in appendices A and B. In tables 2a and 2b below, we exhibit public school characteristics for Michigan and California by the number of charter schools per district.

---

none of the missing observations contain more than one charter school. All missing observations for both states are due to missing data in the original sources listed above.

[8] At least three of the variables discussed above are measured with error: teacher salaries, graduation rates, and the number of private schools. We tried various symmetric truncations, but found that the results were unaffected.

Table 2a: Mean of Independent Variables Categorized by Number of Charter Schools in Michigan School Districts (1998)

NUMBER OF CHARTER SCHOOLS IN A DISTRICT (Observations)	0 (487)	1 (47)	2 (12)	3 (4)	4 (2)	5 (2)	36 (1)
Med. House Value (thous.)	54.44	62.66	53.14	59.43	77.15	53.00	25.29
Per Capita Income (thous.)	12.43	13.69	12.84	15.72	14.01	12.21	9.44
Med. Household Income (thous.)	29.23	31.65	27.78	33.38	26.71	26.75	18.74
Perc. of Children in Poverty	14.87	13.10	21.06	16.40	24.33	25.20	44.29
Perc. of Children Black	2.63	4.94	16.59	34.95	30.45	25.94	82.09
Perc. of Children Hispanic	2.22	2.83	4.97	1.97	8.38	9.67	3.28
Herfindahl Index for Race	88.60	85.62	74.49	56.26	47.84	44.93	69.29
Perc. of Adult with 12 Grade	23.66	20.98	23.17	19.33	17.82	21.73	37.15
Perc. of Adult with High School	38.10	35.72	30.19	26.28	23.02	26.96	28.36
Perc. of Adult with Some College	26.21	28.52	30.53	31.54	27.91	32.79	25.58
Perc. of Adult ≥ Bach. Deg.	12.03	14.78	16.11	22.84	31.25	18.52	8.92
Average Years of Schooling	13.01	13.25	13.35	13.81	14.25	13.55	12.68
Herfindahl Index for Education	30.50	29.26	28.22	27.30	34.26	26.29	29.18
MEAP Math Score 4th Grade	43.26	44.59	38.89	34.03	42.55	30.30	27.30
Graduation Rate	90.23	80.85	83.63	65.45	69.65	34.60	71.60
Expend. Per Student (thous.)	4.30	4.55	4.99	6.36	5.95	5.07	5.29
Special Educ. Exp./Stud. (thous.)	0.32	0.36	0.56	0.44	0.52	0.728	0.61
Pupil-Teacher Ratio	20.19	19.60	20.83	17.75	18.50	21.50	22.00
Avg. Teacher Salary (thous.)	30.51	33.18	35.13	39.14	34.06	31.75	36.24
Students Per Sq. Kilometer	51.64	45.51	122.77	189.77	114.65	227.77	510.17
Number of Private Schools	1.43	1.96	5.50	10.50	14.00	23.00	98.00
Total K-12 Enrollment (thous.)	2.29	3.72	7.57	10.16	14.65	26.49	183.15
Grad. Rate of K-12 Enrollment	-0.11	0.54	-0.98	0.41	-0.67	-2.17	-0.57
Perc. of Public Schools in Cities	3.32	10.59	36.44	25.00	98.33	69.44	94.60

Table 2b: Mean of Independent Variables by Number of Charter Schools in California School Districts (1998)

NUMBER OF CHARTER SCHOOLS IN A DISTRICT (Observations)	0 (683)	1 (73)	2 (11)	3 (7)	4 (2)	5 (2)
Med. House Value (thous.)	161.0	155.6	193.5	166.1	195.0	143.9
Per Capita Income (thous.)	15.92	15.17	16.01	15.02	16.84	14.07
Med. Household Inc. (thous.)	35.84	33.16	36.48	34.92	35.66	35.36
Perc. of Children in Poverty	16.91	15.65	12.64	17.44	13.03	10.55
Perc. of Children Black	2.91	5.00	3.48	6.71	1.87	2.41
Perc. of Children Hispanic	28.98	24.33	19.17	26.96	31.98	34.03
Herfindahl Index for Race	60.15	56.22	56.43	59.51	58.69	51.71
Perc. Of Adult with 12 Grade	27.03	24.03	17.52	24.61	21.10	23.45
Perc. Of Adult with High School	25.02	25.55	25.04	25.46	21.60	28.47
Perc. of Adult with Some College	30.07	33.16	36.85	32.78	34.18	34.12
Perc. of Adult with Bachelor Degree or Higher	18.06	17.47	20.74	17.31	23.22	14.13
Avg. Years of Schooling	13.44	13.59	13.82	13.42	13.96	13.37
Herfindahl Index for Education	31.65	29.54	28.22	28.42	27.34	27.30
MEAP Math Score 4th Grade	618.9	615.3	611.5	614.2	618.7	616.9
Graduation Rate	1.41	2.39	1.61	3.45	3.60	1.93
Total Expend. Per Stud. (thous.)	18.02	42.83	37.69	33.33	147.0	37.80
Special Educ. Exp/Stud. (thous.)	0.10	0.11	0.107	0.10	0.11	0.12
Pupil-Teacher Ratio	22.92	23.55	24.96	25.47	23.43	23.71
Average Teacher Sal. (thous.)	25.11	24.87	24.54	26.23	23.90	25.11
K-12 Enrollment Density	81.59	87.82	41.03	67.32	146.2	57.52
Number of Private Schools	2.36	5.62	4.67	4.75	26.06	45.04
Total K-12 Enroll. (thous.)	3.65	8.91	7.10	6.94	21.14	27.65
Gr. Rate of K-12 Enrollment	1.92	2.09	3.22	6.03	1.77	3.39
Perc. of Public School in Cities	13.22	20.21	27.92	18.54	0.64	6.33

Table 2b (cont'd)

NUMBER OF CHARTER SCHOOLS IN A DISTRICT (Observations)	6 (2)	11 (1)	12 (1)	14 (1)	34 (1)
Med. House Value (thous.)	81.52	87.09	172.1	183.2	226.5
Per Capita Income (thous.)	11.64	10.38	14.77	16.26	15.34
Med. Household Inc. (thous.)	26.55	22.77	27.13	32.05	30.85
Perc. of Children in Poverty	29.91	32.04	29.71	21.04	27.47
Perc. of Children Black	4.99	2.73	51.36	14.13	13.71
Perc. of Children Hispanic	39.20	10.72	17.84	28.52	59.20
Herfindahl Index for Race	38.90	76.23	34.13	29.13	41.08
Perc. of Adult with 12 Grade	31.48	19.06	25.51	17.04	35.64
Perc. of Adult with Diploma	24.95	20.48	21.01	21.58	19.72
Perc. of Adult – Some College	29.12	39.76	27.90	35.07	25.14
Perc. of Adult ≥ Bachelor Deg.	14.74	21.04	25.68	26.40	19.71
Avg. Years of Schooling	13.15	13.98	13.81	14.11	13.33
Herfindahl Index for Education	26.86	27.90	25.30	26.83	26.75
MEAP Math Score 4th Grade	614.1	629.9	594.0	616.1	601.3
Graduation Rate	5.85	0.00	8.80	4.40	13.28
Total Expend. Per Stud. (thous.)	158.6	2.51	231.5	571.3	3756.8
Special Ed. Exp./Stud. (thous.)	0.12	0.20	0.11	0.12	0.11
Pupil-Teacher Ratio	24.82	16.22	22.61	23.24	24.56
Average Teacher Sal. (thous.)	25.43	23.91	27.46	24.93	28.70
K-12 Enrollment Density	203.82	68.23	350.9	243.4	430.8
Number of Private Schools	16.51	1.00	54.00	94.0	593.0
Total K-12 Enroll. (thous.)	38.84	0.37	51.25	125.1	639.9
Gr. Rate of K-12 Enrollment	1.44	8.03	0.40	1.70	1.23
Perc. Public Schools in Cities	36.57	0.00	94.47	100.0	79.74

The empirical analysis involves regressing the number of charter schools in each district on the characteristics of the schools and students in those same districts. Therefore, the dependent variable occurs in non-negative integer amounts and OLS is inconsistent. One of the methods created to deal with such issues is Poisson regression, developed by Hausman, Hall and Griliches (1984) and Cameron and Trivedi (1986). Poisson regressions have also been used by Papke (1991) in the context of manufacturing firm start-up and by Downes and Greenstein (1996) in the context of private school start-ups in California. Two of the potential weaknesses of this approach are: 1) violation of the independence assumption; and 2) "overdispersion" in the data. We use the Huber-White robust standard errors to correct for dispersion. [See Huber (1967), White (1982).]

There are two potential simultaneity problems in this framework. Charter schools, private schools, and public schools of choice are substitutes for each other. Entry of charter schools in a district is determined simultaneously with entry of private schools and/or public schools of choice within the same district. Moreover, public schools may respond to charter school entry by changing their behavior. We deal with both these problems by regressing the number of charter schools in 1998 on school district characteristics in 1992, the year before charter school policies began.

The second simultaneity issue stems from the fact that students can cross district boundaries to attend charter schools.[9] This means that the number of charter schools in district i , C_i, depends not only on the characteristics of that district, X_i , but also on the

[9] Kelejian and Prucha (1998) state that "cross-sectional spatial models frequently contain a spatial lag of the dependent variable as a regressor or a disturbance term that is *spatially autocorrelated.* The first of these topics is discussed here. We assume there is no spatial autocorrelation in our model.

characteristics of neighboring districts, X_j, including the number of charter schools, C_j. This yields

$$C_i = f(X_i, X_j, C_j) \quad . \qquad (1)$$

District j may be a composite of information for many districts because there may be multiple districts nearby. This definition is important because it defines market size, or the geographic area over which it is possible to attract students. We start by excluding neighboring district data altogether, focusing only on home district characteristics. Next, we substitute in for C_j in (1) to obtain a reduced form that is a function only of X_i and X_j.

In most cases where nearby districts are included, composite variables are created that account both for the number of students and physical distance of the border districts relative to the home district. For variables that are hypothesized to be positively (negatively) related to the number of charter schools, increasing the distance and decreasing the proportion of students in the border districts is expected to decrease (increase) the composite variable.

A common econometric issue is identifying the simultaneous equations of supply and demand. The usual simple model is not appropriate in this context because the price is exogenously fixed by the government, rather than being endogenously determined by markets. This yields two equations, but only one endogenous variable: the number of charter schools in a district. Therefore, we combine them into a single equation, which can be estimated consistently without additional changes to the estimation procedure.[10]

---

[10] Consider the following structural equations: and $C_d = \gamma_1 h + \gamma_2 v + \varepsilon_d$ and $C_s = \beta_1 p + \varepsilon_s$. The demand for charter schools C_d is a function of the variation in preferences for various

IV. Results

In this section, specific variables are introduced that relate to various aspects of horizontal and vertical product differentiation. We hypothesize specific relationships and interpret these results for both Michigan and California, which are presented below in tables 4 through 7.

There are two reasons to expect that the results will be different across the two states. Both relate to state education policy. First, California education spending on traditional public schools is significantly constrained in wealthier school district to be below actual desired levels. (See, for example, Fernandez and Rogerson, 1999). Michigan also limits spending in high-income districts; however, the limits appear to allow wealthier districts to come closer to their desired spending levels. In addition, Michigan has recently redistributed substantial funding to low-income districts, resulting in a substantially higher average spending level (Papke, 2000).

A second key policy difference is that California's charter school policy limits chartering authority to local school districts, implying that charters cannot start without some support from the public schools. In Michigan, school districts may serve as chartering authorities, but this permission is also extended to universities and other organizations outside the district in which the charter schools reside.[11] The possible implications of these policy differences for our results are discussed below.

---

horizontal characteristics h, the strength of preferences for quality v, and a disturbance ε_d. The supply of charter schools C_s is a function the price of inputs and output, and a disturbance ε_s. Again, prices do not show up on the demand side as they usually would because price is fixed at zero. After imposing the equilibrium condition $C = C_d = C_s$, it now appears that we have two separate equations trying to explain the same phenomenon. Therefore, we instead estimate $C = \pi_1 h + \pi_2 v + \pi_3 p + \varepsilon$.

[11] The vast majority of schools are authorized by universities.

The demand for charter schooling is related to the size of the market. We measure the size of the market by the number of children enrolled in public schools or by the number of school-age children in the school district (ages 5-17). We would expect the number of charter schools to be positively related with this measure of "market-size," other things being equal. If entry decisions by charter schools are forward looking, trends in market size might matter as well. We therefore include the growth rate of enrollment as an independent variable. Many districts have been growing at a rapid rate, especially in California. In addition, the fact that only school districts can authorize charters in California implies that they are most likely to occur in growing districts that are adding schools. Building new charter schools, instead of traditional schools, allows the district to expand while decreasing regulatory burdens.

All of these hypotheses about market size are supported by the results in table 3a (Michigan) and 3b (California). The coefficient on the number of students is consistently positive and significant in both states. Enrollment growth is positive and significant in California, but insignificant in Michigan. This provides support for the impact of differences in state charter policy.

We use various measures of income.[12] Assuming education is a normal good, we might expect both median family income and median house value to be positively related with charter school entry.[13] This is especially true in California where state equalization policy has constrained public school spending in high-income districts. On the other

[12] Chambers (1999) calculates school district-level cost indexes using a hedonic wage model. Many variables in our regressions are denominated in dollars, however, deflating them has very little impact on the results. All reported results are not deflated.
[13] The assumption that education is a normal good relates only to the vertical dimension of education characteristics. We have no theory about the relationship between income and any horizontal characteristic.

hand, low-income households may have fewer opportunities to move their residence, implying a negative relationship between income and charter entry. In other words, low-income households may demand less of a normal good, but they may also be further away from their most preferred bundle.

We use both median household income and median property values, but the results are unaffected by this choice. We use household income and find that the coefficient is negative for both states. It is occasionally significant for Michigan, and uniformly significant for California. The negative coefficients contradict the evidence for California from Grutzik et al (1995), who finds a positive relationship between income and charter location. However, Grutzik excludes adult education levels, which is highly correlated with income. The magnitude of the effect for California (table 3b, column 7) suggests that a one standard deviation increase in median income ($7,781) decreases the expected number of charter schools by 67 percent. In Michigan, the magnitude of this effect is smaller.

Table 3a: Regression Results for Michigan
Dependent Variable: Number of Charter Schools Per Michigan School District, 1998-1999
(Robust standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)
Med. Household Income	-0.0329 (0.0123) **	-0.0182 (0.0182)	-0.0248 (0.0182)	-0.0413 (0.0199) **	-0.0276 (0.0212)
Percentage of Children in Poverty	--	0.0059 (0.0253)	-0.0004 (0.0241)	-0.0067 (0.0246)	-0.0028 (0.0247)
Average Yrs. of Schooling	1.1902 (0.1918) **	1.0257 (0.2398) **	0.9872 (0.2455) **	0.9954 (0.2438) **	0.9729 (0.2594) **
Herf. Index for Ad. Education	--	-0.1248 (0.0487) **	-0.1233 (0.0472) **	-0.0960 (0.0505) *	-0.1084 (0.0503) **
Herf. Index for Race	--	-0.0326 (0.0108) **	-0.0312 (0.0106) **	-0.0265 (0.0111) **	-0.0249 (0.0112) **
MEAP Math Score for 4th Grade	-0.0431 (0.0121) **	-0.0190 (0.0103) *	-0.0149 (0.0120)	-0.0176 (0.0128)	-0.0161 (0.0125)
Productive Efficiency	--	--	-3.9482 (4.5992)	-3.8725 (5.0275)	-3.9039 (4.8826)
Graduation Rate	-0.0065 (0.0044)	-0.0026 (0.0032)	-0.0024 (0.0030)	-0.0024 (0.0029)	-0.0021 (0.0029)
Tot. Exp. Per Student	0.1997 (0.0948) **	0.0326 (0.0899)	0.0886 (0.0931)	-0.0745 (0.1295)	0.0485 (0.1160)
Avg. Special Educational Expenditure Per Student	--	--	--	1.5656 (0.8082) *	1.5624 (0.8056) *
Pupil-Teacher Ratio	--	--	--	--	-0.0151 (0.0363)
Avg. Teacher Salary	--	--	--	0.0469 (0.0260) *	--
Expenditure minus Foundation Grant	--	--	--	--	--
Number of Priv. Schools	--	--	--	--	--
Total Enrollment	0.0267 (0.0010) **	0.0263 (0.0029) **	0.0274 (0.0030) **	0.0254 (0.0033) **	0.0264 (0.0032) **
Gr. Rate of Enrollment	-0.0379 (0.0608)	-0.0196 (0.0622)	-0.0201 (0.0632)	0.0202 (0.0700)	-0.0003 (0.0750)
City	--	--	--	--	--
Constant	-15.1547 (1.9957) **	-7.7136 (3.6071) **	-3.8374 (5.4191)	-5.8819 (5.6816)	-4.6103 (5.7761)

Table 3a (cont'd)

	(6)	(7)	(8)	(9)	(10)
Med. Household Income	-0.0303 (0.0214)	-0.0374 (0.0213) *	-0.0208 (0.0188)	-0.0231 (0.0179)	-0.0134 (0.0175)
Percentage of Children in Poverty	-0.0042 (0.0243)	-0.0050 (0.0242)	-0.0001 (0.0242)	-0.0151 (0.0260)	-0.0064 (0.0245)
Average Yrs. of Schooling	0.9721 (0.2538) **	0.9543 (0.2624) **	1.0215 (0.2465) **	0.9105 (0.2430) **	0.7771 (0.2552) **
Herf. Index for Ad. Education	-0.1126 (0.0509) **	-0.1025 (0.0519) **	-0.1183 (0.0471) **	-0.1188 (0.0352) **	-0.1150 (0.0340) **
Herf. Index for Race	-0.0340 (0.0108) **	-0.0244 (0.0102) **	-0.0319 (0.0099) **	-0.0265 (0.0099) **	-0.0232 (0.0094) **
MEAP Math Score for 4th Grade	-0.0173 (0.0122)	-0.0174 (0.0127)	-0.0145 (0.0120)	-0.0148 (0.0121)	-0.0115 (0.0121)
Productive Efficiency	-3.3967 (4.8807)	-3.7854 (4.8789)	-3.9480 (4.6431)	-2.4928 (5.3160)	-1.6921 (5.4030)
Graduation Rate	-0.0027 (0.0030)	-0.0022 (0.0028)	-0.0029 (0.0031)	-0.0038 (0.0037)	-0.0012 (0.0022)
Tot. Exp. Per Student	-0.1275 (0.1596)	--	--	--	--
Avg. Special Educational Expenditure Per Student	--	1.5336 (0.8081) *	--	--	--
Pupil-Teacher Ratio	-0.0372 (0.0418)	-0.0175 (0.0310)	--	--	--
Avg. Teacher Salary	0.0526 (0.0268) **	0.0371 (0.0193) *	--	--	--
Expenditure minus Foundation Grant	--	--	0.1853 (0.2596)	0.2440 (0.2502)	0.3354 (0.2573)
Number of Priv. Schools	--	--	--	--	0.0461 (0.0054) **
Total Enrollment	0.0273 (0.0031) **	0.0255 (0.0033) **	0.0276 (0.0029) **	0.0253 (0.0030) **	--
Gr. Rate of Enrollment	-0.0436 (0.0768)	-0.0020 (0.0772)	-0.0141 (0.0632)	-0.0096 (0.0662)	-0.0157 (0.0689)
City	--	--	--	0.0093 (0.0031) **	0.0091 (0.0033) **
Constant	-3.8675 (5.4394)	-5.2281 (5.7439)	-3.9424 (5.3989)	-3.9508 (5.6821)	-4.0476 (5.7591)

Note: One asterisk (*) indicates 90 percent significance.
Two asterisks (**) indicates 95 percent significance.

