Doctoral Dissertation

Three Essays on the Economics of Education

A dissertation submitted to the

School of Public Policy and Management, H. John Heinz III College
and Tepper School of Business

Carnegie Mellon University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Economics and Public Policy

by

Abby Clay Turner

Dissertation Committee:

Lowell Taylor, Carnegie Mellon University
Maria Marta Ferreyra, Carnegie Mellon University
Dennis Epple, Carnegie Mellon University
David Deming, Harvard University

August 14, 2013

Introduction

This thesis is composed of three essays on the economics of education. Chapter 1 analyzes the effectiveness of a battery of formative assessments, 4Sight, which are broadly aligned with annual assessments required by NCLB. These formative assessments are designed to provide teachers with feedback on student performance throughout the year in order to raise end of year student test scores. Methodologies for evaluating the effect of using 4Sight on test score outcomes are drawn from the program evaluation literature, and include individual and school-level OLS, quartile regressions, and probit regressions, as well as matching at the school-level. Micro-econometric results show that 4Sight had no discernible effect on math scores and a small negative effect on reading scores in its first year in Pennsylvania. Policy recommendations include continuing a smaller-scale trial period of 4Sight for several more years in conjunction with careful, improved alignment between 4Sight and PSSA, ongoing empirical analysis of its effects, and incorporating more opportunity for feedback to teachers and students to improve 4Sight as a formative assessment. Perhaps the most surprising finding of this paper is that while 4Sight has a very small effect on student outcomes, students receiving tutoring under NCLB are significantly and between four and nine percent less likely to pass their exams, even after controlling for all covariates, including eligibility for tutoring. This implies that the publicly-funded tutoring is actually disadvantaging those students who receive it. These results should serve as a cautionary example to school, district, and state-level policy-makers when choosing interventions designed to improve student performance on tests tied to NCLB.

Teacher strikes and the right of public employees to collectively bargain are topics of frequent and heated debate in the public sphere, with little research available to inform the debate. In firms, the negative relationship between labor unrest and reduced productivity is well-documented; the purpose of this study is to explore whether there exists a similar, measurable relationship between labor strife and productivity in public schools. In Chapter 2, I use regression analysis to analyze data that includes teacher strikes and expired contracts over a seven-year period in Pennsylvania, and I find that the pass rates on a district-level cohort's math tests decrease by about 1-2% in the year of a strike and by about 0.5% during a year that teachers work under an expired contract. Additionally, cohorts experiencing a strike during their 11th-grade year realize about a 2% decrease in their graduation rate. In addition to improving upon the methodologies of previous teacher strike papers, this paper distinguishes between productivity loss due to strikes and that due to lengthy ongoing labor disputes that do not necessarily end in strike. Policy implications include making administrators aware of the possible effect of a strike on graduation rates and the need for better collection of data on collective bargaining by state agencies.

The past decade has seen enormous growth in the for-profit higher education industry, and along with it, enormous debate over the relative costs and benefits of such an education. Utilizing the rich data from the NLSY97 Geocode merged with institutional data from IPEDS, in Chapter 3 I empirically analyze data on individuals with two-year degrees, estimate the average marginal earnings gain from a two-year degree, and compare the effects of degrees across institutional sector and across major area of study using OLS with family background and extensive demographic controls. I find evidence of selection at three levels: selection into college, selection into type of college, and selection into major area of study. Any estimates of labor market returns to these degrees will be biased until future research unravels and models these selection mechanisms and processes. This chapter provides a first look into the differential inputs and outputs of for-profit and public two-year degree programs. I find that a two-year degree results in an 8.1 percent average marginal earnings gain over a high-school diploma, and that the sector of the degree-granting institution alone does affect this gain. I also find that earnings gains vary greatly by major; individuals with "academic" degrees experience no significant earnings gains while individuals with "vocational/technical" degrees on average experience a 32.7 percent earnings gain. I find statistical differences in the marginal earnings gains across institutional sector within major fields of study, suggesting that attending a for-profit does matter when major field of study is taken into account. Policy-makers should take note that this preliminary analysis of the returns (which can be thought of as the private benefits) to public and for-profit degrees does not provide unambiguous evidence in favor of one sector over another, but rather a first look into the "black box" in which students, institutions, and major areas of study come together and jointly determine labor market outcomes.