Table 3b: Regression Results for California
Dependent Variable: Number of Charter Schools Per California School District, 1998-99
(Robust standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)
Med. Household Income	-0.0861 (0.0300) **	-0.0739 (0.0212) **	-0.0731 (0.0232) **	-0.0695 (0.0240) **	-0.0702 (0.0245) **
Percentage of Children in Poverty		0.0127 (0.0234)	0.0274 (0.0200)	0.0329 (0.0210)	0.0259 (0.0214)
Average Yrs. of Schooling	0.8558 (0.2737) **	0.7799 (0.3307) **	0.9871 (0.3707) **	0.9842 (0.4839) **	0.9486 (0.4117) **
Herf. Index for Ad. Education		-0.2144 (0.0698) **	-0.2137 (0.0646) **	-0.2031 (0.0784) **	-0.2159 (0.0728) **
Alternative Herf. Index for Race		0.0131 (0.0082)	0.0193 (0.0111) *	0.0146 (0.0112)	0.0196 (0.0113) *
STAR Math Score for 4th Grade	-0.0052 (0.0158)	-0.0083 (0.0177)	-0.0078 (0.0185)	-0.0071 (0.0184)	-0.0087 (0.0181)
Productive Efficiency			-3.3665 (1.1353) **	-2.9579 (3.7429)	-3.1755 (3.4613)
Dropout Rate	0.0176 (0.0328)	-0.0026 (0.0379)	-0.0042 (0.0370)	0.0038 (0.0350)	-0.0039 (0.0358)
Tot. Exp. Per Student	-0.0170 (0.0014) **	-0.0163 (0.0022) **	-0.0150 (0.0025) **	-0.0138 (0.0031) **	-0.0153 (0.0032) **
Avg. Special Educational Expenditure Per Student				-1.0158 (8.3636)	-0.6271 (7.1063)
Pupil-Teacher Ratio					-0.0262 (0.0671)
Min. Teacher Salary				-0.1276 (0.0969)	
Expenditure minus Foundation Grant					
Number of Priv. Schools					
Total Enrollment	0.1076 (0.0089) **	0.1028 (0.0131) **	0.0960 (0.0151) **	0.0894 (0.0182) **	0.0976 (0.0190) **
Gr. Rate of Enrollment	0.1276 (0.0557) **	0.1444 (0.0504) **	0.1325 (0.0486) **	0.1363 (0.0473) **	0.1297 (0.0482) **
City					
Constant	-7.4269 (7.9253)	0.4003 (9.4383)	-0.6781 (8.9617)	1.6724 (9.7679)	0.8964 (9.9190)

Table 3b (cont'd)

	(6)	(7)	(8)	(9)	(10)
Med. Household Income	-0.0648 (0.0260) **	-0.0903 (0.0301) **	-0.0995 (0.0292) **	-0.0997 (0.0278) **	-0.0864 (0.0301) **
Percentage of Children in Poverty	0.0277 (0.0194)	0.0549 (0.0176) **	0.0191 (0.0195)	0.0191 (0.0194)	0.0330 (0.0191) *
Average Yrs. of Schooling	0.8936 (0.4198) **	1.4456 (0.4045) **	0.9530 (0.3109) **	0.9575 (0.2931) **	0.8956 (0.3394) **
Herf. Index for Ad. Education	-0.1994 (0.0656) **	-0.2582 (0.0697) **	-0.1864 (0.0565) **	-0.1865 (0.0568) **	-0.2127 (0.0572) **
Alternative Herf. Index for Race	0.0149 (0.0112)	-0.0045 (0.0138)	0.0018 (0.0151)	0.0017 (0.0143)	-0.0005 (0.0158)
STAR Math Score for 4th Grade	-0.0095 (0.0206)	0.0036 (0.0184)	0.0100 (0.0175)	0.0100 (0.0178)	0.0091 (0.0179)
Productive Efficiency	-1.7487 (1.7257)	-6.8301 (3.8641) *	-3.6933 (1.1975) **	-3.6840 (1.2445) **	-3.3845 (1.2546) **
Dropout Rate	0.0059 (0.0376)	0.0271 (0.0446)	-0.0703 (0.0607)	-0.0704 (0.0602)	-0.0105 (0.0498)
Tot. Exp. Per Student	-0.0146 (0.0028) **				
Avg. Special Educational Expenditure Per Student		-8.4139 (6.9365)			
Pupil-Teacher Ratio	-0.0452 (0.0671)	0.0619 (0.0644)			
Min. Teacher Salary	-0.1329 (0.0982)	-0.1528 (0.0800) *			
Expenditure minus Foundation Grant			-0.1037 (0.0204) **	-0.1044 (0.0205) **	-0.0574 (0.0243) **
Number of Priv. Schools					0.0237 (0.0071) **
Total Enrollment	0.0939 (0.0167) **	0.0080 (0.0012) **	0.0357 (0.0062) **	0.0359 (0.0060) **	
Gr. Rate of Enrollment	0.1316 (0.0459) **	0.1408 (0.0477) **	0.1228 (0.0434) **	0.1228 (0.0434) **	0.1278 (0.0443) **
City				-0.0003 (0.0042)	0.0032 (0.0049)
Constant	4.3352 (10.7343)	-5.0846 (10.4561)	-9.6069 (9.1573)	-9.6076 (9.1597)	-8.3214 (9.3238)

Note: One asterisk (*) indicates 90 percent significance.
Two asterisks (**) indicates 95 percent significance.

We include education of the adult population as an indicator of family preferences. Parents with higher education levels might receive greater utility from the education quality of their children, making them more eager to choose a charter school if traditional public schools perform below par. The measures we use are fractions of the population with: (i) less and high school degree; (ii) high school degrees only, (iii) some college, and (iv) college degree. From these, we calculate a measure of average parent education level. The coefficient on this variable is consistently positive and significant in both states, which is consistent with the above hypothesis and the results of Downes and Greenstein (1996). The magnitude for Michigan (table 3-1, column 7) suggests that a one standard deviation increase in average years of schooling (0.6 years) raises the expected number of charter schools by 61 percent.

There are at least three ways in which population diversity may be associated with parents' demand for charter schools. First, diversity may imply greater dispersion of preferences and student needs. Second, parents may desire schools whose students have characteristics similar to those of their own children. Third, the median voter model suggests that if any group has greater than 50 percent of the population, then this group may be able to control school policy through voting. Populations with less than 50 percent of the population may, therefore, seek to open schools that more closely match their preferences.

We include three measures of population diversity, based on race, income, and education of the adult population. Herfindahl indices are used both for race and education. A higher (lower) Herfindahl index implies a more (less) homogeneous population. Therefore, if diversity leads to more charter schools, then the coefficients on these variables should be negative. This hypothesis is generally supported by tables 3a and 3b, except that racial mix does not appear to be a factor in California. One reason for this may be the more restrictive charter-granting policies in California. If minority groups cannot affect school district policies, then these groups can seek charters on their own in Michigan. In California, they must work through the same school districts that have apparently already failed them. In Michigan, the magnitude of the effect suggests that a one standard deviation increase in either Herfindahl index yields a 29 percent decrease in the number of charter schools.

Available data does not allow for easy calculations of a Herfindahl index for income at the school district level. As an alternative, we include a measure of lower tail of the income distribution, in addition to median household income. Controlling for median income, a larger portion of the population in poverty implies a less equal income distribution. The coefficient on poverty in tables 3a and 3b are generally positive, as expected, but they are often insignificant.

The geographic size of the district may be important due to transportation costs. The number of students per square mile, controlling for the number of schools, indicates the average distance to school, which we would expect to be negatively related to charter school entry. Similarly, the number of public schools is expected to be negatively related with charter school location, since a large number of public schools would be associated with lower transportation costs and greater horizontal differentiation, controlling for the number of students.[14] It turns out that these variables are consistently insignificant,

[14] It may also be the case that people prefer schools with few students.

therefore, we omit them from the results reported here.[15] However, whether the schools are in cities, versus suburbs and rural areas, is important. The coefficient on the proportion of schools located in a central city is insignificant for California, but positive and significant in Michigan, regardless of whether density is included.

The intent of creating charter schools was to increase choice and competition in schooling. We expect charter schools to enter more frequently where school choice is limited. A larger number of private schools implies a larger number of substitutes to public schooling and, therefore, a higher degree of competition. We might expect then that the number of private schools (and public schools) will be negatively related with charter school entry.[16] However, charter schools could also locate near private schools and provide a similar type of education without charging tuition. In this case, charter school location might be an increasing function of the size of the private sector.[17] Our estimation results provide a test of which effect is dominant. The results in table 3-1 and 3-2 indicate that more private schools are associated with more charter schools, providing support for the second hypothesis.

The quality of public schools might be an important determinant of charter school entry. If public school quality is low, dissatisfied parents might be more inclined to choose the charter school alternative. Two of our measures of public school quality are measures of output: graduation rates and student test scores.[18] Survey research suggests

[15] They were also highly correlated, therefore, we did not include both in any given specification.
[16] Cross-district schools-of-choice programs, which were described earlier, also measure choice, however, these do not have a significant impact.
[17] Unfortunately, we do not have data on private school tuition and other private school characteristics.
[18] For Michigan, the student test is the 4th grade Michigan Educational Assessment Program (MEAP). For California, the student test is the 4th grade Standardized Testing

that when test scores are low, parents might consider alternatives to public education and we therefore expect these scores to be negatively related with charter school entry (Finn, 2000). The negative point estimates here support this hypothesis, but they are generally not significant.

Student and family characteristics vary widely across districts. Previous research indicates that these differences have an important impact on education outcomes.[19] Therefore, test scores and drop out rates alone will not accurately indicate the contribution to education made by the schools themselves. The ideal measure of school quality (vertical differentiation) is value-added. We calculate a measure of public school efficiency using frontier regressions of educational production functions in which the dependent variable is 7th grade math scores.[20] The estimated functions are shown in appendix C. The best any district can do is produce on the production frontier. The further the actual test score is from the production frontier, the more inefficient is the district. If markets are relatively efficient, then we would expect a negative relationship between this measure of efficiency and the number of charter schools. This is exactly what we find. The efficiency coefficients are insignificant in both states, though they are consistently negative only in Michigan. However, the low significance levels are at least partially caused by the high correlation between test scores and this measure of efficiency.[21]

---

and Reporting (STAR). We use the test scores for math since there is some evidence that math scores are more sensitive to school quality, whereas reading scores are more dependent on interaction with parents in the home. The results are unaffected by the choice of reading and math tests.

[19] See, for example, Coleman (1966) and Harris (2000a).

[20] For Michigan, these scores are from 1993. For California, the scores are from 1998, which is the oldest available.

[21] The correlation range is 0.7-0.9, depending on the frontier specification.

Our data set includes many variables that measure the inputs of public school districts, including expenditure levels, student-teacher ratios, teacher salaries, and special education spending. To the extent that parents residing in a district do not pay the costs of education, we expect that fewer inputs will induce more charter schools to enter.

The interpretation of two input coefficients requires additional discussion. First, controlling for class size, teacher salaries, and special education spending, higher expenditures per pupil implies more spending on other inputs, such as after-school programs and administrators.[22] Second, variation in teacher salaries may reflect differences in compensating differentials across districts (e.g. crime) that are not accounted for by other included variables.

The results for school inputs are different across the two states. The positive relationship between charter schools and special education spending in Michigan is especially interesting, given the survey evidence that charter schools attract a disproportionate number of students with special needs compared with public schools.[23] The special education programs at traditional public schools often involve labeling kids and placing them in "pull-out" programs that separate kids from the mainstream. Also, Cullen (1999) finds that fiscal incentives lead traditional public schools to include more students in special programs. Charter schools, in contrast, have far lower funding levels

---

[22] We use two measures of teacher salaries, namely average teacher salary and the contractual starting salary for a teacher with a bachelor's degree and zero teaching experience.
One possible exception is teacher salaries, as indicated earlier.
[23] Finn (2000, p.79) states that "many charter schools attract youngsters with more problems and deficits than the conventional schools to which they are compared." Also, 20 percent of their survey respondents indicated that they chose a charter school because "my child's special needs [were] not met at [the] previous school."

and rarely offer pull-out programs. Therefore, one interpretation is that many parents prefer to keep their kids in mainstream classrooms and programs.

In Michigan, the coefficient magnitude for special education expenditures suggests that a one standard deviation increase in this category ($156 per pupil) increases the expected number of charter schools by 24 percent. For teacher salaries, a one standard deviation increase ($6,079) raises the number of charter schools by 23 percent. A similar increase in class size decreases the number by 6 percent.

The discussion of coefficient magnitudes above focuses on Michigan. In California, the magnitudes are larger for the demographic variables. For example, a one standard deviation increase in median household income is associated with a 117 percent drop in the number of charter schools. This may be due to differences in state equalization policy, as discussed earlier. The one demographic variable that has a smaller impact in California is the Herfindahl index for race, which here implies only a 10 percent decrease in the number of charters, compared with 30 percent in Michigan. The effects of most other variables are quite similar across the two states.

On the supply side, revenue will be a key factor regardless of a school's objective function. In both states, schools receive the per pupil foundation grant for the district in which the school is located. This grant also indicates the districts' minimum total spending for public schools, as guaranteed by the state government. Without going into detail about how this is calculated, the important characteristic of the funding system is that public school districts with a low foundation grant also tend to have total spending that is exactly equal or slightly above the foundation grant. This means that charter schools will have an easier time competing with public schools in low-spending districts.[24] We incorporate the revenue effect by including a variable that measures expenditures minus foundation grant, which should reveal a negative relationship with the number of charters. For California, the coefficient is negative, but insignificant. The same coefficients for Michigan are positive and insignificant. Therefore, this aspect of education policy does not seem to play an important role in explaining location.

In many ways, Detroit is an unusual school district in Michigan. It is by far the largest district in Michigan with an enrollment of about 180,000, which exceeds mean enrollment by a factor of 60! The ethnic composition of Detroit is much different from the rest of Michigan, the poverty rate is much higher and test scores are much lower. Perhaps most importantly, a very large proportion of all charter schools are located in Detroit. In order to check to what extent our results are driven by this single observation, we re-run the regression from table 3a column 7, omitting Detroit. We also run regressions omitting some other potential outliers, namely the very small school districts. The results are shown in table 4 and are fairly robust to these changes.

[24] The overall advantage/disadvantage is somewhat difficult to establish for two other reasons: 1) charter schools do not receive capital funds from the government; 2) the foundation grant does not account for grade level, allowing charter schools to focus on "cheaper" student populations. However, these differences should affect all districts in relatively equal ways.

Table 4: Regression Results With Sample Restrictions (Michigan)
Dependent Variable: Number of Charter Schools Per Michigan School District, 1998-1999

	With Detroit	Delete Detroit	Delete Smallest 5% (with Detroit)	Delete Smallest 10% (with Detroit)
Number of Observations	517	516	492	466
Med. Household Income	-0.0374 (0.0213) *	-0.0447 (0.0216) **	-0.0485 (0.0220) **	-0.0493 (0.0219) **
Poverty Rate	-0.0050 (0.0242)	-0.0074 (0.0244)	-0.0121 (0.0262)	-0.0161 (0.0262)
Average Yrs. of Schooling	0.9543 (0.2624) **	0.8418 (0.2511) **	0.9494 (0.2642) **	0.9164 (0.2630) **
Herf. Index for Ad. Education	-0.1025 (0.0519) **	-0.0826 (0.0451) *	-0.1140 (0.0523) **	-0.1095 (0.0510) **
Herf. Index for Race	-0.0244 (0.0102) **	-0.0186 (0.0096) *	-0.0264 (0.0107) **	-0.0280 (0.0109) **
MEAP Math Score 4th Grade	-0.0174 (0.0127)	-0.0169 (0.0130)	-0.0111 (0.0118)	-0.0101 (0.0123)
Production Efficiency	-3.7854 (4.8789)	-1.2361 (5.5450)	-5.8932 (4.8209)	-6.3623 (4.9279)
Graduation Rate	-0.0022 (0.0028)	-0.0004 (0.0010)	-0.0020 (0.0029)	-0.0019 (0.0027)
Avg. Special Educ. Expend. Per Student	1.5336 (0.8081) *	1.2851 (0.9139)	1.5302 (0.8325) *	1.4613 (0.8321) *
Pupil-Teacher Ratio	-0.0175 (0.0310)	-0.0457 (0.0321)	-0.0237 (0.0340)	-0.0258 (0.0338)
Average Teacher Salary	0.0371 (0.0193) *	0.0399 (0.0198) **	0.0396 (0.0200) **	0.0350 (0.0202) *
Total Enrollment	0.0255 (0.0033) **	0.0704 (0.0207) **	0.0267 (0.0035) **	0.0272 (0.0035) **
Gr. Rate of Enrollment	-0.0020 (0.0772)	0.0108 (0.0802)	0.0138 (0.0826)	0.0083 (0.0841)
Constant	-5.2281 (5.7439)	-6.7197 (5.9850)	-2.5142 (5.6140)	-1.3709 (5.6048)

Many charter schools attract students not only from the district in which they are located but also from neighboring districts. In order to account for this, we regress the number of charter schools not only on characteristics of the district in which the charter school is located, but also on characteristics of neighboring (contiguous) districts. There are two possible effects of neighboring districts: First, the parents may send their children to charter schools in other districts. If this were the only effect, then we would expect the neighboring coefficients to be of the same sign as the home district coefficients. Furthermore, if the charter schools expected most of their students to come from the home district, then the border-district coefficients should have lower magnitudes.

The home district coefficients in table 5 reveal results quite similar to those in table 3a. Also, unlike Downes and Greenstein (1996), we find that some characteristics of neighboring districts have a significant impact on the number of charter schools. We estimated the model in table 5 with various weighting schemes for the independent variables. The weights were based on the proportion of students in the local market and the distance between the population centers of the contiguous districts to the population center of the home district. (We hypothesized that districts that are far away and have few students would have less impact than nearby districts with many students.) We also estimated various models using different definitions for market size, adding nearby districts that were not contiguous. None of these variations in estimation had a significant impact on the results.

Table 5: Regression Results with Contiguous Districts
Dependent Variable: Number of Charter Schools Per Michigan School District, 1998-1999

INDEPENDENT VARIABLE	HOME	CONTIGUOUS
Number of Observations	516	
Med. Household Income	-0.0376 (0.0279)	0.0158 (0.0639)
Poverty Rate	0.0110 (0.0247)	-0.0833 (0.0535)
Average Yrs. of Schooling	1.0497 (0.2887) **	0.2305 (0.4131)
Herf. Index for Ad. Education	-0.1187 (0.0438) **	-0.1243 (0.1224)
Herf. Index for Race	-0.0238 (0.0110) **	0.0409 (0.0253)
MEAP Math Score 4th Grade	-0.0076 (0.0135)	-0.0914 (0.0375) **
Production Efficiency	-1.2660 (5.7479)	-9.3094 (4.8134) *
Graduation Rate	-0.0032 (0.0034)	-0.0043 (0.0066)
Special Educ. Expend/ Student	0.3621 (0.8689)	5.7005 (1.6936) **
Pupil-Teacher Ratio	-0.1105 (0.0381) **	0.4073 (0.0956) **
Average Teacher Salary	0.0157 (0.0217)	-0.0783 (0.0526)
Total Enrollment	0.0194 (0.0036) **	0.0088 (0.0071)
Gr. Rate of Enrollment	-0.0985 (0.0878)	0.3104 (0.1903) *
Constant	-3.9033 (4.9852)	--

V. Conclusion

This paper provides evidence about the equity and efficiency effects of charter schools by studying their location patterns. The results reinforce the conclusions of existing research that finds a significant impact of racial diversity. However, we clarify this result by including variables that measure other forms of diversity, including income and adult education levels. The effect of population diversity on the demand for charter schools may reflect differences in preferences for education programs, not just preferences for homogeneous student populations.

We find mixed support for the effect that charter schools have on school efficiency. Charter schools do appear to locate more in school districts with less efficient public schools. However, they also locate in districts that should already be competitive because of the presence of private schools. Therefore, instead of providing competition, charters may simply shift resources to students who previously went to private schools.

The results here also suggest that state policies toward charter and public schools do impact the location of charter schools. This means that state education policy at least partially determines the effects that charter schools have on equity and efficiency. Several other issues are left for future research. One is the impact that charter schools have on public school performance. Also, table 5 here indicates that the characteristics of border districts seem to impact charter location. While this specification does not seem to alter the general role of home district characteristics, the general issue of geography and market size warrants further attention.

APPENDICES

Appendix A

Table 6a: Summary Statistics (Michigan)

VARIABLE	MEAN	STD. DEV.	MIN	MAX
Med. House Value (thous.)	55.16	23.57	19.00	230.38
Per Capita Income (thous.)	12.57	3.96	6.22	48.34
Med. Household Income (thous.)	29.39	9.40	9.81	88.92
Perc. of Children in Poverty	15.00	9.61	0.00	51.71
Perc. of Children in Black	3.68	11.70	0.00	97.04
Perc. of Children in Hispanic	2.38	3.27	0.00	35.29
Herfindahl Index for Race	87.48	12.45	37.76	100.00
Perc. of Adult with 12 Grade	23.37	7.53	2.57	51.06
Perc. of Adult with High School	37.54	7.14	6.61	60.19
Perc. of Adult with Some College	26.56	5.90	3.19	50.76
Perc. of Adult ≥ Bachelor Degree	12.52	8.49	0.00	60.17
Average Years of Schooling	13.05	0.64	11.94	16.11
Herfindahl Index for Education	30.32	2.73	25.33	44.44
MEAP Math Score 4th Grade	43.13	13.51	4.30	100.00
Graduation Rate	88.73	115.03	0.00	2570.40
Total Expenditure Per Student (thous.)	4.36	1.14	0.00	14.44
Special Educ. Exp./Student (thous.)	0.33	0.16	0.00	0.86
Pupil-Teacher Ratio	20.14	3.60	8.00	41.00
Average Teacher Salary (thous.)	30.92	6.08	0.00	51.44
K-12 Enrollment Density	55.32	124.08	0.02	1068.50
Number of Private Schools	1.92	5.05	0.00	98.00
Total K-12 Enrollment (thous.)	3.03	8.36	0.00	183.15
Grad. Rate of K-12 Enrollment	-0.08	2.44	-11.95	15.07
Perc. of Public School located in City	5.54	21.75	0.00	100.00

Table 6b: Summary Statistics (California)

VARIABLE	MEAN	STD. DEV.	MIN	MAX
Med. House Value (thous.)	159.49	98.78	0.0000	490.98
Per Capita Income (thous.)	15.71	7.70	4.05	64.99
Med. Household Income (thous.)	35.22	13.50	11.41	122.91
Perc. of Children in Poverty	16.61	12.53	0.00	76.11
Perc. of Children in Black	3.16	6.29	0.00	71.51
Perc. of Children in Hispanic	28.37	23.82	0.00	100.00
Herfindahl Index for Race	59.58	18.16	25.98	100.00
Perc. of Adult with 12 Grade	26.66	16.18	2.12	90.19
Perc. of Adult with High School	25.07	7.24	0.00	87.50
Perc. of Adult with Some College	30.37	7.92	0.00	47.89
Perc. of Adult ≥ Bachelor Degree	17.90	12.93	0.00	70.53
Average Years of Schooling	13.41	0.98	11.20	16.67
Herfindahl Index for Education	31.26	6.89	25.25	81.83
STAR Math Score	618.25	19.67	555.20	681.90
Drop-Out Rate	1.80	4.26	0.00	68.50
Total Expenditure Per Student (thous.)	5.17	6.00	0.00	137.14
Perc. of Students in Special Education	23.94	84.86	0.21	1637.93
Pupil-Teacher Ratio	22.30	4.97	1.40	60.60
Entry Teacher Salary (thous.)	25.09	2.42	15.91	44.36
K-12 Enrollment Density	80.75	149.42	0.00	1450.75
Number of Private Schools	3.46	19.22	0.00	593.00
Total K-12 Enrollment (thous.)	4.98	21.57	0.01	639.78
Gr. Rate of K-12 Enrollment (thous.)	2.33	4.90	-14.45	67.10
Perc. of Public School in Cities	15.59	36.17	0.00	200.00

Appendix B

Table 7: The Educational Production Frontier Function (Michigan)
Dependent Variable: MEAP Math Test Score in 7th, 1993

VARIABLE	MLE ESTIMATES
Per Capita Income (Deflated by Teacher Cost Index)	0.1762 (0.1126)
Percentage of Children in Poverty	-0.3925 (0.1509) **
Percentage of Children in Black	0.0219 (0.0442)
Percentage of Children in Hispanic	0.0015 (0.0040)
Herfindahl Index for Race	-0.4558 (0.1338) **
Average Years of Schooling	-0.2908 (0.0379) **
Pupil-Teacher Ratio	0.2453 (0.0384) **
Total Expenditure Per Student (Deflated by Teacher Cost Index)	2.0855 (3.1521)
Total K-12 Enrollment	0.2079 (0.0357) **
K-12 Enrollment Density	-2.1753 (2.7757)
Average Teacher Salary (Deflated by Teacher Cost Index)	0.0051 (0.0032) *
Average Special Educational Expenditure Per Student (Deflated by Teacher Cost Index)	0.2974 (0.2445)
Constant	60.1321 (9.6782) **

BIBLIOGRAPHY

BIBLIOGRAPHY

Armor, D.J., & Peiser, B.M. (1998). "Interdistrict Choice in Massachusetts," In *Learning From School Choice*, Peterson and Hassel, editors. Washington, DC: The Brookings Institution, pp. 157-186.

Bettinger, E. (2000). "The Effect of Charter Schools on Public Schools," *Working Paper*, Massachusetts Institute of Technology.

Cameron, A.C. & Trivedi, P.K. (1986). "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests." *Journal of Applied Econometrics*, pp. 29-53.

Chambers, J. (1999). "Differences Across the United States: Introduction to a Teacher Cost Index," *American Institutes for Research.*

Chubb, J.E. & Moe, T.M. (1990). *Politics, Markets and America's Schools*, Washington, DC: The Brookings Institution.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfield, F.D., & York, R.L. (1966). *Equality of Educational Opportunity.* Washington, DC: U.S. Government Printing Office.

Coons, J.E. and S.O. Sugarman. (1978). *Education by Choice: The Case for Family Control.* Berkeley: University of California Press.

Cullen, J. (1999). "The Impact of Fiscal Incentives on Student Disability Rates," *NBER Working Paper No. 7173.*

Downes, T.A., & Greenstein, S.M. (1996). "Understanding the Supply of Non-Profits: Modeling the Location of Private Schools," *Rand Journal of Economics,* Vol. 27, pp. 365-390.

Epple, D. & Romano, R. (1996). "Ends Against the Middle: Determining Public Sector Provision When There are Private Alternatives," *Journal of Public Economics.*

Fernandez, R., & Rogerson, R. (1999). "Education Finance Reform and Investment in Human Capital: Lessons from California," *Journal of Public Economics*, Vol. 79, No. 3, pp. 327-50.

Filer, R.K. and D. Munich (2000). "Education Reform in the Post-Communist Czech Republic," *Manuscript.*

Finn, C.E., Manno, B.V., & Vanourek, G. (2000). *Charter Schools in Action.* Princeton, NJ: Princeton University Press.

Fitz, J., D. Halpin, & Power, S. (1993). *Grant Maintained Schools Education in the Market Place.* London: Kogan Page.

Friedman, M. (1962). *Capitalism and Freedom.* Chicago: University of Chicago Press.

Gifford, M., M. Ogle, and L. Solmon. (1998). "Who is Choosing Charter Schools? A Snapshot of Geography and Ethnicity of Charter School Students," *Manuscript.*

Glomm, G., D. Harris, and T. Lo (2000). "Are Urban School Districts Inefficient?" *Working Paper*, Michigan State University.

Grutzik, C., D. Bernal, D. Hirschberger, and A. S. Wells. (1995). "Resources and Access in California Charter Schools." Paper presented at the *American Educational Research Association Meetings*, San Francisco, 1995.

Harris, D. (2000). "New Approaches to Meta-Analysis with Applications to Education Production Functions," *Doctoral Dissertation*, Michigan State University.

Harris, D., S. Oliver, and D. Plank (2000). "International Lessons from School Choice," *Education Policy Center at Michigan State University.*

Hassel, B.C. (1998). "The Case for Charter Schools," In *Learning From School Choice*, Peterson and Hassel, editors. Washington, DC: The Brookings Institution, pp. 30-45.

Hausman, J., Hall, B.H., & Griliches, Z. (1984). "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica*, Vol. 52, pp. 909-38.

Hotelling, H. (1929). "Stability in Competition." *Economic Journal*, Vol. 39, pp. 41-57.

Hoxby, C. (1997). "Does Competition Among Public Schools Benefit Students and Taxpayers? Evidence from Natural Variation in School Districting," *NBER Working Paper No. 4979.*

Hoxby, C. (1994). "Do Private Schools Provide Competition for Public Schools?" *NBER Working Paper No. 4978.*

Huber, P.J. (1967). "The Behavior of Maximum Likelihood Estimates under Non-Standard Conditions." In Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics, Vol. 1, University of California, Berkeley.

Keljian, H., & Prucha, I. (1998). "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, Vol. 17, No. 1, pp. 99-121.

Manski, Ch. (1992). "Educational Choice (Vouchers) and Social Mobility." *Economics of Education Review*, vol.11, pp. 351-69.

Mintrom. M. (1998). "Michigan's Charter School Movement." *Manuscript.*

Nathan, J. (1996). *Charter School Creating Hope and Opportunity for American Education*, San Francisco: Tossey-Bass Publishers.

Nechyba, T. (1996). "Public School Finance in a General Equilibrium Tiebout World: Equalization Programs, Peer Effects and Private School Vouchers," *NBER Working Paper No. 5642.*

Papke, L. (1991). "Interstate Business Tax Differentials and New Firm Location. Evidence from Panel Data." *Journal of Public Economics*, Vol. 45, pp. 47-68.

Papke, L. (2000). "Michigan Applied Public Policy Research Project on K-12 Education," *Manuscript.*

Paris, D.C. (1998). "Standards and Charters: Horace Mann Meets Tinker Bell." *Educational Policy*, Vol. 12, pp. 380-396.

Peterson, P. E. (1999). "Vouchers and Test Scores," *Policy Review*, No. 93, pp. 10-15

Prescott, E. and M. Visscher. (1977). "Sequential Location Among Firms with Foresight." *Bell Journal of Economics*, Vol. 8, pp. 378-393.

Public Sector Consultants, Inc. and MAXIMUS, Inc. (1999) "Michigan's Charter School Initiative: From Theory to Practice." *Final Report.*

Vergari, S. and M. Mintrom. (1996). "Charter School Laws Across the United States." *Manuscript.*

Weinschrott, D.J., & Kilgore, S.B., (1998), "Evidence from the Indianapolis Voucher Program." In *Learning From School Choice*, Peterson and Hassel, editors. Washington, DC: The Brookings Institution, pp. 307-334.

Wells, A. S., Grutzik, C., & Carnochan, S. (1996). "The Multiple Meanings of US Charter School Reforms: Exploring the Politics of Deregulation." Paper presented at the British Educational Research Association Annual Conference, Lancaster, UK, 1996.

White, H. (1982). "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, Vol. 50, 1-26.

Whitty, G., Power, S., & Halpin, D. (1998). *Devolution and Choice in Education.*

Wohlstetter, P., Wenning, R., Briggs, K.L. (1995). "Charter Schools in the United States: The Question of Autonomy." *Educational Policy*, Vol. 9, December, 331-358.

Wooldridge, J.M. (1992). "Quasi Likelihood Methods for Count Data." *Journal of Economics*, pp. 352-406.

CHAPTER 2

New Approaches to Meta-Analysis with Applications to Education Production Functions [1]

I. Introduction

It is quite common for different tests of the same hypotheses to yield different conclusions. In these situations, some type of research synthesis is necessary, though often not sufficient, to reconcile conflicting results and reach valid conclusions. However, common approaches to synthesis, or "meta-analysis," are prone to producing imprecise and sometimes false conclusions. One reason for this appears to be that methods vary across studies, yet these differences are not appropriately accounted for in meta-analysis. The general purposes of this paper are to provides rough estimates of the degree to which methodological differences explain variation in results, and to provide alternative approaches to meta-analysis that appropriately account for methodological differences in primary research studies, improving the likelihood of reaching valid conclusions.

The usual focus of attention in applied research is the statistical significance of coefficients, which generally involves determining the confidence with which we can say a coefficient is positive, negative, or zero. While this provides some useful information, economists are often interested in the *average or expected effect magnitude*. For

[1] The author thanks Ron Fisher, Gerhard Glomm, John Goddeeris, David Neumark, David Plank, Peter Schmidt, Jeff Wooldridge, and participants in the MSU Labor Economics Seminar and the 2000 Midwest Economic Association meetings for useful comments on this paper. Financial support was provided by the U.S. Department of Education's North Central Regional Education Laboratory.

instance, in education, it is useful to know how student test scores are likely to be improved by a reduction in class size compared with an increase in teacher salaries.

In meta-analysis, it is possible to use a sample of individual estimates as data for *meta-estimators;* that is, creating new estimators from a series of other estimators. The simplest and most common example is the sample mean of a series of point estimates, i.e., *vote-counting*. The strength of this approach is that it can decrease the effects of sampling error on individual estimates. It also provides a straightforward estimate of the expected effect, which is useful for policy and welfare analysis. The weakness is that some studies use poor econometric methodologies, therefore, biasing both the individual and average estimates. Nonetheless, all of the coefficients are given one vote, regardless of the estimators' statistical properties.

This same vote counting philosophy underlies common approaches to statistical significance. Coefficients are again given equal weight and are each placed into one of two categories: significant or insignificant.[2] The portion in each category is then compared with the expected number under certain levels of significance. Hedges and Olkin (1985) show that when the effect is actually non-zero, the expected portion in each category depends on both the actual parameter value and sample sizes, a fact that is often not recognized in applications. These authors also state that common applications of "vote counting can fail for sample sizes and effect sizes that most commonly appear in educational and psychological research" (1985, p.51).

Hanushek (1986, 1996) and Greenwald, Hedges and Laine (GHL, 1994, 1996) have published well-known meta-analyses of the literature on education production functions (EPF). These papers attempt to estimate the marginal product of various school inputs in producing student outcomes, such as student test scores and future wages. Both sets of meta-analyses start with similar samples of estimates, labeled *H(test)* in table 1 below.[3] This sample only includes estimates in which student test score is the dependent variable. The independent variables of interest in this literature are school inputs, such as total school expenditures, teacher salaries, teacher education levels, and class size.

Two new samples are used in this paper: *C(test)* represents the current sample, which also focuses on students' test scores. Students' future wages are also commonly studied. This sample is labeled *C(wage)*. Each number in table 1 reflects the portion of studies in the respective samples whose estimates fall within the given categories. (If "money matter," then the expected coefficient for class size is negative, whereas the others are positive.)

Table 1: Signs and Significance Levels of Coefficients from Three EPF Samples
(Portion of Estimates in Each Category)

SCHOOL INPUT	SAMPLE	SIGNIFICANT		INSIGNIFICANT	
		Positive	Negative	Positive	Negative
Total Expenditures	H(test)	0.24	0.06	0.46	0.24
	C(test)	0.10	0.00	0.76	0.13
	C(wage)	0.20	0.00	0.72	0.06
Teacher Years of Education.	H(test)	0.09	0.08	0.38	0.46
	C(test)	0.03	0.03	0.26	0.14
	C(wage)	0.00	0.00	0.61	0.39
Teacher Salaries	H(test)	0.25	0.02	0.42	0.31
	C(test)	0.95	0.01	0.01	0.03
	C(wage)	0.81	0.00	0.14	0.00
Class Size	H(test)	0.15	0.10	0.47	0.27
	C(test)	0.09	0.11	0.28	0.52
	C(wage)	0.05	0.33	0.16	0.45

[2] This implicitly assumes a one-sided test. Two-sided tests are also possible.
[3] Note that these portions ignore estimates in Hanushek's sample that he reports as having "unknown signs."

Hanushek concludes from his sample that "there appears to be no strong or systematic relationship between school expenditures and student performance" (1986, p.1162). In addition, he states that "it seems somewhat unimportant to investigate the size of any estimated effects" (1997, p.144). Given the obvious importance of parameter magnitudes in policy and welfare analysis, he appears to mean there is too much variation in estimates to provide any precise information about actual parameter magnitudes.

Greenwald, Hedges and Laine (GHL, 1996) perform a different type of meta-analysis. First, rather than counting votes, they put more weight on studies that use "gold standard" econometric techniques. They also use a different standard for the number of estimates that must be positive and significant, compared with the portions in the other two categories. Despite using roughly the same sample of estimates as Hanushek, GHL indicate that "school resources are systematically related to student achievement and that these relationships are large enough to be educationally important" (1996, p.384).

One reason for these different conclusions is that the authors take different approaches to the variation in methodology. Hanushek acknowledges that some estimates may be biased, but suggests that it is difficult to identify quality estimates. As a result, his conclusions are based on the implicit assumption that *all* variation in estimates is due to sampling error or variation in the underlying parameters across samples. In contrast, GHL explicitly exclude certain types of estimates based on education and econometric theory.

The above discussion of the EPF meta-analyses highlights the fundamental problem of meta-analysis in non-experimental settings: We simply do not know the correct econometric specification. Researchers are forced to make assumptions that are, at best, loosely based on evidence. This paper makes four contributions toward improving research synthesis: First, in section II, I describe how regression can be used in meta-analysis. In these *meta-regressions*, the dependent variables are point estimates and other statistics from primary research studies, and the independent variables describe the methodologies used in estimation. One use of these models is to create new, and perhaps less biased, parameter estimators. A second use is to estimate the degree to which econometric methodology is systematically related to EPF estimates, i.e., how much "methodology matters." The approach also provides evidence about the role of specific methodological issues, such as data aggregation and omitted variables. These ideas have been discussed and partially implemented in the past, most notably by Jarrell and Stanley (1989, 1990). However, they misinterpret some of their results, and miss many important considerations.

Second, a particular advantage of meta-regression is that it allows for better corrections for publication bias, which can bias the results of meta-analysis. Berlin et al (1989) present a test for publication bias. Hedges (1992) proposes a correction that can be applied if the sample fails this test. However, both the test and correction include the assumption that point estimates are normally distributed *independent of methodology*. I show in section II that this assumption can be relaxed with meta-regression.

Third, meta-regressions are used to create a new estimator based on out-of-sample predictions from estimated meta-regression models. I discuss the statistical properties of this meta-regression estimator (MRE) in section II, leaving more formal proofs to the appendix. In section III, I introduce the literature on education production functions, as well as the two new samples of EPF estimates. Meta-regression is applied to these samples in section IV.

Fourth, GHL (1996) and other meta-analyses exclude studies based largely on econometric and economic theory. For the EPF literature, I show first that experimental evidence on class size reductions provides a narrow range of statistically significant effects. These precise and robust results imply that the information can be helpful in estimating education production functions. I propose one approach involving restrictions on estimation in primary econometric studies. In addition, I use the experimental evidence in the current meta-analysis to search for "gold standard" econometric methodologies. In brief, this approach involves identifying econometric point estimates that are close to estimates from experiments. This process and results are presented in section V.

II. An Introduction to Meta-Regression

A small number of economics publications have used meta-regression. Examples include Card and Kreuger (1995), Doucouliagos (1995), Jarrell and Stanley (1989, 1990), and Ashenfelter, Harmon, and Oosterbeek (1999). In addition, Glass, McGaw, and Smith (1981) and Hedges and Stock (1983) use meta-regression in other contexts. The application of meta-analysis, and meta-regression in particular, is more common in experimental research than in other fields.[4] Several textbooks provide discussion in the

[4] Meta-analysis, including meta-regression, appears to be more widely accepted in sciences other than economics. This is a curious fact because: (a) primary research in these fields does not often utilize regression (sometimes for good reason); and (b) experimental studies often include fewer methodological differences than is common in economics, making meta-regression comparatively less useful.

experimental context that can be extended to non-experimental work, including Glass, McGaw, and Smith (1981) and Hedges and Olkin (1985).

A somewhat controversial element of meta-analysis is the nature of the data. First, it appears to be generally impossible to achieve a random sample. Glass and Smith (1981) were some of the first researchers to focus on this issue, discussing the role of publication bias and other forms of selection bias. The publication bias problem can be addressed with the corrections suggested by Hedges (1992), and the improvements presented later in this paper.

A second issue is that the nature of the data may more accurately fit the assumptions of Bayesian statistical theory, as opposed to classical theory. Science is generally seen as a progression of knowledge, each study building on the others. This implies that each research project (observation) is conditional on those coming before it. Regardless, the assumption throughout this paper is that the assumptions of classical theory do in fact hold in this unique context.

II.A. The Meta-Regression Process. Any study including an econometric estimate of a parameter of interest is a part of the literature "population." This requires that the dependent variable in each paper represent the same construct, i.e., the same concept is being measured. For instance, in the EPF literature, two common dependent variables are wages and test scores, which clearly measure different things. Different studies also measure academic ability with different tests, however, academic ability may be considered a single construct that is measured in multiple ways.

Each study provides multiple estimates/observations, which individually serve as observations used in the meta-regressions. Each estimate contains multiple statistics of interest, which serve as the dependent variables. I refer to these collectively as the "original statistics" to distinguish them from "meta-statistics" calculated in the course of the meta-analysis. The focus of this analysis is on point estimates. For possible applications to other statistics, such as standard errors, see Harris (2000).

Each estimate in a given literature involves an underlying relationship that takes the form

$$y^{*} = a^{*}(x^{*}, \delta^{*}, \varepsilon^{*}), \tag{1}$$

where y^{*} indicates the original dependent variable of interest (e.g. student test scores in the EPF literature), x^{*} is the vector of independent variables, δ^{*} is a vector of parameters, and ε^{*} represents a disturbance.

Each estimate of equation (1) takes the form

$$y_i = a_i(v_i, c_{-i}, \hat{\delta}, e_{i1}), \tag{2}$$

where x^{*} from equation (1) is separated into two vectors, distinguishing the independent variables of interest v_i from the control variables c_{-i}. There are N total estimates in the sample, which are indexed by $i=1 \ldots N$. In theory, all of the elements of x^{*} should be included in either v_i or c_{-i}, however, most individual estimates will exclude one or more of them, causing omitted variable bias. In addition, functional form $a^{*}(\cdot)$ may not match $a_i(\cdot)$ for some i. The coefficient vectors are represented by $\hat{\delta}_i$ and $\hat{\delta}_{-i}$ where each element indicates the estimated relationship between one independent variable and the dependent variable.

The original statistics S are calculated through the estimation of (2). The relationship between these statistics, and the methods used to calculate them, is represented by

$$S^{*} = A^{*}(M^{*}, V^{*}, C^{*}, \beta^{*}, \eta^{*}) \tag{3}$$

where the independent variables in meta-regression M^{*} are the methodological properties of the associated estimates, including characteristics of the data, functional forms, and others. The term V^{*} is a vector of included variables of interest, while C^{*} is a vector of controls. The importance of this distinction will become more clear in the section V. The term β^{*} represents a vector of parameters, each of which relates one independent variable with the original statistic. Any characteristic that could influence these statistics should be considered a method variable.

Each estimate of (3) takes the form

$$S_i = A_i(M_i, V_i^{I}, C_{-i}^{I}, \hat{\beta}, e_{i2}), \tag{4}$$

where M^{*} in (3) is separated into three sub-vectors: The symbols V^{I}_{i} and C^{I}_{-i} are vectors of indicator variables that reflect whether these variables were included in the respective estimates. The term M_i indicates other methodological characteristics, many of which will also be coded as indicators. Each observation in (2) is coded to fit the form in (4), thus the indexes (i and k) are identical in both equations. All other notation in (4) is analogous to that in (2).

So far, the introduction to meta-regression here is simply a formalization of the discussion in Jarrell and Stanley (JS, 1989). These authors also discuss such issues as the subjectivity required in defining method variables and the normalizations required to make comparisons across studies. For instance, the dependent variable is often measured in different units across studies. Two key examples in this context are monetary variables, which must be adjusted for inflation, and test scores, which must be normalized based on the test score standard deviation. Different studies also use different functional forms, a point not mentioned by JS. For the purposes of meta-analysis, a specific functional form should be chosen with all estimates being normalized (e.g., at the means) before meta-regression models are estimated. The original functional form may still impact the normalized coefficient. Method variables should be used to indicate the original functional form used to identify this effect.

Hedges and Olkin (1985) show that heteroskedasticity is likely to exist when point estimates are dependent variables. The elasticity of the standard error with respect to degrees of freedom equals negative one when there is no publication bias. This implies a systematic relationship between the dependent variable and the error term in (4). Hedges suggests using a weight matrix in a generalized least squares framework, in which the diagonals are the associated degrees of freedom. However, this ignores the possibility that errors may vary systematically across studies for other reasons. It may therefore be reasonable to assume that all estimates within studies have the same error distribution, but that the variance may differ across studies, i.e., errors are clustered by study. The results should be similar if there are no important unobserved differences across studies.

II.B. Evidence From Meta-Regression. The existing literature on meta-regression, especially JS (1989, 1900) suffers from several weaknesses: ignoring important methodological issues, misinterpreting results, and ignoring potentially useful information. These issues are discussed below in four categories: (a) identifying the existence of non-linearities, structural change, and differences in underlying parameters; (b) estimating the effects of specific methodological differences and the relative roles of methodology, unobserved data set differences, and differences in underlying parameters in explaining variation in estimates; (c) testing and correcting for publication bias; and (d) creating new parameter estimators based on out-of-sample predictions from meta-regressions.

II.B.1. Non-Linearities, Structural Change, and Other Differences in Underlying Parameters. Meta-regression allows for tests of non-linearities, even when none of the original estimates includes such tests. This is accomplished by including the sample means of independent and dependent variables in the meta-regression. If a coefficient estimate varies systematically with a variable's sample mean, then this provides evidence of a non-linearity, ceteris paribus. The sign and magnitude of the meta-point estimate indicate the shape of the relationship.

Meta-regression can be used to test for time-dependent "structural change" by including time-related variables. For instance, in the EPF literature, the year of graduation from high school is included as a method variable. In may be that people who graduated in previous decades obtained greater gains from smaller class sizes, perhaps because of such societal changes as television viewing.

At any point in time, there may also be differences in underlying parameters based on the race, income, or other characteristics of the sample. These differences can be identified by including variables that measure the proportion of the sample that includes a particular characteristic. Other differences in underlying parameters may be identified by including variables that indicate characteristics of the sample. In section IV, for instance, I include variables for the proportion of each sample that comes from various racial groups.

II.B.2. The Roles of Methodology and Other Factors. One use of meta-regression involves the *meta-point estimates,* which indicate the relationship between study characteristics (M) and the point estimate of interest.[5] The discussion of Hanushek's work highlights the importance of also understanding the degree to which variation in results can be explained by general categories of factors: methodological differences, differences in underlying parameters, and sampling error. The specific variables used to identify these categories have already been discussed, except for sampling error, which is viewed here as a residual factor.