Acknowledgements

Chapter 1 utilizes confidential student achievement information obtained under a signed confidentiality agreement with the Pennsylvania Department of Education. The authors wish to thank Jinxiang Liu for programming and database support. The authors benefited from discussions with Maria Ferreyra, Margaret McMacken, John Garrow, Dave Davare, Brian Junker, participants at the October 3, 2008 seminar on applied microeconomics at the Tepper School, participants at the May 23, 2009 seminar at Heinz College, and comments received at the 2009 Annual Meeting of the American Economic Finance Association and the 2009 Summer Meeting of the North American Econometric Society. Financial support from the Heinz Endowments and the William Penn Foundation is gratefully acknowledged. Responsibility for the findings and interpretation of this paper rests solely with the authors.

Regarding Chapter 2, I would like to thank Lowell Taylor, Maria Marta Ferrerya, and Mel Stephens, Heinz and Tepper seminar participants, Melanie Zilora, Billie Morrow Davis, John Gardner, and Abby Linn. I also thank Dave Davare, Anne Herald, and the staff of the PSBA.

Chapter 3 has benefited from the advice of Lowell Taylor, Melanie Zilora, Billie Morrow Davis, Aaron Turner, my committee, and comments received at my dissertation proposal. I also thank Stephanie Cellini and Latika Chaudhary for their help in replicating their paper. I gratefully invite the comments and feedback of any readers of this version. This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. The views expressed here do not necessarily reflect the views of the BLS.

Table of Contents

Three Essays on the Economics of Education **1**
Introduction **2**
Acknowledgements **4**
Table of Contents **5**
List of Tables **6**
List of Figures **8**
Chapter 1: With the Best of Intentions: School-Level Interventions and the Pursuit of Proficiency **9**
1.1 Introduction and Research Questions 10
1.2 Some Background on Student Achievement Testing and Testing in Pennsylvania 12
1.3 Overview of 4Sight and its Use in Pennsylvania 14
1.4 Evaluation Methodology and Data Requirements 15
1.5 Empirical Results 20
1.6 Conclusions, Implications, and Future Research 29
1.7 References 30
1.8 Tables and Figures 33
1.9 Appendix 59
Chapter 2: Labor Strife in Public Schools: Does it Affect Education Production? **64**
2.1 Introduction 65
2.2 Institutional Details 67
2.3 Conceptual Framework 69
2.4 Data 72
2.5 Empirical Model 77
2.6 Results 79
2.7 Conclusions and Policy Implications 80
2.8 References 84
2.9 Tables and Figures 88
2.10 Appendix 99
Chapter 3: Selection Bias and For-Profit Associate Degrees: Evidence from the NLSY Geocode **103**
3.1 Motivation 104
3.2 Conceptual Framework 107
3.3 Data 110
3.4 Empirical Methodology 115
3.5 Results 116
3.6 Replication of Cellini and Chaudhary (2012) 120
3.7 Discussion 123
3.8 References 125
3.9 Tables and Figures 128