Adding categories of variables in a step-wise fashion, and studying the *meta-R^2*, provides one obvious approach understanding the relative role of each. The problems with the step-wise approach are well documented. For instance, omitted variable bias implies that some of the included variables are really explaining the effects of an excluded variable. This difficulty is somewhat exacerbated in meta-regression because the only way to know whether a variable is omitted is to read the original studies. For example, in estimating an original EPF, family characteristics appear to play a large role in explaining student success in EPF studies. If family income were omitted from an EPF estimate, researchers would then recognize its potential importance and hypothesize about the effect this would have on school input coefficients. However, in meta-analysis the relationship between independent (method) variables is completely arbitrary. Jarrell and Stanley (1990), in their application of meta-regression to the non-union wage gap, appear to miss these *meta-R^2* considerations when interpreting their results.

JS (1990) include indicator variables relating to authors and data sets. These play a role similar to region indicators in primary research, in which the indicators capture the net effects of many possible, but unknown factors. In the case of meta-regression these include unobserved methodological differences and measurement error. It is useful to include these "catch-all" variables, however, it is quite difficult to interpret the resulting change in the *Meta-R^2*. The inclusion of these variables may explain why JS are able to explain to such a large portion of the variance in estimates.

II.B.3. Testing and Correcting for Publication Bias. Publication bias is a common concern in meta-analysis because it implies that particular types of studies are systematically excluded from publication, usually based on the significance levels of key coefficients. Berlin, Begg, and Louis (1989) present a test to identify publication bias that has been implemented by Ashenfelter, et al (1999), Card and Krueger (1995), and Neumark and Wascher (1998). However, the Berlin approach assumes that the population of estimates is normally distributed *independent of methodology*. Meta-regression allows for formal tests of role of methodology to determine whether this normality assumption is realistic. In brief, this involves estimating the meta-regression model, identifying a base econometric methodology, and normalizing the coefficients to this base. Consider, for example, the effect of publication bias on the EPF expenditure

---

[5] These meta-point estimates say nothing about whether a characteristic is "correct."

coefficient. Suppose that the gold standard methodology calls for a constant elasticity functional form, but that one estimate is linear. The estimated meta-regression model would provide an estimate of the impact of this misspecification, which could then be added to the original estimate. Completing this process for each coefficient, and each difference in methodology, would produce a new distribution of coefficients, conditional on the base methodology. The Berlin test for publication bias could then be performed on this transformed sample. Implementation of this test is left for future research.

Hedges (1992) shows that publication bias based on significance levels will generally result in a sample of published point estimates that are biased upwards in absolute magnitude. In short, controlling for the standard error, a higher coefficient will produce a higher *t*-ratio, which is more likely to be published and observed in the meta-analysis. Hedges presents a correction for publication bias, but he makes the same normality assumption described above in the context of Berlin's test for bias.[6] The same normalization described above can be used in both cases.[7]

II.B.4. Statistical Properties of MRE. In this section, I discuss the statistical properties of the vote counting estimator (VCE) and the meta-regression estimator (MRE) with a focus on unbiasedness. If multiple methods are used in a given sample of estimates, which is likely, then many of the estimates are likely to be biased. The VCE, which is the mean or median of a sample of estimates, counts both the biased and the unbiased estimates the same way. It is obvious that the average of such a sample is also

---

[6] In brief, Hedges' correction involves estimation of the probabilities of observing point estimates with various levels of significance. These probabilities are then used as weights in a maximum likelihood framework to estimate the true parameter distribution.

[7] There may be less variation in methodology in experimental relative to non-experimental work, which may be why Hedges (1992, p.249) states that "meta-analyses

generally be biased. More importantly, the exclusion of biased estimates may not improve the situation if any biased estimates remain in the sample.

The weakness of this simple form of meta-analysis motivates the need to eliminate the biases of individual estimates. The MRE provides such an option. The MRE is calculated by developing vectors of method independent variables M^* that ideal estimates would include. This vector is then multiplied by the estimated vector of coefficients ($M^* \hat{\beta}$) to obtain expected values under ideal conditions.

The MRE has several advantages over standard forms of meta-analysis. First, the Hedges correction for publication can be directly incorporated in the model. Second, adjustments can be made for non-linearities and differences in underlying parameters. Despite these advantages, I show in the appendix that this estimator is still generally biased when some of the original estimates suffer from omitted variable bias. However, in cases where the most of the existing estimates appear to have significant methodological problems, the MRE may still be least biased.

III. Application: Education Production Functions

Since the *Coleman Report* in 1966, there has been an ongoing debate among academics and policymakers about the role of resource levels in improving outcomes in K-12 education. This industry comprises an increasing share of the U.S. economy with current expenditures levels comprising 4 percent of GDP.[8] Hansuhek and Rivkin (1996) document the large increases in resources in recent decades, especially for smaller class sizes. The EPF literature provides estimates of the marginal products of class size and

---

are unlikely to combine studies of fundamentally different research designs," referring only to experimental work.

other school inputs. These estimates are essential for evaluating recent and future changes in education policies.

III.A. Introduction to the EPF Literature. There are two general types of production functions that vary depending on the student outcome of interest. When student test score is the dependent variable, the assumed relationship is

$$t = \alpha_1 + \delta_1 f + \delta_2 s + \varepsilon \quad , \tag{5}$$

where test scores t are a function of student and family characteristics f, school resources s, and a parameter vector δ. I will refer to the collection of estimates of (5) as the "EPF test score literature."

There are several important differences when student future wage is the dependent variable. In this case, the equation is

$$w = \alpha_2 + \delta_3 f + \delta_4 s + \delta_5 y + \delta_6 j + \varepsilon_2 . \tag{6}$$

The wage w is a determined partially by the same student, family, and school characteristics as in (5). Years of schooling y and job experience j are also included. The parameter associated with schooling is often referred to as the "return to schooling" (RTS). Years of schooling is also assumed to be a function of student, family, and school characteristics. Specifically,

---

[8] See U.S. Census, 1996.

$$y = \alpha_3 + \varphi_3 f + \varphi_4 s + \varepsilon_3. \tag{7}$$

Many researchers therefore estimate *reduced form* equations that excludes attainment from equation (6).

There is a third functional form in the wage literature. This is a two-stage approach in which the first stage model is

$$w = \alpha_5 + \delta_{10} y \cdot r + \delta_{11} f + \delta_{12} j + \varepsilon_5. \tag{8}$$

The region of residence is a vector of indicators *r*, implying that δ_{10} is also a vector. Card and Krueger (1992) use this approach, estimating fifty separate returns to schooling, one for each U.S. state. In the second step, the return to schooling is then regressed on school characteristics, such as

$$\hat{\delta}_{10} = \alpha_6 + \gamma s + \varepsilon_6. \tag{9}$$

Substituting (9) into (8) shows that the two-stage procedure introduces an interaction between school inputs and years of schooling.

In the meta-analysis below, I use indicator variables relating to the above three functional forms: "returns to schooling," "reduced form," and "two-stage," respectively. As with other functional form differences, these must be normalized before coefficients can be compared. For instance, with the two-stage approach, I divide the reported coefficients by average education attainment in the sample to account for the implied interaction between *s* and *y*.

III.B. Methodological Issues in EPF. Several methodological issues have been raised in both the test score and wage literatures: (a) *data aggregation levels*; (b) *measurement error*, e.g., pupil-teacher ratios are often used as proxies for class size; (c) *omitted variables*, e.g., unobserved student and family characteristics may affect student outcomes; and (d) *simultaneity*, e.g., students are put into classes with more resources when they do not perform well, implying that test scores also determine school inputs.

Data aggregation has received considerable attention. There are two possible aggregation issues: student data and school input data. The first can be measured at the student, class, or higher levels. The second may be measured at the class or higher levels. Some studies indicate that more aggregated school input data tend to find larger point estimates and significance levels than studies with lower levels of aggregation. There have been at least two explanations given for the impact of school input data aggregation levels. First, measurement error may attenuate (bias) estimates toward zero. The aggregated school input measures can then be viewed as instruments, and higher estimates may be less biased. Second, Hanushek, Rivkin, and Taylor (1996) show theoretically that there is an interaction between omitted variables and aggregation level that would cause the aggregated estimates to be biased upwards in this literature. However, it is possible that other misspecifications in these models may be interacting with aggregation. The meta-regressions below control for many other study characteristics, providing more evidence on the direction and magnitude of the aggregation effect.[9]

[9] Section V provides evidence on which level of aggregation yields gold standard estimates.

III.C. New Samples of EPF Estimates. The test score sample used here includes seven studies: Akerhielm (1995), Dolan and Schmidt (1998), Eide and Showalter (1987), Ehrenberg and Brewer (1994), Ferguson (1991), Goldhaber and Brewer (1997), and Hanushek, Rivkin, and Taylor (1996). The characteristics of these studies are summarized in table 3. The test score sample includes four different data sets: High School and Beyond (HSB), National Education Longitudinal Survey (NELS), and two state-specific data sets from Virginia and Texas. These yield 87 estimates of the class size-test score parameter, which is the focus of analysis.

The wage sample is also summarized in Table 2. It includes eight studies: Altonji and Dunn (1996), Betts (1995), Card and Krueger (1992a, 1992b), Grogger (1996), and Olson (1998, 1999). These studies yield 85 estimates of the class size-wage parameter.

Table 2: Methods in the New Samples of EPF Studies
(Portion of Sample Using Each Method)

CATEGORY	VARIABLE	TEST SCORE EPF (87 Observations)	WAGE EPF (85 Observations)
Family/Student (C^I)	Income	0.84	0.29
	Parents' education	0.77	0.29
	Race	0.91	0.05
	Ability/IQ	0.14	0.02
School/Teacher (V^I)	Class Size	1.00	1.00
	Expenditures	0.38	0.18
	Teacher Salary	0.25	0.47
	Teacher Test Scores	0.12	0.00
	Teaching in Major	0.19	0.00
	% Master's Degree	0.57	0.58
	College Quality	0.17	0.00
	Experience	0.20	0.27
	School size	0.25	0.03
Test Content (M^I)	Math	0.21	NA
	Math/English	0.57	NA
	English	0.11	NA
	History	0.05	NA
	Science	0.05	NA
Functional Form (Class size only) (M^I)	Linear	0.55	0.00
	Log Depend. Var.	0.00	1.00
	Spline	0.31	0.00
	Squared Term	0.00	0.03
	Elasticity	0.14	0.00
Aggregation: Student Data (M^I)	Student	0.46	1.00
	Class	0.00	0.00
	School	0.15	0.00
	District	0.39	0.00
	Larger	0.00	0.00
Aggregation: School Data (M^I)	Class	0.28	0.61
	School	0.45	0.00
	District	0.27	0.00
	Larger	0.00	0.39
Other (M^I)	Robust t-ratios	0.38	0.75
	Gain Score	0.48	NA
	IV	0.05	0.01

IV. Results: Meta-Regression

IV.A. Meta-Regression and the Class Size-Test Score Coefficient. The focus of attention in this section is on the effect that class size has on student test scores. Table 3 below shows the results of the meta-regression when test score is the dependent variable. The models in columns (5), (6), and (8) have a *meta-R^2* of approximately 0.46, which is the highest of all eight specifications. This is somewhat lower than those found in JS (1990).[10] One likely reason is that there is simply more variation in methodology in the JS study and, therefore, more variables. This is especially true of the "catch-all" variables, such as data set and author indicators, which are likely to capture the effects of unobserved methodological differences. Columns (5), (6), and (8) are considered the preferred specifications below.

Huber-White robust standard errors are reported in table 3 where the weights are based on degrees of freedom.[11] The earlier discussion suggested that a clustered errors approach was warranted, however, the within study correlation is actually negative. Therefore, this approach was not used.

The class size coefficients for these regressions were all normalized to a linear functional form (at the means), and each test score was normalized based on the standard deviation. If decreasing class size has a positive effect on test scores, as we would expect, then the dependent variable should be negative. Thus, negative coefficients in the meta-regression in table 1 imply that the class size coefficient becomes closer to the

[10] Recall the earlier caveats regarding omitted variables, and the inclusion by JS of author indicators. The observable differences may be correlated with important, yet omitted, variable characteristics. Therefore, this is probably an upper-bound on the explanatory power of the model.

[11] See earlier discussion about the relationship between degrees of freedom and coefficient estimates.

expectation, or more negative. For instance, the estimated effect of using an instrumental variable (*IV*) in table 3 column 1 indicates that the use of an instrument changes the coefficient by –0.0021. (As a basis of comparison, the average class size coefficient in this sample is –0.0047.)

The coefficients on class size mean (*CS mean*), use of non-linear spline specifications (*Spline*), and *IV* are all significant and fairly stable across specifications. The elasticity indicator (*Elasticity*) is collinear with other variables and is therefore dropped *(--d--)* in most cases.

The coefficient on student data aggregation (*Agg: Student*) levels is consistently negative and significant in the preferred specifications. The effects of school input data aggregation (*Agg: School*) is ambiguous. This contrasts with the finding of HRT (1996) that higher aggregation leads to larger coefficient magnitudes.

If school inputs affect learning at all grade levels, then we would expect that applying the same low class size across all grades would reveal a larger impact at higher grades – i.e. the longer the treatment the larger the expected effect. Therefore, all the class size coefficients were divided by the grade at which the students took the test. Without this adjustment, the explanatory power of the models is somewhat lower, and the test grade becomes statistically significant with the expected positive sign.

Table 3: Meta-Regression with the Class Size-Test Score Coefficient
Dependent Variable: Class Size Point Estimate from Test Score Sample
(Robust t-ratios in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Meta-R^2	0.23	0.28	0.28	0.28	0.47	0.47	0.28	0.47
GENERAL								
Constant	0.0844 (1.069)	0.1077 (1.130)	0.1078 (1.116)	0.1112 (1.084)	0.0605 (0.547)	0.0605 (0.547)	0.1112 (1.069)	0.0605 (0.538)
CS mean	-0.0031 (1.067)	-0.0038 (1.113)	-0.0038 (1.098)	-0.0038 (1.098)	-0.0038 (1.020)	-0.0038 (1.020)	-0.0038 (1.083)	-0.0039 (1.004)
Spline	0.0058 (1.106)	0.0075 (1.951)*	0.0075 (1.925)*	0.0075 (1.925)*	0.0081 (1.081)	0.0081 (1.081)	0.0075 (1.898)*	0.0081 (1.064)
IV	-0.0021 (2.824)**	-0.0017 (1.647)*	-0.0017 (1.624)	-0.0017 (1.624)	-0.0017 (1.543	-0.0017 (1.543)	-0.0017 (1.602)	-0.0017 (1.520)
Elasticity	-0.0084 (1.575)	-0.0069 (0.463)	-0.0067 (0.446)	--d--	--d--	--d--	--d--	--d--
Agg: Student	-0.0001 (1.208)	-0.0001 (1.701)*	-0.0001 (1.170)	-0.0001 (1.170)	-0.0151 (8.519)**	-0.0151 (8.519)**	-0.0001 (1.154)	-0.0151 (8.397)**
Agg: School	-0.0121 (1.074)	-0.0157 (1.399)	-0.0159 (1.390)	-0.0192 (1.131)	0.0078 (0.411)	0.0078 (0.411)	-0.0192 (1.115)	0.0078 (0.405)
Grade Level	---	-0.0002 (0.243)	-0.0002 (0.239)	-0.0002 (0.239)	-0.0002 (0.211)	-0.0002 (0.211)	-0.0002 (0.236)	-0.0002 (0.207)
Gain Scores	---	-0.0003 (2.177)**	-0.0003 (2.440)**	-0.0003 (2.440)**	-0.0003 (2.318)**	-0.0003 (2.318)**	-0.0003 (2.406)**	-0.0003 (2.282)**

Table 3 (con't)

TEST CONTENT								
Math/English	---	-0.0005 (0.093)	-0.0004 (0.070)	0.0030 (0.912)	-0.0001 (0.027)	-0.0001 (0.027)	0.0030 (0.899)	-0.0001 (0.027)
English	---	-0.0046 (1.227)	-0.0046 (1.211)	-0.0046 (1.211)	-0.0046 (1.150)	-0.0046 (1.150)	-0.0046 (1.194)	-0.0046 (1.132)
History	---	-0.0038 (1.786)*	-0.0038 (1.762)*	-0.0038 (1.762)*	-0.0038 (1.674)*	-0.0038 (1.674)*	-0.0038 (1.737)*	-0.0038 (1.648)*
Science	---	-0.0043 (2.005)**	-0.0043 (1.978)**	-0.0043 (1.978)**	-0.0043 (1.873)*	-0.0043 (1.873)*	-0.0043 (1.950)*	-0.0043 (1.850)*
SAMPLE								
Black	---	---	-0.0004 (2.410)**	-0.0004 (2.410)**	-0.0004 (2.290)**	-0.0004 (2.290)**	-0.0004 (2.376)**	-0.0004 (2.254)**
Hispanic	---	---	-0.0000 (0.148)	-0.0000 (0.148)	-0.0000 (0.140)	-0.0000 (0.140)	-0.0000 (0.145)	-0.0000 (0.138)
OBS. SCHOOL VAR. (x10)	---	---	---	---	2 Drop 6 Signif.	2 Drop 6 Signif.	---	2 Drop 6 Signif.
UNOBS. SCH. VAR. (x5)	---	---	---	---	---	---	3 Drop 2 Signif.	3 Drop 0 Signif.
DATA SET INDIC. (x4)	---	---	---	3 Drop 0 Signif.	All Drop	---	3 Drop 0 Signif.	All Drop

Note: One asterisk (*) indicates 90 percent significance. Two asterisks (**) indicates 95 percent significance.

Some estimates include both a pre-test and a post-test. These *gain score* models appear to make the class size coefficient more negative, other things being equal. The content of the test also seems to be important. Math is the excluded variable in table 4 and all the included subject matter coefficients are negative.[12] This implies that class size has the smallest impact on math scores relative to the other subjects. *Blacks* and *Hispanics* both appear to gain more from smaller class sizes than do whites, even controlling for other family traits, though the results for Hispanics are insignificant

The inclusion of observed school characteristics (*OBSERVED SCH.*), e.g., teacher salary, appears to matter. Unobserved school characteristics (*UNOBS. SCH.*) include either indicators or individual effects (with panel data) for individual teachers, schools, and states.[13] These are collinear with other variables, indicating that some of the significant coefficients discussed above may actually pick up the effects of these excluded variables.

IV.B. Meta-Regression and the Class Size-Wage Coefficient. Many of the variables are different in the wage literature. Most of the original estimates used a log-linear specification. A few used elasticity forms and were transformed to the log-linear form at the sample means. The average class size coefficient in the sample is –0.004, implying that a one student reduction in class size would increase future student wages by 0.4 percent.

[12] Note that the *Math/English* variable represents one test in the sample that included a composite score for both math and verbal skills.
[13] State effects could pick up factors other than school differences.

Table 4: Meta-Regression with the Class Size-Wage Coefficient
Dependent Variable: Class Size Point Estimate from Wage Sample
(robust t-ratios in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Meta-R^2	0.12	0.32	0.55	0.54	0.34	0.37	0.57
GENERAL							
Constant	0.2823 (1.218)	0.4401 (2.058)**	0.2138 (0.401)	0.4902 (3.143)**	0.3815 (0.387)	-0.1462 (1.924)*	0.0349 (0.027)
CS mean	-0.0002 (1.995)**	-0.0002 (1.980)*	-0.0002 (2.094)**	-0.0002 (1.924)*	-0.0002 (1.883)*	-0.0002 (1.996)**	-0.0002 (1.972)*
Year of grad	-0.0001 (1.151)	-0.0002 (1.910)*	-0.0001 (0.339)	-0.0002 (3.033)**	-0.0002 (0.363)	0.0001 (0.103)	-0.0000 (0.014)
Job Experience	-0.0001 (0.782)	-0.0000 (0.625)	0.0000 (0.225)	0.0002 (1.679)	0.0001 (1.235)	0.0000 (0.103)	0.0003 (1.811)*
Average Wage	-0.0015 (0.976)	-0.0005 (0.355)	-0.0022 (1.489)	0.0008 (0.829)	0.0016 (0.221)	0.0003 (0.052)	-0.0002 (0.027)
Return to School.	-0.0036 (1.070)	-0.0010 (0.338)	-0.0055 (2.583)**	0.0007 (0.480)	0.0004 (0.236)	-0.0028 (0.847)	0.0003 (0.193)
Two-Stage RTS	-0.0025 (0.359)	0.0256 (1.800)*	0.0260 (1.331)	0.0166 (1.135)	0.0163 (0.535)	--d--	0.0191 (0.502)
Agg: School	---	-0.0107 (2.355)**	-0.0094 (1.892)*	-0.0089 (1.960)*	-0.0106 (2.211)**	-0.0126 (2.234)**	-0.0092 (1.748)*

Table 4 (con't)

OBSERVED SCH. (7 Total Variables)	---	---	0 Drop 1 Signif.	---		0 Drop 1 Signif.	0 Drop 1 Signif.
UNOBS. SCH. (1 Total Variable)	---	---	0 Drop 1 Signif.	0 Drop 1 Signif.	---	---	0 Drop 1 Signif.
OBSERVED STUD. (6 Total Variables)	---	---	---	---	0 Drop 1 Signif.	0 Drop 0 Signif.	0 Drop 0 Signif.
UNOBS. STUD. (2 Total Variables)	---	---	---	0 Drop 1 Signif.	0 Drop 0 Signif.	---	0 Drop 1 Signif.
DATA SETS (5 Total Variables)	---	---	---	---		2 Drop 0 Signif.	---

Note: One asterisk (*) indicates 90 percent significance; two asterisks (**) indicate 95 percent significance.

Table 4 indicates the results of the meta-regression for the wage literature. Nearly all of the coefficients in table 4 are quite sensitive to the other included method variables. This includes the year of graduation (*Year of Grad*), average potential years of job experience (*Job Experience*), average wage in the sample (*Average Wage*), and the included functional forms discussed in the previous section: returns to schooling (*Returns to School*) and two-stage returns to schooling (*Two-Stage RTS*).

One exception is the class size mean (*CS mean*), which has a negative and significant relationship with the class size coefficient, implying that the effects of class size reductions are larger at the upper end of the class size distribution. Another exception is the school data aggregation level (*Agg: School*), which is consistently negative and significant. This confirms the observation of HRT (1996) that school input effects seem larger with higher levels of data aggregation.

The inclusion of unobserved school differences (*UNOBS. SCH.*) has a statistically significant impact and explains a large portion of the variance in coefficient estimates. Observed school differences (*OBSERVED SCH.*), observed student differences (*OBSERVED STUD.*), and unobserved student differences (*UNOBS. STUD.*) all have statistically significant impacts, but they explain little of variance.

IV.C. Discussion. Since there is a positive correlation between test scores and wages, we would expect some qualitatively similar results in tables 4 and 5.[14] However, this is not what we observe. First, increasing aggregation of school inputs has a positive effect on the class size coefficient in the test score sample, but a negative effect in the wage sample.

---

[14] The magnitudes of these coefficients cannot be usefully compared.

Second, the use of observed school effects explains a large portion of the class size-test effect variance, but a small portion of the class size-wage effect variance. Conversely, unobserved school effects explain a large portion of the variance in class size-wage effects, but a small portion of the class size-test effects. There are several possible explanations for this. First, this may be spurious, due to collinearity between these variables and the data set indicators (which may capture other differences). Alternatively, success in the workplace may indeed depend more on unobserved characteristics of schools. This would have important implications for current policy reforms that attempt to change measurable differences, such as school input levels.

One of the main purposes of the meta-regressions was to establish the degree to which the variation in class size coefficients can be explained by observed methodological differences. The results suggest that approximately 57 percent of the variation can be explained by such differences. While this is smaller than the number observed by JS (1990), all of their regressions include author and data set indicators, which capture the net effects of unobserved differences. Overall, the results here provide evidence of the importance of methodology. The bigger challenge is finding econometric methods that produce unbiased estimates. This is addressed below in section V.

V. Identification of an EPF Econometric Gold Standard

V.A. Experiments in the Production of Education. The production function literature is somewhat unique because there are experimental estimates available for one of the key production function parameters, i.e., the class size effect. The Tennessee STAR experiment took place in the 1980's and included random assignment of 12,000 students to small and large classes for grades kindergarten through three (K-3). The average large class had approximately 24 students and the average small class had 15 students. Nye, Hedges, and Konstantopolous (NHK, 1999) and Finn and Achilles (1999) describe the study in greater detail. Hanushek (1999) focuses on the weaknesses of the project and previous interpretations of results, including attrition and some violations of random assignment.

The Wisconsin SAGE program randomly assigned 5,000 students to small and large classes with average sizes quite similar to the STAR program. The SAGE program, which started in 1996, also included staff development, after-school programs, and a new curriculum for kids in the treatment group. While this might appear to complicate the analysis, Molnar et al (1999) found that these other programs had no significant effect after controlling for class size.

Many other smaller scale experiments have been attempted. These and others have been studied in at least two meta-analyses. Glass and Smith (1979) studied 725 separate estimates that used varying degrees of sample control and randomization, some of which date back to the turn of the century. Hedges and Stock (1983) reanalyzed the Glass and Smith sample with some modifications, but found that these changes did not affect their results. Therefore, I refer to these results collectively as GS/HS.[15]

To compare the results from these three studies, two adjustments were made to the results reported by the authors. First, it is necessary to divide the experiment effect sizes by the average difference in class size between the control and treatment groups.[16]

[15] Mciverin, Gilman, and Tillitski (1989) also perform a meta-analysis of class size effects, but this is not discussed here.

[16] The average treatment sizes are: STAR (9 student reduction), SAGE (9 student reduction), and GS/HS (12 student reduction).

This yields the effect for a one-unit reduction in class size, which is the same form as the econometric estimates. Second, the length of time in the treatment varies and should also affect the outcome. Therefore, the estimates described above are divided by the average length of treatment.[17]

Table 5: Effect Sizes by Study and Sample Characteristics

Study Characteristics	Study		
	STAR	SAGE	GS/HS
Overall Avg.	---	---	-0.00204
Elementary	-0.00201	-0.00083	-0.00158
Secondary	---	---	-0.00312
Black	---	-0.00133	---
White	---	-0.00066	---
Good Controls	---	---	-0.00204
Poor Controls	---	---	-0.00158
Reading	-0.00150	-0.00066	---
Language Arts	---	-0.00094	---
Math	-0.00260	-0.00161	---
Science	-0.00200	---	---

The estimates reported in table 5 are quite similar across studies. For example, the coefficients for "elementary grades" range from –0.00083 to –0.00201. This implies that a one unit decrease in class size for an entire school year would increase test scores up to 0.002 standard deviations.[18] While these experiments certainly did have some weaknesses, there are several factors suggesting that these are near the actual parameter

[17] The average lengths of treatment are: STAR (4,320 hours), SAGE (2,160 hours), and GS/HS (432 hours).
[18] As a basis of comparison, a one standard deviation increase is equivalent to moving a student from the 50th percentile to the 84th percentile.

value. First, the point estimates are quite consistent across the three studies. Second, the effects are generally statistically significant.

A third reason to defend the results above is that they come from random assignment experiments. As Greene (1997, p.3) states, "The belief that we may scrutinize a set of non-experimental data and expect some complex truth to be revealed to us if we only spend enough time manipulating the numbers is hopelessly optimistic. In an experimental setting, we are free to choose the values of the stimuli and move them in whatever way we wish to elicit a change in the response variable . . . At best, [in non-experimental settings] we can . . . assume that the conditions necessary to employ our tools of statistical inference are met." Thus, even a poorly designed experiment is likely yield better estimates than the best non-experimental techniques.

V.B. Using Results from Experiments to Identify a Gold Standard. The discussion above suggests that a clear line exists between experimental and non-experimental research. However, it is possible to combine the two in a way that makes the most of the information available, achieving the precision and accuracy of experiments with the low costs of econometrics. The bridge between the two involves using the results from experiments as "extraneous information" in econometric estimation.

One common approach to using extraneous information is to impose restrictions on econometric models based on the information.[19] These restrictions, if true, help to improve a model's explanatory power. However, it is not helpful in identifying other desirable characteristics of the econometric specification. For instance, what levels of data aggregation are likely to produce unbiased estimates of other parameters?

[19] See Pudney, Deadman, and Pyle (2000) for an example of this type of approach using survey data as extraneous information.

A second, apparently new, approach to incorporating experimental evidence is used here: Starting with a broad sample of EPF econometric specifications, "gold standard" specifications are identified based on how similar the econometric coefficients are to related experimental evidence. To see the potential of this approach, it is useful to consider an example. Suppose that the actual education production function includes class size (*cs*), teacher salaries (*tsal*), teacher test scores (*test*), and a vector of other independent variables (*x*), such that:

$$y = \alpha_7 + \delta_{13} \cdot cs + \delta_{14} \cdot tsal + \delta_{15} \cdot test + \delta_{16} \cdot x + \varepsilon_7 \quad . \tag{10}$$

If teacher tests are excluded, then all the other coefficients will be biased unless the variables happened to be completely uncorrelated with teacher tests scores. Specifically, the coefficients on class size (*cs*) and teacher salaries (*tsal*) will equal

$$\hat{\delta}_{13} = \delta^*_{13} + \rho^*_{test,cs} \cdot \delta^*_{15} \tag{11}$$

$$\hat{\delta}_{14} = \delta^*_{14} + \rho^*_{test,tsal} \cdot \delta^*_{15} \tag{12}$$

where $\rho^*_{i,j}$ is the partial correlation between variables *i* and *j*, conditional on all other included variables. Both of the partial correlations in (11) and (12) could be estimated by a regression of teacher test scores on class size, teacher salaries, and the other variables (represented earlier by *x*).

If the experimental estimates are close to $\hat{\delta}_{13}$ then the bias of the class size parameter, represented by the last term of equation (11), is small. This does not necessarily mean that the last term of equation (12) is also small, which is the implied assumption behind this technique for selecting estimates. However, the assumption may be reasonable depending on how the anecdotal and empirical evidence that might shed light on the partial correlations. For instance, it seems unlikely that teacher test scores would be highly correlated with class size, but not teacher salaries. Similar statements could also be made about student ability, which many researchers point to as an excluded, or at very poorly measured, variable. If low ability students are systematically assigned to small class sizes, then it also seems likely that they are assigned to more experienced, and thus more highly compensated, teachers. Therefore, if one coefficient is close to the actual parameter value, it seems more likely that the others will be near their actual parameters as well.

There are two types of indirect evidence that may shed more light on the above assumptions in different circumstances: First, the coefficient estimates should be similar across studies for each of the included variables (e.g., class size and teacher salaries). An example of a formal test for this criterion is to compare the variance of the gold standard sample of estimates with the variance of the entire sample. Second, the methods used to obtain the gold standard sample should also be similar. A formal procedure might test whether the proportion of studies with particular characteristics is significantly different for the gold standard sample versus the entire sample. This second criterion is somewhat less important because it is possible that multiple methods could produce similar results.

Many of the econometric class size estimates in the current test score sample come close to those in the experiments above. Table 6 shows that there are seven estimates in "Range 1," which includes point estimates within the range for elementary grade students from table 5 ($-0.0020 < \delta > -0.0008$). There are nine estimates in "Range 2," which includes a somewhat wider range of point estimates ($-0.0040 < \delta > -0.0004$). The sample means are presented for noteworthy method variables. For the dichotomous variables, these can be interpreted as the portions of the studies in the sub-sample that have the respective characteristics. The polychotomous variables must be taken on a case by case basis. Student and school input aggregation levels are coded as follows: student level = 1; class level = 2; school level = 3; school district level = 4; and higher levels = 5. Thus, the average for Range 1 gold standard estimates for school input data in the test scores literature is 1.71, or somewhere in between the classroom and district levels. Job experience is measured in years.

In the wage literature, it is somewhat more difficult to find an unbiased estimate with which to compare econometric estimates because experiments do not usually follow students after graduation. One way to approach the issue is to assume that school inputs affect future wages only through academic achievement and attainment. Manksi (1987) calculates the effects of SAT scores on future earnings, controlling for class rank and gender. This is not ideal, since a relatively small percentage of American students take this test, and relatively few control variables are included. Nonetheless, he estimates that the effect of a one standard deviation increase in SAT score produces a 2-4 percentage increase in future wages.[20] This implies that a low end estimate of the effect of class size on student earnings is (0.02)(-0.0008) = -0.000016. The second term (-0.0008) is the lowest of the three class size-achievement coefficient estimates from table 3. The high end would then be (0.04)(-0.0020) = -0.0008. The low end of this range is almost certainly an underestimate of the economic benefits because it only accounts for the role of academic achievement, therefore, the absolute magnitude of the effects in my "Range 1" is somewhat larger: –0.00016 to –0.0008. "Range 2" is –0.0005 to –0.004.

All four of the variables describing family characteristics (the first four variables listed in table 6) are almost always included in each of the gold standard test score estimates, but their effects on class size-wage effect are less clear. These variables are usually included in the other estimates as well, but to a somewhat lesser extent. The differences become more noticeable in Range 2 for the test score sample. Student ability and living in rural areas also appear to improve class size-test estimates, but not class size-wage estimates.

[20] Blau and Kahn (2000) estimate the relationship between wages and scores on the International Adult Literacy Survey (IALS) for the United States and other countries. The results are quite similar to Manski's estimates based on the SAT.

Table 6: Methods of Gold Standard Versus Other Estimates

	TEST SCORE SAMPLE				WAGE SAMPLE			
	RANGE 1		RANGE 2		RANGE 1		RANGE 2	
	Gold	Other	Gold	Other	Gold	Other	Gold	Other
OBSERVATIONS	7	78	14	71	7	81	24	63
DICHOTOMOUS								
Race	1.00	0.87	1.00	0.88	0.14	0.21	0.15	0.21
Parent's Income	1.00	0.79	1.00	0.79	0.43	0.38	0.46	0.40
Parent's Education	0.86	0.77	1.00	0.74	0.43	0.38	0.46	0.40
% Single Parent HH	1.00	0.80	0.93	0.78	0.00	0.11	0.00	0.11
Student Ability	0.14	0.11	0.22	0.09	0.14	0.04	0.05	0.03
Rural	1.00	0.94	1.00	0.95	NA	NA	NA	NA
% Teacher Mast.	0.43	0.44	0.67	0.53	1.00	0.21	0.92	0.20
Relative Teacher	0.00	0.30	0.00	0.30	0.57	0.23	0.62	0.21
Absolute Teacher	0.14	0.40	0.18	0.41	0.00	0.10	0.00	0.10
Teacher Experience	0.43	0.19	0.29	0.30	0.43	0.09	0.31	0.09
Expenditures	0.43	0.32	0.29	0.33	0.00	0.25	0.00	0.27
Teacher Race	0.86	0.46	0.64	0.46	NA	NA	NA	NA
School Race	0.86	0.52	0.71	0.52	NA	NA	NA	NA
School Effects	0.00	0.02	0.00	0.02	0.00	0.12	0.00	0.13
Spline/squared	0.43	0.27	0.50	0.25	0.00	0.02	0.00	0.02
IV	0.14	0.03	0.21	0.01	0.00	0.02	0.00	0.03
Return to School	NA	NA	NA	NA	0.29	0.33	0.42	0.32
Two-stage RTS	NA	NA	NA	NA	0.14	0.20	0.25	0.20
POLYCHOTOMOUS								
Agg: School	1.71	2.1	1.86	2.06	2.42	2.93	2.75	2.92
Agg: Student	1.57	1.52	1.29	1.56	1.00	1.00	1.00	1.00
Job Experience	NA	NA	NA	NA	18.1	19.0	19.3	18.9

The effects of measured school inputs also vary dramatically across the two samples. In the test score sample, the inclusion of percentage of teachers with master degrees, relative teacher salary, absolute teacher salary and expenditures does not seem to improve the class size estimates. In fact, both of the two teacher salary measures are found much less often in the gold standard estimates. In contrast, teacher experience and teacher race both seem to improve the class size effect estimates.

In the wage literature, there are substantial differences in methods in the gold standard sub-sample. Inclusion of the percentage of teachers with master degrees, relative teacher salaries, and teacher experience all seem to clearly improve estimates. None of the gold standard estimates include either expenditures or absolute teachers salaries.

A few other methodological characteristics seem to have some influence. Non-linear and IV specifications seem to improve estimation for the test score sample. None of the gold standard estimates include unobserved school effects.[21] Also, there seems to be no substantial difference between the three functional forms in the wage literature. This is quite surprising because the reduced form model does not estimate structural parameters. This seems to suggest that class size does not have a substantial impact on years of schooling or other variables that explain wages.

The aggregation of student and school variables are consistently lower in the gold standard estimates, but not strikingly so. Average years of job experience are also not substantially different in the gold standard wage sample. Overall, these results suggest that the bias of estimates is determined primarily by common observed variables relating to student and school characteristics.

The results in table 7 may not present a complete picture of the differences between the gold standard and other estimates because characteristics may be inter-related. For instance, race and income may be so highly correlated that only one of them is necessary to achieve a nearly unbiased estimate. It may therefore useful to consider

[21] Recall from the meta-regressions that unobserved school differences seemed to explain much of the variance the wage sample, while observed school differences explained much of the variance in the test score sample. It is quite possible that these factors could explain a large portion of the variance without producing unbiased estimates.

profiles of the individual estimates. In this case, such comparisons do not appear to reveal anything new and are therefore omitted.

Recall that this technique of identifying gold standard estimates should pass two tests. First, there should be consistency in methodology, which was discussed above. Second, the methods should produce a fairly narrow range of estimates. Table 7 below is similar in format to table 6, but instead focuses on the point estimates for the school input parameters. Information about *t*-ratios is provided only for the class size coefficient, since this is the only coefficient for which we have experimental evidence. (The class size effects are consistently significant in the experimental literature.)

Table 7 indicates that the coefficients on expenditures and teacher experience are consistently positive and fall within narrow ranges in each sample. The average coefficient is almost always positive for the other school inputs, but the minimums are sometimes negative. Also, only 12 percent of the coefficients from gold standard estimates summarized in table 7 have unexpected signs compared with 22 percent for entire sample.[22] Many of the variables are missing in the gold standard samples, which is unfortunate for identifying coefficient magnitudes, but still informative about what is necessary to make unbiased estimates.

[22] Both sets of numbers exclude class size coefficients.

Table 7: Statistics from Gold Standard Versus Other Estimates

EPF STATISTIC	VALUE	TEST SCORE SAMPLE				WAGE SAMPLE			
		Range 1		Range 2		Range 1		Range 2	
		Gold	Other	Gold	Other	Gold	Other	Gold	Other
Expenditure/Pupil	Mean	+0.067	+0.014	+0.051	+0.015	NA	+0.015	+0.074	+0.014
Point Estimate (x100)	Min	+0.045	- 0.024	+0.006	- 0.024		- 0.096	+0.074	-0.096
	Max	+0.078	+0.065	+0.078	+0.065		+0.117	+0.074	+0.117
Abs. Teacher Salary	Mean	- 0.002	+0.013	+0.002	+0.015	NA	+0.030	+0.266	+0.282
Point Estimate (x100)	Min	- 0.002	- 0.022	- 0.022	- 0.016		+0.007	+0.008	+0.007
	Max	- 0.002	+0.048	+0.029	+0.048		+0.523	+0.523	+0.179
Rel. Teacher Salary	Mean	NA	- 0.308	NA	- 0.308	+0.872	+0.098	+0.129	+0.088
Point Estimate	Min		- 0.607		- 0.607	+0.853	- 1.014	-0.040	-1.014
	Max		- 0.122		- 0.122	+0.942	+1.668	+0.373	+1.668
% Teacher Master's	Mean	+0.00003	+0.001	- 0.0008	+0.675	+0.037	+0.045	+0.047	+0.043
Point Estimate	Min	+0.00000	- 0.003	- 0.0030	- 1.35	-0.001	- 0.325	- 0.003	-0.325
	Max	+0.00010	+0.020	+0.0001	+3.37	+0.151	+0.310	+0.174	+0.310
School Size	Mean	NA	- 0.007	-0.001	- 0.007	NA	- 0.002	NA	-0.002
Point Estimate	Min		- 0.083	-0.002	- 0.083		- 0.006		-0.006
	Max		+0.250	+0.001	+0.025		+0.000		+0.000
Teacher Experience	Mean	NA	+0.001	+0.001	+0.001	+0.013	+0.021	+0.017	+0.021
Point Estimate	Min		+0.001	+0.001	+0.001	+0.007	- 0.109	+0.007	-0.109
	Max		+0.003	+0.001	+0.003	+0.019	+0.102	+0.022	+0.102
Teacher Test Score	Mean	NA	+0.194	+0.235	+0.191	NA	NA	NA	NA
Point Estimate	Min		- 0.013	+0.235	- 0.013				
	Max		+0.266	+0.235	+0.266				
# School Days/Year	Mean	NA	+0.002	-0.001	+0.002	+0.009	+0.001	+0.054	-0.010
Point Estimate	Min		-0.001	-0.001	- 0.000	+0.003	- 0.281	-0.013	-0.281
	Max		+0.004	-0.001	+0.004	+0.020	+0.594	+0.251	+0.594

Table 7 (con't)

Class Size	Mean	-0.0012	-0.0009	-0.0012	- 0.0009	-0.0012	- 0.0042	-0.0023	-0.0046
Point Estimate	Min	-0.0018	-0.0304	-0.0024	- 0.0304	-0.0016	- 0.0527	-0.0036	-0.0527
	Max	-0.0008	+0.016	-0.0004	+0.0157	-0.0008	+0.0308	-0.0008	+0.0308
Class Size	Mean	2.391	1.099	2.16	1.028	4.380	1.963	2.330	2.107
t-ratio	Min	0.410	0	0	0	0.2	0	0.2	0
(Absolute Values)	Max	4.41	7.71	5.4	7.71	27.3	13.4	27.3	13.4

In table 8 below, I present a summary of the results for the class size coefficients for all the estimators considered here. The first row shows the results from the three experiments discussed earlier. These also reflect the gold standard estimates, which were selected based on the results from experiments. The second row has the VCE, similar to Hanushek's approach, which takes the simple mean and range of estimates in both samples; none are excluded. The next row restricts the current samples based on the GHL (1994) criteria.

The MRE, shown in the last row of table 8, was discussed in section II and in the appendix. The actual M^* includes a series of vectors, not a single ideal. The vectors all include every school input variables and student characteristic control variables. For both literatures, I assume school or district levels of input aggregation (as supported by the earlier analysis) and more recent years (which more closely match most of the experimental estimates and the parameters of interest here). For the wage literature, I assume the reduced form is ideal. For the test score literature, I tried various types of test content and different race samples. The averages for each sample are presented below. Both the test score sample and the wage sample include three preferred specifications in tables 4 and 5, respectively. I used all three in estimating the MRE. In all, I created 45 different specifications (fifteen for each of the three preferred models). Table 8 reports mean minimum and maximum.

Table 8: Comparing Class Size Point Estimates from Different Meta-Analyses

TYPE OF STUDY	VALUE	TEST SCORE SAMPLE	WAGE SAMPLE
Experiments (table 3)	Mean Min Max	- 0.00150 - 0.00201 - 0.00080	- 0.00023 - 0.00042 - 0.00008
VCE (Unrestricted: Hanushek)	Mean Min Max	- 0.00074 - 0.03042 +0.01567	- 0.00397 - 0.05273 +0.03079
VCE (Restricted: GHL Criteria)[23]	Mean Min Max	- 0.00103 - 0.03043 +0.01567	- 0.00048 - 0.00814 +0.01200
MRE	Mean Min Max	- 0.00122 - 0.00147 - 0.00075	- 0.00021 - 0.00043 +0.00034

Both of the VCE approaches yield extremely wide ranges of estimates, though the means are surprisingly close to those from the experiments, except for the unrestricted version in the wage sample. The MRE also performs surprisingly well with much less variation and averages that are extremely close to the experimental results in both samples. This suggests that further work on the MRE is warranted, work which may include adding adjustments for publication bias.