List of Tables

Table 1.1: Student-Level Descriptive Statistics by 4Sight Math Use....33
Table 1.2: Mean Student-Level 2005-06 PSSA Scaled Scores by 4Sight Math Use....34
Table 1.3: Student Descriptive Statistics by 4Sight Reading Use....35
Table 1.4: Mean Student-Level 2005-06 PSSA Scaled Scores by 4Sight Reading Use....36
Table 1.5: School-Level Descriptive Statistics by 4Sight Math Use....37
Table 1.6: Mean School--Level 2005-06 PSSA Scaled Scores by 4Sight Math Use....38
Table 1.7: School-Level Descriptive Statistics by 4Sight Reading Use....39
Table 1.8: Mean School-Level 2005-06 PSSA Scaled Scores by 4Sight Reading Use....40
Table 1.9: Stochastic Specifications....41
Table 1.10: Individual-Level OLS Results: Math....43
Table 1.11: Individual-Level OLS Results: Reading....44
Table 1.12: Individual-Level Quartile Regressions: Math....45
Table 1.13: Individual-Level Quartile Regressions: Reading....46
Table 1.14: Individual-Level Probit Results: Math....47
Table 1.15: Individual-Level Probit Results: Reading....48
Table 1.16: Individual-Levle Tobit Regression Results:....49
Table 1.17: Individual-Level Tobit Regression Results:....50
Table 1.18: Individual-Level Tobit Regression Results: Math Reporting Category with Controls....51
Table 1.19: Individual-Level Tobit Regression Results:....52
Table 1.20: School-Level OLS Results: Math....53
Table 1.21: School-Level OLS Results: Reading....54
Table 1.22: School-Level Nearest-Neighbor Matching Model Results: ATT of Math 4Sight on Mean Math PSSA Scaled Scores....57
Table 1.23: School-Level Nearest-Neighbor Matching Model Results: ATT of Reading 4Sight on Mean Reading PSSA Scaled Scores....58
Table 2.1: District-Level Math PSSA Pass Rate Descriptive Statistics....88
Table 2.2: District-Level Reading PSSA Pass Rate Descriptive Statistics....90
Table 2.3: Descriptive Statistics: Striking vs. Non-Striking Districts....92
Table 2.4: Percent of Districts with Expired Contracts or Striking Each Year....94
Table 2.5: Cohort Regression Results: 8th-11th Grade‡....96
Table 2.6: Regression Results: Graduation Rate†....97
Table 2.7: Cohort Regression for One-Year First Differences, 4th-8th Grades†....98
Table 3.1: Descriptive Statistics....128
Table 3.2: Descriptive Statistics for Two-Year Completers, by Institutional Status....129
Table 3.3: Outcome Descriptive Statistics, Age 25+....130
Table 3.4: Outcome Descriptive Statistics for Two-Year Completers, by Institution Type, Age 25+....131
Table 3.5: Area of Major for Two-Year Completers, by Institution Type....132
Table 3.6: Selected Descriptive Statistics for Two-Year Completers, by Major and Institution Type....133
Table 3.7: OLS Regression Results: Effects on Log Weekly Earnings....134
Table 3.8: OLS Regression Results: Effects on Other Outcomes....135
Table 3.9: Two-Year Completers: Industry Working In At Age 26, by Major and Institution Type....136
Table 3.10: Descriptive Statistics for Fixed-Effects Analysis....137

Table 3.11: Pre- and Post-Education Summary Statistics 138
Table 3.12: Fixed-Effects Analysis: Returns to College Attendance, Log Weekly Earnings[a] 139
Table 3.13: Fixed-Effects Analysis: Returns to College Attendance and Completion, Log Weekly Earnings[a] .140
Table 3.14: Fixed-Effects Analysis: Effects of College Attendance on Other Labor Market Outcomes[a] 141
Table 3.15: Fixed-Effects Analysis: Effects of College Attendance and Completion on Other Labor Market Outcomes[a] 142
Table 3.16: Fixed-Effects Analysis: Returns to College Attendance by Major, Log Weekly Earnings 143
Table 3.17: Fixed-Effects Analysis: Returns to College Attendance and Completion by Major, Log Weekly Earnings 144

List of Figures

Figure 1.1: Propensity Scores for Math 4Sight Use† 55
Figure 1.2: Propensity Scores for Reading 4Sight Use† 56
Figure 2.1: Time-Trend in Math PSSA Pass Rates........ 89
Figure 2.2: Time Trend in Reading PSSA Pass Rates 91
Figure 2.3: Map of Striking Districts........ 93
Figure 2.4: Test Score Data Available, By Cohort 95

Chapter 1: With the Best of Intentions: School-Level Interventions and the Pursuit of Proficiency

with Robert P. Strauss

Abstract

This paper analyzes the effectiveness of a battery of formative assessments, 4Sight, which are broadly aligned with annual assessments required by NCLB. These formative assessments are designed to provide teachers with feedback on student performance throughout the year in order to raise end of year student test scores. Methodologies for evaluating the effect of using 4Sight on test score outcomes are drawn from the program evaluation literature, and include individual and school-level OLS, quartile regressions, and probit regressions, as well as matching at the school-level. Micro-econometric results show that 4Sight had no discernible effect on math scores and a small negative effect on reading scores in its first year in Pennsylvania. Policy recommendations include continuing a smaller-scale trial period of 4Sight for several more years in conjunction with careful, improved alignment between 4Sight and PSSA, ongoing empirical analysis of its effects, and incorporating more opportunity for feedback to teachers and students to improve 4Sight as a formative assessment. Perhaps the most surprising finding of this paper is that while 4Sight has a very small effect on student outcomes, students receiving tutoring under NCLB are significantly and between four and nine percent less likely to pass their exams, even after controlling for all covariates, including eligibility for tutoring. This implies that the publicly-funded tutoring is actually disadvantaging those students who receive it. These results should serve as a cautionary example to school, district, and state-level policy-makers when choosing interventions designed to improve student performance on tests tied to NCLB.

1.1 Introduction and Research Questions

The enactment of the No Child Left Behind (NCLB) Act in 2002 has hastened the spread of a culture of testing in the K – 12 public schools of our nation. Since its enactment, Section 1111 of NCLB has required each state to devise academic standards on which to base their local school curriculum and annual assessments in math, reading, writing, and science. Year after year, NCLB requires an increasing fraction of public school students to pass these assessments as a condition to the state, district, and school continuing to receive federal funding, which makes up 9% of public school funding nationwide.[1] Failure at the school level to meet such targets over time can lead to sanctions and mandatory school reform, and each summer the release of school-level test results in every state has substantial repercussions for local superintendents, school boards, teachers, and principals. Political consequences and push back at sanctions are substantial, and the pressure to improve test scores increasingly reaches every classroom and lesson plan on a daily basis in public education.

While there is growing focus and concern about the consequences of the culture of testing on students, teachers, schools, and other stakeholders,[2] there is far less focus on the tests themselves, and on the procedures that states and school districts are using to raise their performance. As a practical matter, it is extremely difficult to devise a test of 6th grade math of several hours duration that will test the knowledge and skills of what a 6th grade student is expected to know about math.[3] Historically, teachers within a school or district, taking advice from organizations such as the National Council of Teachers of Mathematics (NCTM) or the National Council of Teachers of English (NCTE), would devise their own set of 6th grade math or reading standards, upon which they would base their teaching, curriculum, and assessments. With the advent of federally-mandated state-wide standardized testing, schools and districts have had to realign their curricula and teaching to address the same academic standards that are assessed in the state exams. The principle of teaching to and assessing the same set of academic skills is called alignment.

[1] As of 2004-05. (Hoover Institution 2006)
[2] News articles such as "Schools Found Likely to Miss NCLB Targets" (Cavanagh and Hoff 2008) and "No Easy Answers About NCLB's Effect on 'Poverty Gap'" (Viadero 2007) summarize some of the research and public opinion regarding the effects of NCLB.
[3] See Table A.1 in the appendix for the list of Pennsylvania's math academic standards for 6th graders.

In response to increasing pressure for students to perform well on state assessments, many states and districts are turning to intra-school year tests which demonstrate how each student is progressing towards preparation for the end-of-year state assessment and enable meeting deficiencies before the high stakes test. These interim tests are sometimes referred to as formative assessments, "ongoing assessments designed to make students' thinking visible to both teachers and students" (National Research Council 2000, p.24), or as benchmark assessments, which seek to predict a student's score on an upcoming assessment, such as the state assessments required under NCLB. This paper provides analysis of the alignment, use, and level of success of one such set of tests, 4Sight, a "benchmark assessment" which is written and sold to school districts in a variety of states by the Success for All Foundation of Maryland. In broader terms, the use and analysis of 4Sight should be considered an example of a type of reform that schools, districts, or states may choose to purchase and invest human resources into as part of the push to increase student test scores motivated by the sanctions and reform mandated by NCLB.

Our analysis of 4Sight examines the Pennsylvania version of the reading and mathematics tests at the 6th grade level. Our purpose is threefold. First, we examine the alignment of 4Sight with the academic standards of the state of Pennsylvania. Second, we examine the way Success for All and the Pennsylvania Department of Education (PDE) suggest that 4Sight is used. Finally, we statistically analyze the effects of the initial use of 4Sight in Pennsylvania and the subsequent performance of students on the state NCLB-approved examination, the Pennsylvania State System of Assessment (PSSA).