What do the coefficient magnitudes reported in tables 8 and 9 imply about the level and allocation of school resources? To help answer this question, chapter 3 includes a calibrated partial equilibrium model that is used to estimate the magnitude of the EPF estimates required to justify current spending patterns. The results suggest that the current level of resources would be justified, based solely on increases in human

[23] GHL exclude studies that do not have the following characteristics: "the outcome measure is some form of academic achievement," "the level of aggregation is as the level of school districts or smaller units," and "the model controls for socioeconomic characteristics or is either longitudinal (including a pretest and a post-test) or quasi-

capital, if the coefficient on expenditures were 0.18.[24] None of the expenditure coefficients in the current sample are this large. (See table 7 above.) The portion of current resources allocated to class size (teacher salaries) would be justified if the class size (teacher salaries) coefficient were –0.43 (0.016). Some of the teacher salary estimates in the current sample, including some in the gold standard subsample, are at least as large as these required magnitudes. However, the same cannot be said of the class size coefficients, all of which are substantially smaller in the current sample.

Combining the results here with chapter 3, therefore, suggests that the current level and allocation is sub-optimal. While the focus of both papers is on future earnings, the gap between optimal and actual levels is so large that it is difficult to imagine that other economic benefits would be large enough to make up the difference. In addition, the allocation of resources seems inappropriately skewed toward smaller class sizes and away from higher teacher salaries. This conclusion also holds when the focus is on test scores. Chapter 3 also uses the results found here to determine the most cost-effective ways of improving test scores. Again, the results provide evidence that improvements in achievement could be obtained at lower cost with increases in teacher salaries.

---

longitudinal (including IQ or a measure of earlier achievement as an input)." I applied the second and third criteria to the wage sample.

VI. Conclusion

The ongoing debate about education production functions highlights many of the problems that arise when synthesizing disparate results from a large number of studies. First, there are uncertainties about the correct econometric specification, which are compounded by data limitations. One possible response, taken by Hanushek, is to assume that methodology differences explain only a very small portion of the variance in estimates. The meta-regression results in section IV suggest that 47-57 percent of the variance in EPF class size estimates can be explained by observed differences, implying that Hanushek's assumptions are likely to lead to false conclusions. A second approach, taken by GHL, is to develop criteria to identify methods that are likely to minimize bias. The results in section V support the idea that particular methods are better than others, but that theory alone, and the GHL theories in particular, are unlikely to provide enough guidance.

A second contribution of this paper is showing the consistency of results from experiments on class size. Previous studies have focused on the strengths and weaknesses of particular studies, such as Tennessee's STAR, however, there appear to be no other studies that compare the magnitude of the reported effects across studies.

Third, I show how information from experiments can be usefully combined with econometric research to identify gold standard econometric methods. This approach combines the accuracy of experiments with the low cost of econometrics. The results of this approach to selecting studies indicates that the effects of school inputs are positive, but that they do not justify the current level or allocation of resources. Other studies,

---

[24] In other words, the marginal change in the presented discounted value of future earnings equals the marginal change in the present discounted value of school costs at

such as GHL, have come to this same conclusion, but through methods that have numerous weaknesses.

Several extensions would help to clear up remaining uncertainties. First, the information from experiments could be used in other ways, perhaps by restricting coefficients in the estimation of education production functions based on experimental evidence. Second, more econometric studies would be useful to eliminate some of the collinearity in the current sample of studies. Third, many of the ideas presented have not been implemented, e.g., transforming the samples of coefficients in tests and corrections for publication bias, and exploring the possible advantages and disadvantages of MRE under various contexts.

Perhaps the most important extension of this paper is that these techniques could be usefully applied to other research literatures for which there are large numbers of studies, both experimental and non-experimental. The ever-expanding number of research studies will expand the list of possible applications. However, without proper synthesis, these studies will add more confusion than they do knowledge.

---

this point.

APPENDICES

Appendix A

Statistical Properties of MRE

This appendix provides a proof that the meta-regression estimator (MRE) is generally biased for the case of omitted variables bias in the original estimates. It may be useful to review the notation in equations (1)-(4) in section II, which is also used here.

Suppose there are three variables in x^*: the variable of interest (v_1) is included in each estimate; c_2 and c_3 are controls that may or may not be omitted in any given estimate. Further, suppose observation #1 has only the variable of interest (v_1); observation #2 includes v_1 and c_2; and observation #3 includes v_1 and c_3. After coding these as indicator variables, we have the matrix of method variables M such that

$$M = \begin{bmatrix} v_{11} & c_{12} & c_{13} \\ v_{21} & c_{22} & c_{23} \\ v_{31} & c_{32} & c_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \tag{10}$$

The actual vector of original parameters is represented by

$$\delta^* = \begin{bmatrix} \delta_1^* & \delta_2^* & \delta_3^* \end{bmatrix} \tag{11}$$

Again, the first parameter is the one of interest. Suppose we have three estimates of this parameter represented by the vector

$$\hat{\delta}_1 = \begin{bmatrix} \hat{\delta}_{11} \\ \hat{\delta}_{12} \\ \hat{\delta}_{13} \end{bmatrix} \tag{12}$$

The elements δ^*_2 and δ^*_3 reflect the effects of the control variables on the original dependent variables. While these are not the parameters of interest, they are necessary for determining the bias caused by the omission of these variables from the original regression. The omitted variables cause each element of $\hat{\delta}_1$ to be biased. For instance,

$$\hat{\delta}_{11} = \delta_1^* + \delta_2^* \rho_{21} + \delta_3^* \rho_{31} \neq \delta_1^* \tag{13}$$

where ρ_{ij} represents the partial correlation between the omitted variables (indexed by i) and the included variables (indexed by j). These partial correlations can be estimated from a regression of the omitted variable on all the included variables. Thus, ρ_{21} comes from the regression of c_2 on v_1. The other two estimates take the forms

$$\hat{\delta}_{12} = \delta_1^* + \delta_3^* \rho_{321} \neq \delta_1^* \tag{14}$$

$$\hat{\delta}_{13} = \delta_1^* + \delta_2^* \rho_{213} \neq \delta_1^* \tag{15}$$

In equation (14), the term ρ_{321} could be estimated by a regression of c_3 on c_2 and v_1. The term ρ_{213} in equation (15) is defined analogously.

Returning to the meta-regression, assume the true meta-regression model for δ_1^* is

$$\delta_1^* = M^* \beta^* + \eta^* = \alpha_7^* + \beta_1^* C_2^{l*} + \beta_2^* C_3^{l*} + \eta^* . \tag{16}$$

The OLS estimator of β^* is

$$\hat{\beta} = (M' M)^{-1} (M' \hat{\delta}_1). \tag{17}$$

Substituting in M from equation (10) yields

$$\hat{\beta} = \begin{bmatrix} \hat{\alpha}_7 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \delta_1^* + \delta_2^* \rho_{21} + \delta_3^* \rho_{31} \\ -(\delta_1^* + \delta_2^* \rho_{21} + \delta_3^* \rho_{31}) + (\delta_1^* + \delta_3^* \rho_{321}) \\ -(\delta_1^* + \delta_2^* \rho_{21} + \delta_3^* \rho_{31}) + (\delta_1^* + \delta_2^* \rho_{213}) \end{bmatrix} \tag{18}$$

The second and third elements of this vector can be interpreted as the effect of including the given variable when it was previously excluded, ceteris paribus.

The MRE is calculated by multiplying $\hat{\beta}$ by M^*. I assume that all three variables are relevant, therefore, M^* is a vector of ones. This yields

$$\delta_1^{MRE} = M^* \hat{\beta} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \hat{\alpha}_7 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \delta_1^* + (\delta_2^* \rho_{213} - \delta_2^* \rho_{21}) + (\delta_3^* \rho_{321} - \delta_3^* \rho_{31}). \tag{19}$$

This equation implies that the MRE is biased in this simple case, but that the bias may be small depending on the partial correlations between the included and excluded variables. This results is somewhat intuitive: it says essentially that the effect of one omitted variable depends on which other variables are omitted.

Appendix B

Table 9: Profiles of Gold Standard Estimates (Range 1 from Table 6)

	TEST SCORE SAMPLE						
	1	2	3	4	5	6	7
AUTHOR	A	A	A	E	E	E	D
DICHOTOMOUS							
Race	1	1	1	1	1	1	1
Parent's Income	1	1	1	1	1	1	1
Parent's Education	1	1	1	1	1	1	0
% Single Parent HH	1	1	1	1	1	1	1
Student Ability	0	0	0	0	0	0	1
Rural	1	1	1	1	1	1	1
% Teacher Mast.	0	0	0	1	1	1	0
Relative Teacher	0	0	0	0	0	0	0
Absolute Teacher	0	0	0	0	0	0	1
Teacher Experience	0	0	0	1	1	1	0
Expenditures	0	0	0	0	0	0	0
Teacher Race	1	1	1	1	1	1	0
School Race	1	1	1	1	1	1	0
School Effects	0	0	0	0	0	0	0
Spline/squared	1	1	1	0	0	0	1
IV	1	1	1	0	0	0	0
Return to School	-	-	-	-	-	-	-
Two-stage RTS	-	-	-	-	-	-	-
Gain Score	0	0	0	1	1	1	0
POLYCHOTOMOUS							
Agg: School	1	1	1	2	2	2	3
Agg: Student	0	0	0	3	3	3	2
Job Experience	-	-	-	-	-	-	-

A = Akerheilm (year)

E = Ehrenberg (year)

D = Dolan and Schmidt (year)

B = Betts (1995)

O = Olson (1998)

BIBLIOGRAPHY

BIBLIOGRAPHY

Akerhielm, K. (1995) "Does Class Size Matter?" *Economics of Education Review*, Vol. 14, No. 3, pp. 229-241.

Altonji, J., and Dunn, T. (1996) "Using Siblings to Estimate the Effect of School Quality on Wages." *Review of Economics and Statistics*, pp.665-671.

Ashenfelter, O., Harmon, C., & Oosterbecek, H. (1999) "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias." *Working Paper #425, Princeton University, Industrial Relations*.

Berlin, J., Begg, C., & Louis, T. (1989) "An Assessment of Publication Bias Using a Sample of Published Clinical Trials." *Journal of the American Statistical Society*, Vol. 84, No. 406, pp. 381-92.

Betts, J. (1995) "Does School Quality Matter? Evidence From the National Longitudinal Survey of Youth," *The Review of Economics and Statistics*, Vol. 77, No. 2, pp. 231-251.

Blau, F., & Kahn, L. (2000) "Do Cognitive Test Scores Explain Higher U.S. Wage Inequality?" *Preliminary Draft*.

Card, D., and Krueger, A. (1992a) "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, Vol. 100, pp. 1-40.

Card, D., and Krueger, A. (1992b) "School Quality and Black-White Relative Earnings: A Direct Assessment," *Quarterly Journal of Economics*, Vol. 107, No. 1, pp. 151-200.

Card, D. & Krueger, A. (1995) "Time Series Minimum Wage Studies: A Meta-Analysis," *AEA Papers and Proceedings*, Vol. 85, No. 2, pp. 238-243.

Card, D. & Krueger, A. (1996) "School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina," *Journal of Economic Perspectives*, Vol. 10, pp. 31-50.

Dolan, R.C. and Schmidt, R.M. (1987) "Assessing the impact of expenditure on achievement: some methodological and policy considerations," *Economics of Education Review*, Vol. 6, pp. 285-299.

Doucouliagos, C. (1995) "Worker Participation and Productivity in Labor-Managed and Participatory Capitalist Firms: A Meta-Analysis." *Industrial and Labor Relations Review*, Vol. 49, No. 1, pp. 58-77.

Eide, E. & Showalter, M. (1998) "The Effect of School Quality on Student Performance: A Quantile Regression Approach," *Economics Letters*, Vol. 58, pp. 345-350.

Ehrenberg, R.G. & Brewer, D.J. (1994) "Do School and Teacher Characteristics Matter? Evidence from High School and Beyond," *Economics of Education Review*, Vol. 13, No. 1, pp. 1-17.

Ehrenberg and Brewer (1995) "Did Teachers' Verbal Ability and Race Matter in the 1960's? Coleman Revisited," *Economics of Education Review*, Vol. 14, No. 1, pp. 1-13.

Ferguson, R. (1991). "New Evidence on How and Why Money Matters," *Harvard Journal on Legislation*, Vol. 28, pp. 465-498.

Ferguson, R. (1991) "Paying for Public Education: New Evidence on How and Why Money Matters," *Harvard Journal on Legislation*, Vol. 28, pp. 465-498.

Glass, G.V. & Smith, M.L. (1979) "Meta-Analysis of Research on Class Size and Achievement," *Educational Evaluation and Policy Analysis*, Vol. 1, No. 1, pp. 2-16.

Glass, G., McGaw, B., & Smith, M. (1981) *Meta-Analysis in Social Research.* Beverly Hills: Sage Publications.

Goldhaber, D., and Brewer, D. (1997) "Why Don't Schools and Teachers Seem to Matter?" *The Journal of Human Resources*, Vol. 32, No. 3, pp. 504-523.

Greene, W. (1990). *Econometric Analysis.* Prentice Hall, New Jersey.

Greenwald, R., Hedges, L., and Laine, R. (1996) "The Effect of School Resources on Student Achievement," *Review of Educational Research*, Vol. 66, No. 3, pp. 361-196.

Grogger, J. (1996) "School Expenditures and Post-Schooling Wages: Evidence from High School and Beyond." *Review of Economics and Statistics*, Vol. 77, No. 2.

Hanushek, E. (1986) "The Economics of Schooling," *Journal of Economic Literature*, Vol. 24, pp. 1141-1177.

Hanushek, E. (1996) "A More Complete Picture of School Resource Policies," *Review of Educational Research*, Vol. 66, No. 3, pp. 397-409.

Hanushek, E. (1997) "Assessing the Effects of School Resources on Student Performance: An Update," *Educational Research and Policy Analysis*, Vol. 19, No. 2, pp.141-164.

Hanushek, E. (1999) "Some Findings From an Independent Investigation of the Tennessee STAR Experiment and From Other Investigations of Class Size Effects," *Educational Evaluation and Policy Analysis*, Vol. 21, No. 2, pp. 143-163.

Hanushek, E., & Rivkin, S. (1996) "Understanding Twentieth Century Growth in US School Spending," *Journal of Human Resources*, Vol. 32, No. 1, pp. 35-67.

Hanushek, E.A., Rivkin, S.G., & Taylor, L.L. (1996) "Aggregation and the Estimated Effects of School Resources," *Review of Economics and Statistics*, Vol. 77, No. 2, pp. 611-628.

Harris, D. (2000) "Meta-Regression with Standard Errors, Coefficients of Variation, and Other Statistics," *Working Paper*.

Hedges, L. (1992) "Modeling Publication Selection Effects in Meta-Analysis," *Statistical Science*, Vol. 7, No. 2, pp. 246-255.

Hedges, L. & Olkin, I. (1985) *Statistical Methods for Meta-Analysis*. New York: Academic Press.

Hedges, L. & Stock, W. (1983) "The Effects of Class Size: An Examination of Rival Hypotheses," *American Educational Research Journal*, Vol. 20, No. 1, pp. 63-85.

Jarrell, S.B. & Stanley, T.D. (1989) "Meta-Regression Analysis: A Quantitative Method of Literature Surveys," *Journal of Economic Surveys*, Vol. 3, No. 2, pp. 161-170.

Jarrell, S.B. & Stanley, T.D. (1990) "A Meta-Analysis of the Union-Nonunion Wage Gap," *Industrial ad Labor Relations Review*, Vol.44, No. 1, pp. 54-67.

Krueger, A. (1999) "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics*, Vol. 114, No. 2, pp. 497-532.

Manski, C.F. (1987) "Academic Ability, Earnings, and the Decision to Become a Teacher: Evidence from the National Longitudinal Study of the High School Class of 1972," In *Public Sector Payrolls*, ed. David A. Wise. Chicago: University of Chicago Press.

Mciverin, J.D., Gilman, D., & Tillitski, C. (1989) "A Meta-Analysis of the Relation Between Class Size and Achievement," *Elementary School Journal*, Vol. 90, pp. 47-56.

Molnar, A., Smith, P., Zahorik, J. Palmer, A, Halbach, A., & Ehrle, K. (1999) "Evaluating the SAGE Program: A Pilot Program in Targeted Pupil-Teacher Reduction in Wisconsin," *Educational Evaluation and Policy Analysis*, Vol. 21, No. 2, pp. 165-177.

Neumark, D. & Wascher, W. (1998) "Is the Time Series Evidence on Minimum Wage Effects Contaminated by Publication Bias?" *Economic Inquiry*, Vol. 36, pp. 458-470.

Nye, B., Hedges, L., and Konstantopoulos, S. (1999) "The Long-Term Effects of Small Classes: A Five-Year Follow-Up of the Tennessee Class Size Experiment," *Educational Evaluation and Policy Analysis*, Vol. 21, No. 2, pp. 127-142.

Olson, C.A. (1998) "High School Inputs and Labor Market Outcomes for Male Workers in Their Mid-Thirties: New data and New Estimates from Wisconsin," *Preliminary working paper.*

Olson, C.A. (1999) "High School Inputs and Labor Market Outcomes for Male Workers in Their Mid-Thirties: New data and New Estimates from Wisconsin," *Preliminary working paper.*

Pudney, S., Deadman, D. & Pyle, D. (2000) "The Relationship Between Crime, Punishment, and Economic Conditions: Is Reliable Inference Possible when crimes are under-reported?" *Journal of the Royal Statistical Society*, Series A, Vol. 163, No. 1, pp. 81-97.

Sander, W. (1999) "Endogenous Expenditures and Student Achievement," *Economics Letters*, Vol. 64, pp. 223-231.

CHAPTER 3

Optimal School Inputs [1]

I. Introduction

Spending for primary and secondary education in the U.S. in 1996 was approximately $280 billion, or 4 percent of GDP.[2] Real school resources have been increasing steadily in recent years, allowing schools to decrease class sizes and increase teacher salaries. President Clinton has proposed federal subsidies to further reduce class size, similar to the $1 billion per year program recently implemented by the State of California.

These changes in the level and allocation of resources have been going on for several decades. Table 1 reveals that the level of real resources going toward instruction has increased significantly, but that non-instructional expenditures have grown even more. This reflects large increases in spending for special education, support services, and administration.

---

[1] The author wishes to thank Scott Adams, Dale Ballou, Ron Fisher, Gerhard Glomm, John Goddeeris, David Plank, Jeff Wooldridge, and participants in the MSU Labor Economics Seminar for valuable comments on this paper. Financial support was provided by the U.S. Department of Education's North Central Regional Education Laboratory.

[2] See U.S. Census, 1996.

Table 1: School Resource Allocation in U.S. States, 1960 and 1990
(1997 dollars)

Year	Operating $ per Pupil	Class Size	Teacher Daily Wage	Proportion of Total Spending		
				Teacher Salaries And Benefits	Other Instructional	Administration Other Costs
1960	$2,593	25.8	$152	0.68	0.07	0.25
1970	$4,273	22.3	$191	0.67	0.11	0.22
1980	$5,189	18.7	$176	0.60	0.15	0.25
1990	$7,070	17.2	$225	0.61[3]	0.12	0.27

Source: U.S. Department of Education, *Digest of Education Statistics* (1999).

These and other school reforms are responses to general public criticism and reports, such as the *Coleman Report* (1966) and *A Nation at Risk* (1983). However, as Hanushek points out, "it is startling how little any of the reports, or the reform movement itself, draw upon economic principles in formulating new plans" (1996, p.29). One possible explanation is that economists have produced relatively little direct evidence about the best mix of education resources. While there is some evidence that incentive-based plans have modest effects on school efficiency[4], there is little useful research regarding the level and allocation of resources. How much should we spend? What portion should go to hiring more teachers and reducing class size? What salaries should be paid to teachers? The purpose of the current research is to address these and other questions.

Economists have published an extremely large number of studies relating to education. For instance, the literature on "education production functions" (EPF) and "returns to schooling" (RTS) includes hundreds of articles that estimate the statistical relationship between various school inputs and student outcomes, e.g., the change in

[3] This number is estimated because information about teacher benefits was not available for 1990.

students' future wages associated with a $1,000 increase in per-pupil expenditures. These coefficients provide estimates of the marginal product of school inputs that are essential to calculating *optimal inputs*, i.e., the point at which marginal benefit equals marginal cost.

It has been difficult to draw conclusions from this literature because coefficients are often insignificant and coefficient magnitudes vary widely. Many estimates even imply that more resources lead to worse outcomes. Meta-analyses by Hanushek (1986, 1994) and Greenwald, Hedges, and Laine (1994a, 1994b) were not able to reconcile these disparate results. However, the meta-analysis in chapter 2 above provides a different perspective. The paper reveals that 47-57 percent of the variance in EPF coefficient estimates can be explained by methodological differences, such as omitted variables. In addition, the paper uses results from random-assignment experiments on class size reductions, which are shown to be quite robust, to identify the effects that various inputs have on student outcomes. The results suggest that properly specified models are more likely to yield coefficients that reflect positive marginal products from school inputs that are also statistically significant.

The estimates from chapter 2 are used here in a calibrated Cobb-Douglas production function to calculate the marginal benefits of school inputs. This function is incorporated into a partial equilibrium simulation model in which the government maximizes social welfare, subject to the technology constraint. One of the main benefits from schooling is human capital, as measured by students' future wages. The primary cost of education is teacher time, which can increased either by hiring more teachers, thus decreasing class size, or by increasing teacher salaries. Other costs are also considered.

---

[4] See, for example, Hoxby (1994, 1997).

The increased role of U.S. courts in determining funding implies that the government may make decisions about resource levels in a first stage process. A second stage decision may then be made to allocate those resources among various inputs. To account for this possibility, the above welfare maximization problem is solved with various constraints on resource levels. This approach also has the advantage of showing how resource allocation should vary across schools, districts, and states that actually face these different constraints. For instance, it may be that higher resource levels imply higher portions going toward teacher salaries.

There are at least two potential problems in calculating the marginal benefit function. First, there is still uncertainty about the appropriate production function parameter values, even with recent advancements by chapter 2. Therefore, calculations are provided regarding how large the parameters would have to be to justify particular reforms, allowing for useful comparisons with the hundreds of empirical EPF estimates.[5] Second, even if the parameters were known, there may be gains that are not captured in human capital. For instance, society may value academic ability beyond what it contributes to productivity in the workplace. Therefore, the cost minimizing approach to improving test scores is also calculated. Attempts are also made to account for external benefits, including decreased crime and improved public health, which have been highlighted by Haveman and Wolfe (1984, 1994), Cohen (1998), Levin (1989), and others.

[5] Hanushek states that the effects of smaller class sizes "will be completely overshadowed by differences in teacher quality" (1997, p.144), and that "there is a good case for holding overall spending constant" (1996, p.30). Betts (1995, p.235) makes similar statements. However, these statements apparently rely on back-of-the envelope calculations rather that do not account for the key costs and benefits involved. A key purpose of this research is filling this void in the literature.

The three economic models employed here are explained in section II. In section III, key parameter values and model calibrations are discussed. The calibrated models are used in section IV to address several questions: First, what is the optimal level and allocation of education inputs? Second, how large would the production function parameters have to be for current input levels to be optimal? Third, how does the optimal allocation vary based on exogenously imposed resource level constraints? Fourth, what is the cost-minimizing method of achieving specific improvements in test scores?

II. Models

II.A. The Government's Problems. Suppose that a central government maximizes a social welfare function, and that the optimal government choice is to provide full public funding for education. Further, suppose that impact of other markets on the government's optimal behavior occurs only through constant shadow prices, so that the government's decisions regarding education can be separated from decisions about health care and other goods.[6] Thus, one of the government's objectives is maximizing the present discounted value of the net benefits to education. Specifically, the government solves

$$\max_{O,\ Tsal,\ CS} \sum_{t=0}^{\infty} (1-\delta)^t (TB - TC)$$

$$\text{s.t.} \quad TB = TB(CS,\ Tsal,\ O,\ \theta,\ \phi_B)$$
$$TC = TC(CS,\ Tsal,\ O,\ p,\ \phi_C),$$

[6] Many of the assumptions in this partial equilibrium setting are similar to those in cost-benefit analysis. See, for example, Levin and McKewan (2000), Mishan (1988), and Layard and Glaister (1994) for detailed explanations. A key difference between standard cost-benefit analysis and this paper is that the former usually focuses on fixed projects, e.g. reducing class size by five students per class. Rather than yielding optimal decisions, this approach provides information only about very specific options.

where total benefits *TB* are a function of class size *CS*, teacher salaries *Tsal*, other expenditures *O*, a vector of education production function parameters θ, and other parameters ϕ_B. Total costs *TC* are defined similarly with the addition of input shadow prices *p*. Total benefits and total costs are discounted by δ. Neither are indexed according to time *t*, reflecting the fact that only permanent changes in policy are considered and that technology does not change.[7] Issues relating to the transition path are discussed later.

The government may set funding levels (total costs) first and then find the optimal allocation. The second stage problem is

$$\max_{O,\ Tsal,\ CS} \sum_{t=0}^{\infty} (1-\delta)^t TB$$

$$\text{s.t.} \quad TB = TB(CS, Tsal, O, \theta, \phi_B)$$
$$\overline{TC} = TC(CS, Tsal, O, p, \phi_C).$$

The government may also set a desired level of academic achievement and then solve for the optimal allocation of resources, yielding the problem

$$\min_{O,\ Tsal,\ CS} \sum_{t=0}^{\infty} (1-\delta)^t TC$$

$$\text{s.t.} \quad \bar{a} = a(CS, Tsal, O, \theta)$$
$$TC = TC(CS, Tsal, O, p, \phi_C),$$

where academic achievement is produced according to the production function $a(\cdot)$.

[7] See Brewer, et al for predictions of future changes for many school variables, such as the number of students.

All three government problems share the following characteristics and caveats. First, discounting will play a very important role, since many of the effects occur far into the future. For instance, labor market outcomes occur as many as 60 years after the program starts (from kindergarten to retirement from work). Compounding this is the fact that benefits are considered here for student cohorts who have not yet entered school, pushing the costs and benefits even further into the future.

Second, marginal policy changes may be difficult to make in practice. Suppose a state government decides to decrease average class size in the third grade from 26 to 18 students. If an individual school has exactly 26 third graders, then the likely reduction would involve splitting the class into groups of 13. This means that many reforms that may seem simple at the state and national levels are "lumpy" at the school level. This problem is not as great as it appears for policies that include local flexibility, allowing for averaging across grades and schools.[8] However, it would be a factor with more rigid policies.

Third, the education production function parameters are best interpreted as average effects, which vary across students. For instance, there is evidence that minority students benefit from class size reduction more than white students (Nye, Hedges, and Konstantopoulos, 1999, and Harris, 2000a). Therefore, any estimate of the effect size is simply the average effect for those students and schools in the sample. The current

[8] The rules of the policy will certainly affect the costs of reforms, as indicated by Brewer, et al (1999). These rules can be accounted for directly in the simulations. It may also be difficult for higher levels of government to induce desired behavior by lower levels of government. For instance, if the federal government attempted to increase teacher salaries with categorical grants, that money may simply displace local funds for the same purpose, resulting in no net increase teacher salaries. For simplicity, it is assumed throughout the analysis that the cost of enforcing policy is zero. This assumption is

analysis might then conclude that the average actual class size is nearly optimal even if the optimal class size in each school is sub-optimal, or vice versa. Some analysis is presented later to determine whether these scenarios are likely.

II.B. The Production Function. A basic premise of this research is that school inputs contribute to the formation of human capital, which improves welfare. Specifically, it is assumed that output is a function of human capital h and physical capital K, such that

$$y = K \cdot h^{\phi} = K \cdot (r_1^{\theta_1} \cdot r_2^{\theta_2})^{\phi} \tag{1}$$

where r_1 and r_2 represent school inputs (resources) and $K>0$ is a constant. The parameters ϕ, θ_1, and θ_2 reflect decreasing returns to scale ($0<\phi, \theta_1, \theta_2 <1$). This means that output displays decreasing returns in human capital, and human capital displays decreasing returns in school inputs.[9] The model used here does not identify the three parameters separately, focusing instead on the products $\phi\theta_1$ and $\phi\theta_2$.

II.C. Resource Levels. There are two ways to calculate optimal resource levels in the unconstrained welfare maximization problem: (a) using total expenditures only; and (b) separating total expenditures into the three input categories (class size, teacher

---

unlikely to significantly affect the results unless the required enforcement costs comprise a large portion of expenditures.

[9] Heckman et al (1996) focus solely on human capital h and support the use of the function $h_t = h_{t-1}(1+\theta r)$ that reduces to $h_t=h_0(1+\theta r)^a$ after recursive substitution, where a is years of educational attainment. However, this function displays increasing returns to scale in school inputs, which presents problems in solving the three optimization problems. In addition, there is empirical evidence of diminishing returns, contradicting their assumption. See, for instance, chapter 2 and Olson (1999).

salaries, and others). If the EPF estimates are accurate, then the first "single input" approach should produce the same results as the second "multiple input" approach.

The single input approach implies that total costs are solely a function of expenditures per pupil E. The present discounted value of E for all current and future cohorts is

$$TC(E) = \left[\sum_{f=-\infty}^{0} \sum_{g_1=0}^{12} (1-\delta)^{g_1 - f} + \sum_{c=1}^{12} \sum_{g_c=c}^{12} (1-\delta)^{g_2 - c} \right] \cdot S \cdot MCF \cdot E, \qquad (2)$$

where S represents the number of students and MCF is the marginal cost of funds, which reflects the efficiency loss of taxation required to fund education.[10] The two terms in brackets refer to the summing and discounted of costs across cohorts and grades. Student cohorts currently in school are indexed by $c = 1 \ldots 12$. Cohorts that will enter school in the future are indexed by $f = -\infty \ldots 0$. All students are assumed to finish thirteen grades, which are indexed by $g = 0 \ldots 12$. Some current cohorts would only receive these resources for a few years because they would have already started school at the time of policy implementation, therefore, the discounting depends on grade levels and cohort numbers (indexed by c).

When $c=12$, this refers to the $c(12)$ cohort, which is the group of students starting the twelve grade in the current time period. Permanent changes in spending policy will have little effect this group because they will leave school the following year. The $c(0)$ cohort is the group of students starting kindergarten in the current period, while $c(-1)$ refers to the group starting kindergarten in the next period, and so on.

Total benefits from expenditures are

$$TB(E) = \left[\sum_{f=-\infty}^{0} \sum_{t_1=13}^{59} (1-\delta)^{t_1-f} + \sum_{c=1}^{12} \sum_{t_2=13-c}^{59-c} (1-\delta)^{t_2-c} \right] \cdot S \cdot (K \cdot E^{\theta_E}) \tag{3}$$

where the EPF estimate for expenditures is θ_E. It is assumed that students begin work after they graduate from high school and continue until age 65. Assuming students finish school at age 18, and start work at age 19, this implies a career of 46 years from period 13 through period 59. The parameter K in the production function is calibrated to the U.S. economy, as discussed in section III.[11]

It is assumed above that all students start work after high school and work every year, yet many students will instead choose college or other options. The decision to enter college would not affect the analysis here if the higher education market were perfectly competitive, i.e., if the present discounted value of future earnings for each worker did not depend on whether the student chose college or work. While this assumption appears unrealistic, the costs and benefits involved in higher education occur far in the future and are probably small relative to those for the thirteen years of primary and secondary education.[12]

---

[10] The MCF is formally defined as the change in consumer welfare divided by the change in government revenue.

[11] The above equation is a slight simplification of the actual equation. The effects of the permanent policy on those students currently in school are smaller than for future cohorts. This is accounted for in the simulation by assuming that these cohorts received current average resource levels in previous grades.

[12] Education may also increase productivity in non-work activities, increasing utility for people even when they are not working. Again, this implies that the decision not to attend college may not have a significant impact on the results.

II.D. Resource Allocation to Class Size and Teacher Salaries. Schools can allocate resources to various types of inputs. Instructional expenditures refer to those going toward teacher salaries and benefits, in addition to "other instructional" spending, such as textbooks and computer software that are used in classrooms. Table 1 reveals that the proportion of total spending going toward instruction has decreased somewhat over the past several decades, but that the proportion of instructional spending going toward teacher compensation has decreased substantially. The numbers in table 1 imply that an extra $12 billion per year would now be provided for teacher compensation if the spending proportions had remained the same as that in 1960.

The changes indicated above are somewhat difficult to justify based on the considerable evidence that teacher quality is a key, if not primary, determinant of student outcomes.[13] However, even if teacher quality were the key factor, the appropriate level and proportion of spending going for teacher compensation would still be unclear. A key purpose of this research is to fill this void in the existing literature.

There are two ways to calculate the optimal level of resources, as described earlier. There are also two ways to calculate the optimal *allocation*: (a) choosing all inputs simultaneously; and (b) fixing some inputs at specified levels, while solving for the optimal level of remaining inputs. The simultaneous approach appears to be more reasonable, given the interrelationships between the inputs. However, this problem is also somewhat difficult to solve, as discussed later. Both approaches are attempted, though the equations below relate to the second approach.[14]

---

[13] See, for example, Sanders (1999) and Ehrenberg and Brewer (1995).

[14] These equations are easily adapted to the simultaneous version by summing the total costs functions, adding variables to the production function, and re-calibrating.

It is impossible to separate the decisions about class size from those about teacher salaries. Smaller classes require more teachers and, therefore, higher teacher salaries, unless there is a teacher surplus. The salary "necessary" required to attract a sufficient number of teachers at any given class size is labeled $Tsal_n$. Education production function estimates consistently show that the cognitive ability of teachers has an impact on student performance, and changing salaries alters both the number and the ability of teachers who apply.[15] Therefore, it may also be desirable to raise teacher salaries above $Tsal_n$. Such "discretionary" changes are labeled $Tsal_d$. Total teacher salary $Tsal$ is the sum of these two salary components.

The measurement of class size *CS* is somewhat complicated by the common use of the student-teacher ratio (*S-T*) in empirical work. There are at least two reasons why this ratio may not accurately capture class size: First, it is possible for classes to include teacher aides, in addition teachers. This may effectively reduce class size even though the number of students in the room remains unchanged. Second, teachers use some school time for activities outside the classroom. Therefore, I include a parameter α to account for the difference between the two, implying that the number of teachers equals

$$T(CS) = \frac{S}{\alpha \cdot CS}, \tag{4}$$

which is simply a re-arrangement of the student-teacher ratio.

---

[15] Harris (2000) considers teacher standards as another option for increasing teacher ability. See chapter 2, Brewer et al (1999), Ehrenberg and Brewer (1994) and Ballou (1996) for more detail on the effects of both teacher salaries and teacher test scores. Section III in this paper includes discussion of some of this literature.

The labor supply elasticity ε_{TS} implies that

$$\frac{dTsal_n}{Tsal_n} = \frac{dT}{T \cdot \varepsilon_{TS}} \qquad (5)$$

where $Tsal_n$ represents the teacher salary required to maintain current average teacher ability. Existing research indicates that as many as 40 percent of people trained to be teachers actually teach for less than five years in their careers (Ballou, 1996). Also, unions may increase wages above market clearing levels. These facts imply that there may be a surplus of workers who cannot find a teaching job at the current wage.[16] It is unclear how large such a pool might be in reality. However, to account for the effect, the right hand side of (6) is multiplied by a parameter λ, assuming that the surplus is a fixed proportion of teacher employment.

Taking the first derivative of (4) with respect to CS yields dT/dCS, which can be substituted for dT in (5). Multiplying (5) by $Tsal_n$ and λ then yields

$$\frac{dTsal_n}{dCS} = \left[\frac{-S}{\alpha \cdot CS^2}\right] \cdot \lambda \cdot \frac{Tsal_n}{T \cdot \varepsilon_{TS}} \qquad (6)$$

which represents the required change in teacher salary associated with a small change in class size. Equation (6) is used to define the function for total teacher salary,

$$Tsal(CS) = Tsal_d + Tsal_n = Tsal_d + Tsal_0 + \frac{dTsal_n}{dCS} \qquad (7)$$

[16] Most of the rest of this group has probably left the profession after teaching for a time and realizing they do not wish to be teachers. These workers would obviously not be

where $Tsal_0$ is the current teacher salary.

Below, the above equations are used to create the cost and benefit functions. Again, since class size and teacher salaries are inseparable, the equations include both inputs. A category for school inputs "other than teachers," labeled O, is also added to complete the accounting identity, forcing total expenditures to be equal to the sum of the three separate inputs. The total cost of a given class size and teacher salary is

$$TC(CS,Tsal,O) = MCF \cdot \sum_{t=0}^{\infty}(1-\delta)^t \cdot [T(CS) \cdot Tsal(CS) + O \cdot S] \tag{8}$$

(Note that this function is somewhat simpler than that for expenditures. The reason is that all the variables are expressed in the form commonly reported: expenditures *per pupil*, other inputs *per pupil*, and salary *per teacher*.)

Total benefits from the three inputs are represented by

$$TB(CS,Tsal,O) = \left[\sum_{c_1=-\infty}^{0}\sum_{t_1=13-c_2}^{60-c_2}(1-\delta)^{t_1} + \sum_{c_2=1}^{12}\sum_{t_2=13-c_2}^{60-c_2}(1-\delta)^{t_2}\right] \cdot S \cdot \left[B \cdot CS^{\theta_{CS}} \cdot Tsal^{\theta_{Tsal}} \cdot O^{\theta_O}\right] \tag{9}$$

where the EPF parameters are represented by θ_{CS}, θ_{Tsal}, and θ_O, respectively.

One school input that has not been mentioned is physical capital, primarily classroom space. One might expect that each additional class would require one additional classroom. This probably overstates actual classroom costs because there are

---

considered a part of any teacher surplus pool because they are unwilling to work at the often extra classrooms available that have little economic value. For instance, there may be safety risks to renting out open space to non-school organizations. In addition, capital costs comprise only 10 percent of total education spending.[17] Therefore, classroom costs are initially left out of the current analysis. If the results had shown that class size reductions were too large, then it would have been necessary to include capital costs in our calculations. However, the results presented below suggest just the opposite. Adding capital costs to the model would make class size reductions compare even less favorably with increases in teacher salaries.

III. Parameters

III.A. Education Production Function Parameters. Table 2 below summarizes the results of chapter 2 for the effects of each school input on student test scores and wages. Two sets of numbers are reported for each: The "gold standard" estimates are those that are found to be preferred, based on comparisons with experimental evidence.

For the test score sample, the coefficients reported in table 2 reflect the change in test scores (measured in standard deviation units) associated with a unit change in the school input. For instance, the mean effect of expenditures implies that a $100 increase in expenditures per pupil would increase test scores by 0.067 standard deviations.

---

current wage.

Table 2: Summary of EPF Estimates from Chapter 2

EPF STATISTIC	VALUE	TEST SAMPLE		WAGE SAMPLE	
		Gold	Other	Gold	Other
Expenditure/Pupil Point Estimate (x100)	Mean	+0.067	+0.014	+0.074	+0.014
	Min	+0.045	- 0.024	+0.074	-0.096
	Max	+0.078	+0.065	+0.074	+0.117
Abs. Teacher Salary Point Estimate (x100)	Mean	+0.002	+0.013	+0.266	+0.028
	Min	- 0.022	- 0.022	+0.008	+0.007
	Max	+0.029	+0.048	+0.523	+0.179
Rel. Teacher Salary Point Estimate	Mean	NA	- 0.308	+0.872	+0.098
	Min		- 0.607	+0.853	- 1.014
	Max		- 0.122	+0.942	+1.668
Class Size Point Estimate	Mean	- 0.017	- 0.012	- 0.016	- 0.055
	Min	- 0.023	- 0.395	- 0.021	- 0.685
	Max	- 0.010	+0.208	- 0.010	+0.400

For the wage sample, the coefficients reflect the percentage change in wages associated with a unit change in the school input. Therefore, at the mean value, a one hundred dollar increase in expenditures per pupil would increase students' future wages by 7.4 percent. This seems somewhat large, therefore, some comparisons may be useful: First, how large would the effect on wages be if a $100 increase in per pupil expenditures were all put into teacher salaries? With a pupil-teacher ratio of 23, the increase in expenditures could increase teacher salaries by $2,300 per year. At a base salary of $34,500, and average wages in other profession of $37,800, an increase in teacher salary of $2,300 would increase relative teacher salary from 0.91 to 0.97. According to table 2, this translates into a 0.052 percentage increase in future student wages (or 0.06 times 0.872), which is remarkably close to the expenditure coefficient of 0.074.

[17] According to the U.S. Census, 1996, total state and local spending on capital was $28 billion, while total expenditures, including capital, totaled $280 billion.

Recall that the values in table 2, and the discussion above, represents only a starting point, given uncertainty about actual parameter values. Later sections include estimates of how large actual parameters would have to be to justify current resource choices. For now, the parameter values discussed in the above paragraph are used as the base values in the simulations. Economic theory provides some justification for focusing on the relative teacher salary parameters, since absolute salaries ignore compensating differentials and opportunity cost. Thus, relative salaries are likely to produce less biased results than absolute teacher salary. In addition, the coefficients on relative teacher salaries in chapter 2 are more robust, and the average magnitude of the coefficient on absolute teacher salary is extraordinarily large.[18]

Table 3 below translates the mean coefficients from table 2 into elasticity forms, which are required for the Cobb-Douglas education production function. For the test score sample, these are not true elasticities because the interpretation would involve percentage changes in standard deviations. Therefore, instead, the test score "elasticities" reflect the standard deviation change in test scores associated with a one percent increase in the school input. The last two columns reflect true elasticities, i.e., the percentage change in the student's future wages associated with a one percent increase in the school input.

Table 3 also includes some differences in the teacher salary coefficients compared with table 2. As stated earlier, the relative teacher salary coefficients appear to be more reliable, therefore, the "absolute teacher salary estimates" reflect the relative teacher

[18] The average of 0.256 implies that a $100 increase in teacher salary leads to a 25.6 percent increase in student future wages.

salary values transformed into an absolute form. All transformations are performed at the mean values.

Table 3: Transformed EPF Estimates

	FORM	TEST SAMPLE	WAGE SAMPLE
Expenditure/Pupil Estimates (x 100)	Table 3	+0.067	+0.074
	Elasticity	+0.045	+0.050
Abs. Teacher Salary Estimates (x 100)	Table 3	+0.002	+0.007
	Elasticity	+0.010	+0.035
Class Size Estimates	Table 3	-0.017	-0.016
	Elasticity	-0.004	-0.004

There are at least two ways to measure the relationship teacher salaries and teacher productivity. The *direct approach* involves estimating the relationship between teacher salaries and student outcomes, e.g., coefficients from the EPF literature. The *indirect approach* involves a two-step procedure: First, estimates are made of the relationship between teacher salaries and some other measure of teacher productivity, such as teacher verbal ability. Second, estimates can be made of the relationship between this measure of teacher ability and student outcomes, a relationship that can be combined with the results from the first step. For instance, Manski (1987) finds that a one percent increase in the teacher's wage yields a 0.0098 standard deviation increase in teacher SAT scores.[19] Figlio (1997) finds similar results, measuring teacher ability with the selectivity of the students' undergraduate college. He finds that a one percent increase in metropolitan teacher salary is associated with a 1.58 percent increase in the probability of

[19] The average SAT scores is approximately 1003 with a standard deviation of 100. Therefore, a ten percent increase in the wage would increase ability by 0.098*100=9.8 points, yielding a new score of 1012.8.

attracting a teacher who graduated from a highly selective institution. Based on the average SAT scores of highly selective colleges, this translated into a parameter value of 0.0022, which is somewhat lower than Manski's estimate.[20] (This difference may be caused by differences in tastes and constraints of low ability students who are excluded from the Figlio estimates, but included by Manki.)

The second step in the indirect approach requires an estimate of the relationship between teacher test scores and student test scores. In other words, how much do the skills measured by these tests help teachers in developing those same skills in their students? Chapter 2 discusses 26 estimates of this relationship. The mean effect is 0.20, meaning that a one standard deviation increase in teacher test scores produces a 0.20 standard deviation in student test scores. Combining this with the results of the previous paragraph, a one percent increase in teacher salaries should increase student ability by 0.0019 standard deviations (or 0.2 multiplied by 0.0098). This is almost identical to the mean estimate in table 3, which used the direct approach.

III.B. Teacher Opportunity Cost. In a simple competitive economy with many firms and homogeneous actors, the opportunity cost equals marginal product and the market wage. However, the market for teachers clearly differs from this simple model: (1) teaching includes compensating differentials that are quite different from other professions, especially the low number of working hours per year; (2) workers differ in labor productivity and these differences are hard to observe; (3) labor unions negotiate wages that are not related to productivity and that extract economic rents; and (4)

[20] In addition, Ballou and Podgurksy (1994) report a baseline ability-salary coefficient of 0.032, while Ballou and Podgursky (1992) report a baseline of 0.0028.[20] The average of these two estimates is similar to Manski's results. The reasons for the relatively wide range of values from these two authors are complex and are not discussed here.

governments may pay wages that do not equal marginal product, since they do not maximize profits. These factors imply that the teacher wage may not represent teacher opportunity cost. Table 4 provides information about teaching and non-teaching occupations that is suggestive of the appropriate range of values that should be considered in the analysis.

Table 4: Salaries in Teaching and Various Professions[21], 1998

Profession	Annual Income	Hours/ Week	Weeks/ Year	Hours/ Year	Implied Hourly Wage	Wage Adj. For Fringe Benefits
K-12 Teaching	34,801 34,200(1)	44 ---	40 ---	1,760 ---	\$19.77 ---	\$24.71 ---
All	37,803	43	50	2,150	\$17.50	\$21.88
Managerial/ Professional	54,030	44	50	2,200	\$24.56	\$30.70
Technical, Sales, Admin. Support	32,873	42	50	2,100	\$15.65	\$19.56
Licensed Practical Nurses	29,463	41	50	2,050	\$14.37	\$17.96
Officials and Administrators	47,329	43	50	2,150	\$22.01	\$27.51
Accountants and Auditors	42,323	42	50	2,100	\$20.15	\$25.19
Management-related	42,193	41	50	2,050	\$21.09	\$26.36

A large proportion of teachers are women. Two decades ago the occupational choices of women were much more limited (Hanushek and Rivkin, 1996), and it may have been reasonable to use nursing and administrative support as estimates of

[21] All data from the *Current Population Survey*, 1998, and *The Condition of Education*, 1998.

opportunity cost. However, such limits would certainly be inappropriate in the current economy.

Fringe benefits represent approximately 20 percent of total compensation, which is not reflected in the "implied hourly wage."[22] The last column of table 5 is adjusted accordingly. Therefore, a relatively wide range is considered: $18-30 per hour. The base value is $25 because it is near both the teacher hourly wage average and the median of the professions shown above.

III.C. Teacher Labor Supply Elasticity. There are several different types of labor elasticities estimated in published research. The "total elasticity" of supply in teaching describes the economic behavior of all workers, which is the parameter of interest in this model. Teacher supply elasticities can also be calculated for various sub-groups of the labor force. The "certification elasticity" applies primarily to college students who are selecting academic majors. In education, this choice is especially important due to the importance of official certification. The certification elasticities are likely to be greater than total elasticities because there are few costs to changing majors compared with changing professions later in life. This also means that the "retention elasticity," which applies to people who have already chosen teaching as a profession, probably underestimates the total elasticity. Table 5 provides a summary of the literature.

[22] Chamber of Commerce (1985).

Table 5: Summary of Teacher Supply Elasticities

Study	Elasticity Type	Estimation Technique	Alternative Wage	Estimates
Manksi (1987)	Certification	Simulation	Non-teachers	2.34-3.18
Ballou (1997)	Certification	Simultaneous equations Current and lagged wage	College-educated workers	1.32
Zabalza (1979)	Certification	One equation Permanent income	College-educated workers	2.46
Ballou and Podgursky (1994)	Certification Retention	Literature Review	NA NA	1.25-2.00 0.06-0.33
Currie (1991)	Total	One equation District level	Bordering school districts	3.65-5.62

This sample of the literature is useful in establishing a reasonable range of values for the total elasticity. The certification elasticity range is 1.25-3.18. The retention elasticities are substantially lower, as expected. Currie's estimate is the only total elasticity, but it is quite different from the others, in that, the wage observations occur at the district level, and the alternative wage is that for teachers in bordering districts. The other estimates use broader definitions, such as salaries for college-educated workers and salaries for the entire state or region. Also, it appears that the large estimate found by Currie is due to her use of district level data, compared with the county or state levels used in the other studies. It is less costly for teachers (and other workers) to switch jobs across geographically small regions. Thus, Currie's estimate is inappropriate for the state/national analysis performed here.

A final consideration is that all of these estimates are static. Ideally, we would have an impulse response function, indicating the effect of a policy change on teacher behavior in each future period.[23] In reality, it is unclear whether the static estimates are above or below the actual value for any given time span. It is assumed here that the static estimates overestimate the short-run effects and underestimate the long run effects.

The middle of the certification range is 2.22, which is probably too high, as suggested earlier. However, the total elasticity is almost certainly higher than any of the retention elasticities. This implies that a base of 1.0 would be a reasonable base value. The range considered in the sensitivity analysis is 0.5-1.5. Even if the actual parameter were out of this range, it is unlikely that the allocation results would be affected. Any biases in this parameter affect the results in the same direction for each input. (In the case of class size, an overly large elasticity would allow for lower reform costs because it would appear easier to attract new teachers to fill these new positions. The same is true of teacher salaries.) However, the optimal *level* of resources, as well as the cost minimization results, would be affected by any bias in this parameter.

III.D. Calibration. The parameter K in the production function is calibrated to the U.S. economy such that the production function in (1) reproduces 1970 GDP at 1970 input levels and base EPF values. This approach of matching current output with past school inputs is used because the average worker in today's economy was educated in the late 1960's and early 1970's. One weakness of this approach to calibration is that any conclusions about the level of school resources appear to pertain only to the optimality of school inputs in 1970, which is of little direct interest.

There are at least two reasons to justify this approach despite the above problem. First, there is no alternative that is obviously better; for instance, matching current inputs

[23] Such a function might be estimated from a natural experiment, such as an unanticipated, large change in government policy toward teacher salaries.

with expected future outputs would introduce the problem of unpredictable future changes in the productivity of capital and labor, and other macroeconomic factors. (This approach was attempted and produced nearly identical conclusions.) Second, using previous inputs highlights the possibility that recent increases in school inputs may be justified based on improvements in the productivity or labor or school inputs.

There are also several reasons why this problem is not as great as it may seem. First, conclusions using this approach are based on essentially the same political institutions as existed in the 1970's. There is no obvious reason why these decisions would have improved over time. Second, conclusions about the *allocation* of resources could be extended to the current period as long as it is reasonable to assume that the marginal rate of technical substitution is relatively constant.[24]

III.E. Other Parameters. The welfare cost of taxation is one element of total input costs. A common assumption is that taxes are lump sum, which implies under certain circumstances that there is no deadweight loss and the marginal cost of funds (MCF) is one. Empirical estimates of the MCF range from approximately 0.9-1.4 for income and sales taxes (Browning, 1976 and Stuart, 1984). The U.S. Office of Management and Budget uses 1.25, which is used here as the base value.

The parameter α reflects the fact that the student-teacher ratio is not the same as class size. One data set that includes actual class size is the U.S. *School and Staffing Survey* (SASS). This data is used, along with the pupil-teacher ratio, to estimate α. This was done by dividing class size by pupil-teacher ratio for each state, yielding a range of 0.65-0.81. This implies that teachers have more non-class time than do students on average, though the degree differs across states. The weighted average across states is 0.73 and is used as the base value.

A base value for λ was chosen such that the teacher surplus is a constant 10 percent of currently employed teachers. The range for sensitivity analysis is 0-30 percent. Table 6 below summarizes the parameter values discussed in this section.

Table 6: Summary of Parameter Values

PARAMETER	SYMBOL	RANGE	BASE
Discount Rate	δ	0.00 – 0.06	0.03
Teacher's Opp.Cost	Tsal	\$17.96-30.70	\$25
Teachers' Supply Elast.	ε_{TS}	0.5-1.5	1.0
Teacher Surplus	λ	0.7-1.0	0.9
Class size adjustment	α	0.73	0.65-0.81
Marginal Cost of Funds	MCF	0.9-1.4	1.25

The simulations in section IV below use the base values. Sensitivity analysis using the range values is reported in appendix A.

[24] Note that the information provided earlier regarding teacher salaries in 1998 is adjusted in the simulation to account for this focus on past inputs.

IV. Results

IV.A. Maximizing the Social Utility of Education. Three government problems were presented in Section II. The current sub-section deals with the first: Maximizing the social utility of education without restrictions on the level of education resources.

The first phase of the simulation involves solving for the optimal input level. At the base EPF values, the optimal expenditure is approximately $3,000 per pupil. This is considerably lower than the 1970 input levels shown in table 1. Since there is uncertainty about actual EPF values, input levels were also fixed at 1970 levels and the model was solved for the EPF parameter "required" to justify the 1970 input levels. This yielded an estimate of 0.07 compared with a base value of 0.05, as shown in table 7.

The same model was also solved using the multiple input approach where total spending is separated into three categories: class size, teacher salaries, and other inputs. The error minimization routine produced unstable solutions to this problem, depending on starting values. Therefore, the optimal level of each input was calculated by fixing the others at 1970 levels. These results are presented in the right-hand columns of table 7. The parameters required to justify 1970 individual levels are also included, along with the base values used in the simulation.

Table 7: Optimal Inputs and Required Parameters
(Unconstrained Maximization)

		TOTAL EXPEND.	CLASS SIZE	SALARIES
EPF	Base	0.050	-0.004	0.035
	Required	0.090	-0.115	0.066
Input Levels	At Base Parameters	\$3,060	112	\$28,891
	1970 Input Levels [25]	\$4,547	22.3	\$30,500

Notice that the base EPF values are all lower than those required to justify current spending patterns. Put another way, actual input levels all appear to be too large. Policy regarding teacher salaries appears to be closest to the optimum.

The calculations at base parameter values can also be used to make the "multiple input" calculation of total expenditures. (The above discussion focused on the multiple input approach to calculating the *allocation* of resources.) This approach yields total expenditures of \$1,968 per student, approximately \$1,000 less than the one input approach. While clearly different, these are not dramatically different and, therefore, provide some support for the chosen EPF base values.

Sensitivity analysis for table 7 is reported in appendix A. Only one parameter is changed in each experiment with the others remaining at their base values. The parameter ranges given in table 6 are used as the extreme ends. For the EPF parameters, I use ± 50 percent of the parameter value. For instance, the base expenditure coefficient is 0.05 with a range of 0.025-0.075. The results appear to be most sensitive to the EPF parameters and the discount rate. Overall, the sensitivity analysis strongly reinforces the earlier conclusions. The highest optimal input level is still below current levels. In

[25] The cost function yields expenditures per pupil of \$4,547 at 1970 levels of class size and teacher salaries, which similar to the value in table 1. This value is used in table 7.

addition, any given combination of parameter values still implies that resources for teacher salaries are too small.

IV.B. Maximizing Welfare Subject to Exogenous Resource Constraints. This section focuses on the same maximization problem discussed above, except that an exogenous cost constraint is added. Again, there was no closed form solution. However, with the added constraint, the multiple-input results were somewhat more stable in the error minimization routine, therefore, these numbers are reported in Table 8.[26]

Table 8: Optimal Input Results
Maximization with Resource Constraints

INPUT LEVELS	OPTIMAL INPUTS	
	CLASS SIZE	SALARIES
$4,000	36	$37,542
$7,000	30	$63,348
$10,000	26	$82,005

Table 8 indicates that actual class sizes were to small and actual teacher salaries were too small at 1970 total resource levels (approximately $4,000). This is the same conclusion reached earlier in the unconstrained problem.

IV.C. Cost Minimization for Improved Test Scores. There is substantial evidence that academic ability is, at best, a small portion of schooling increases human capital (Manski, 1987). In addition, society's desire to improve academic ability may be separate from its effect on worker productivity. Therefore, the cost minimizing approach to achieving various specific improvements in test scores is also estimated. The results in

[26] The single-input results are not reported for this government problem because this would leave only one unknown. For example, fixing spending levels, teacher salaries, and other inputs, would leave class size as the only unknown. Such calculations would not be very informative.

table 9 below were obtained by, first, finding the required change in inputs necessary to reach the objective, and then calculating the cost of the input change from the cost functions in section II. For instance, an additional $100 in spending per pupil yields a 0.067 standard deviation increase in test scores. Therefore, achieving a 0.1 standard deviation increase in test scores requires $150 per pupil.

Table 9: Input Levels and Costs for Test Score Gains

	TEST SCORE GAIN	TOTAL EXPEND.	CLASS SIZE	TEACHER SALARIES
Current Input Level	---	$4,547	22.3	$30,500
Required Input Level	0.1 σ	$4,697	21.4	$34,500
	0.5 σ	$5,297	19	$53,500
	1.0 σ	$6,047	17	$70,500
Additional Expenditures per pupil	0.1 σ	$150	$435	$200
	0.5 σ	$750	$1,876	$1,000
	1.0 σ	$1,500	$4,652	$2,000

The results are quite similar across the three variables. If these estimates are accurate, this implies that schools are quite efficient in allocating resources for the purposes of improving test scores. The same cannot be said of the results for future wages.

V. Conclusion

The results presented above are quite consistent across the different approaches. First, the results are qualitatively similar for both the unconstrained model (table 7) and constrained models (table 8). Second, the allocation results are similar regardless of whether the dependent variable of interest is student scores or future student wages. (Compare tables 7 and 8 with table 9.)

Several interesting results are obtained by comparing the "required" EPF values with the chapter 2 sample. First, none of the EPF preferred expenditure estimates in the sample is large enough to justify the level of inputs, i.e., not one of them is greater 0.07, which is shown in table 7. Considering the entire sample, including non-preferred estimates, only one expenditure coefficient meets this standard. Since the base EPF expenditure estimate seems unrealistically large, the fact that it is still not large enough is especially striking.

There are several areas in which additional research could be informative. First, the partial equilibrium nature of the model implies that results far away from current values are likely to be inaccurate. However, it is important to keep in mind that the main conclusions here are not based on specific optimal input estimates. Rather, the main conclusions focus on the direction of change, which is consistent across the three approaches. For instance, the optimal class size is probably not 112, as indicated in table 7, but the optimal class size probably is larger than current values. It seems fairly clear that a more complicated general equilibrium model would come to this same general conclusion, given the same parameter values.

Second, it may be that external benefits from education, combined with the benefits of human capital, would justify the current level and allocation of resources. This seems unlikely in the case of class size, due to the large difference between optimal and actual values. However, it is quite possible that conclusions about the level of resources would change if all the benefits of education were included.

Third, structural changes in either productivity of human capital, or the productivity of school inputs, could account for recent increases in real resources. Macroeconomic evidence for the past decade certainly suggests such changes in the case of human capital, though it is far less clear whether schools have become more efficient. These issues are also closely related to the assumption that the results for 1970 school inputs also hold for current school inputs.

Fourth, the role of class size as a compensating differential to teachers is not incorporated into this model. It is possible that this plays a substantial role in teacher supply decisions. Educators often argue that students have become less disciplined, requiring more teacher attention. In addition to taking time away from teaching, this may impose substantial disutility on teachers. Rather than compensating for this through salary, it may indeed be optimal to accomplish this through class size reductions.

Debates about school policy often center on resource levels. This paper helps to advance this debate in three ways: First, Hanushek appears to be quite right that economic principles are lacking from the debate on education policy. However, while it is unwise to ignore economic theory and principles, decisions based on *untested* theories may none the wiser. This paper provides such tests for the appropriate use of school resources.

Second, the paper deals with the literature on education production functions in a new way. This appears to be the first example in which any such parameters have been used to estimate the optimal level and allocation of resources in a formal model. In addition, rather than empirically estimating yet another production function, this paper calculates the values that would justify specific input allocations and levels.

Third, and finally, the paper highlights the type of reasoning that is necessary for appropriate school policy decisions. It is probably true that increased resources increase school quality and this may be part of the solution to many education problems. However, there must be some limit to how far this policy can reach. Education resources compete with health care, income security, emergency services, and various other worthy programs in the public sector, in addition to the food and other consumables that are obtained in the private sector. If the model and results presented here show anything, it is that simple analyses of explained variance and back-of-the-envelope calculations are insufficient to guide these important decisions.

APPENDICES

Appendix A

Sensitivity Analysis

Table 7a below is similar to table 7 in the text, but with varying parameter values. Three of the cells are left blank because changing these parameters has no impact on the results. This may seem somewhat confusing in the case of teacher salaries and the teacher supply elasticity. The reason is that this approach fixes class size at 1970 levels, meaning that no additional teachers are required. Rather, the increase in salaries is potentially beneficial because it may attract better teachers.

Table 7a – Sensitivity Analysis for Optimal Input Levels (Base EPF)

PARAMETER	VALUES	EXPENDITURES	CLASS SIZE	TEACHER SALARIES
Table 7 – Optimal Input	Base	\$3,060	112	\$28,891
EPF	Max Min	\$3,924 \$832	90 166	\$52,429 \$11,921
δ	Max: 0.06 Min: 0.00	\$256 \$5,030	163 70	\$15,831 \$69,437
MCF	Max: 1.40 Min: 0.90	\$1,828 \$2,911	120 94	\$27,658 \$40,110
λ	Max: 1.0 Min: 0.7	---	105 131	\$29,008 \$28,787
εts	Max: 1.5 Min: 0.5	---	142 75	---

The range of values (maximums and minimums) for the EPF parameters are ± 50% from the base values. This range is admittedly arbitrary, but helps to show the effects of parameters on results.

BIBLIOGRAPHY

BIBLIOGRAPHY

A Nation at Risk (1983). National Commission on Excellence in Education, Washington, DC.

Ballou, D. (1996). "Do Public Schools Hire the Best Applicants?" *Quarterly Journal of Economics*, Vol. 111, no. 1, pp. 97-133.

Ballou, D. & Podgursky, M. (1994). "Recruiting Smarter Teachers." *Journal of Human Resources*, Vol. 30, no. 2, pp. 326-338.

Ballou, D. & Podgursky, M. (1992). "Buying Quality: Will Higher Pay Improve the Teaching Workforce?" *Working Paper*, University of Massachusetts at Amherst.

Blau, F., & Kahn, L. (2000) "Do Cognitive Test Scores Explain Higher U.S. Wage Inequality?" *Preliminary Draft.*

Brewer, D.J., Krop, C., Gill, B.P., & Reichardt, R. (1999). "Estimating the Cost of National Class Size Reductions Under Different Policy Alternatives," *Educational Evaluation and Policy Analysis*, Vol. 21, no. 2, pp. 179-192.

Browning, E.K. (1987). "On the Marginal Welfare Cost of Taxation," *American Economic Review*, pp. 11-23.

Cohen, A.M. (1998). "The Monetary Value of Saving a High-Risk Youth," *Journal of Quantitative Criminology*, Vol. 14, no. 1, pp. 5-33.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfield, F.D., & York, R.L. (1966). *Equality of Educational Opportunity.* Washington, DC: U.S. Government Printing Office.

Currie, J. (1991). "Employment Determination in a Unionized Public-Sector Labor Market: The Case of Ontario's School Teachers," *Journal of Labor Economics*, Vol. 9, no. 1, pp. 45-66.

Ehrenberg, R. & Brewer, D. "Did Teachers Verbal Ability and Race Matter in the 1960s?" *Economics of Education Review*, Vol. 14, no. 1, pp.1-21.

Figlio, D. (1997). "Teacher Salaries and Teacher Quality," *Economics Letters*, Vol. 55, pp. 267-271.

Hanushek, E. (1997). "Assessing the Effects of School Resources on Student Performance," *Educational Evaluation and Policy Analysis*, Vol. 19, no. 2, pp. 141-164.

Hanushek, E. & Pace, R.R. (1995). "Who Chooses to Teach (and Why)?" *Economics of Education Review*, Vol. 14, no. 2, pp. 101-117.

Hanushek, E. & Rivkin, S. (1996). "Understanding the Twentieth Century Growth in U.S. School Spending," *Journal of Human Resources*, Vol. 32, no. 1, pp. 35-68.

Harris, D. (2000). "Teacher Standards and Teacher Quality," *Working Paper*, Michigan State University.

Haveman, R.H., and Wolfe, B.L. (1984). "Schooling and Economic Well-Being: The Role of Non-Market Effects," *The Journal of Human Resources*, Vol. 14, no. 3, pp. 377-407.

Haveman, R. & Wolfe, B. (1994). *Succeeding Generations: On the Effects of Investments in Children*. New York: Russell Sage Foundation.

Hedges, L.V, Laine, R.D., & Greenwald, R. (1994a). "Does Money matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes," *Educational Researcher*, Vol. 23, no. 3, pp. 5-14.

Hedges, L.V, Laine, R.D., & Greenwald, R. (1994b). "Money does matter somewhere: a Response to Hanushek." *Educational Researcher*, Vol. 23, no. 4, pp. 5-8.

Hoxby, C. (1994). "Do Private Schools Provide Competition for Public Schools?" *NBER Working Paper No. 4978.*

Hoxby, C. (1997). "Does Competition Among Public Schools Benefit Students and Taxpayers? Evidence from Natural Variation in School Districting," *NBER Working Paper No. 4979.*

Layard, R. & Glaister, S. (1994). *Cost Benefit Analysis*. Cambridge University Press: New York.

Levin, H. & McKewan, P. (2000). *Cost Effectiveness Analysis*. Sage: Beverly Hills.

Levin, H. (1989). "Economics of Investment in Educationally Disadvantaged Students." *American Economic Review*, Vol. 79, no. 2, Papers and Proceedings of the Hundred and First Annual Meeting of the American Economic Association, pp. 52-56.

Manski, C.F. (1987). "Academic Ability, Earnings, and the Decision to Become a Teacher: Evidence from the National Longitudinal Study of the High School Class of 1972." In *Public Sector Payrolls*, Ed. David A. Wise, University of Chicago Press: Chicago.

Mishan, E.J. (1988). *Cost-Benefit Analysis: An Informal Introduction.* 4th Edition, Unwin Hyman Ltd: Cambridge, MA

Nye, B., Hedges, L., and Konstantopoulos, S. (1999) "The Long-Term Effects of Small Classes: A Five-Year Follow-Up of the Tennessee Class Size Experiment," *Educational Evaluation and Policy Analysis*, Vol. 21, No. 2, pp. 127-142.

Sanders, W. & Horn, S. (1998). "Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research," *Journal of Personnel Evaluation in Education*, Vol. 12, no. 3, pp.247-56.

United States Chamber of Commerce (1985). *Employee Benefits.* Washington, DC.

Zabalza, A. (1979). "The Determinants of Teacher Supply," *Review of Economic Studies*, Vol. 46, no. 1, pp. 131-147.