Our findings indicate that 4Sight is not fully aligned with the state standards, with the reading tests covering only 40% of the content contained in the Pennsylvania reading standards and the math tests covering 80% of the content contained in the state math standards. We also determine that, according to its proposed use by Success for All and the PDE, 4Sight only provides feedback to teachers regarding their students' performance. By not also providing students with feedback as to their strengths and weaknesses, 4Sight fails to conform to the definition of a formative assessment, as defined by the National Research Council. Finally, our estimates of OLS and matching models of the education production function reveal that 4Sight had little or no statistically significant effect on student performance on the PSSA in its pilot year.

Our results suggest several policy recommendations. The failure of 4Sight to have a significant positive effect on PSSA scores might be caused in part by its incomplete coverage or its failure to conform to the criteria of formative assessment. We suggest that the coverage of 4Sight be expanded to become fully aligned with the Pennsylvania state standards in math and reading. Furthermore, we suggest that Success for All and the PDE include additional tools and training with 4Sight to allow both teachers and students to receive feedback from each test so that it becomes a true formative assessment. Finally, we suggest that careful and thorough statistical analysis of PSSA performance with regard to 4Sight use be continued for several more years, until a reliable pattern of the effects of 4Sight use can be observed.

On a broader scale, our findings suggest that policy-makers within the school system seeking any type of intervention to increase student test scores should proceed with caution. Pennsylvania's large-scale pilot-year implementation of 4Sight was most likely motivated by increasing pressure to increase student performance. However, if the empirical results from the pilot year are representative of the effect 4Sight will continue to have on student performance, this is an intervention whose use it would have behooved the state to embark upon using a smaller and more organized trial period, and this should serve as an example for policy-makers in the future.

1.2 Some Background on Student Achievement Testing and Testing in Pennsylvania

Pennsylvania began testing students state-wide as a result of the School District Reorganization Act (Act 229), which required the State Board of Education to develop an "evaluation procedure designed to measure objectively the adequacy and efficiency of the educational program offered by the public schools of the Commonwealth."[4] The purpose of these tests was to allow districts to appraise their own educational performance and to provide "uniform evaluation" across school districts. In conjunction with Educational Testing Service (ETS), the Pennsylvania Department of Education (PDE) constructed the first state assessment of students in Pennsylvania, which took place in the 1969-70 school year, testing students in grades 5 and 11 in many subject areas; grade 8 testing was added in 1974. This program, which reported only school-level scores, ran through 1988, when the state implemented student-level competency testing, Testing for

[4] Quoted text in this section comes from Act 229. For further detail about the history of standardized testing in Pennsylvania, see Chapter 1 of the *Technical report for the Pennsylvania System of School Assessment: 2006 Reading and mathematics grades 4, 6, and 7* (Data Recognition Corporation 2007).

Essential Learning and Literary Skills (TELLS), designed to identify students in grades 3, 5, and 8 with difficulties in reading or mathematics. TELLS continued until 1991, and in 1992 the current Pennsylvania System of School Assessment (PSSA) began testing reading and mathematics at grades 5, 8, and 11, and writing at grades 6 and 9.

In 1999, the Pennsylvania State Board of Education adopted a new set of academic standards detailing the knowledge and skills students should have at each grade level so that in 2000, the purpose of the PSSA became two-fold: to measure student attainment of academic standards and to assess the extent to which school policies enabled students to achieve proficiency. This change pre-dated the inception of NCLB by two years. As a result, Pennsylvania used the standards and assessments they already had in place in order to meet the requirements of the federal legislation. Testing has expanded over the years since 2002; one requirement of NCLB is that every student[5] be tested in grades 3, 5, 8, and 11. The majority of these students must also have their scores included in the school-level reports documenting the fraction of students within subcategories[6] that have performed at or above a level defined as "proficient."[7] In 2006, Pennsylvania testing was expanded to include grades 4-8 and 11. To meet the requirement of NCLB, each year a school must make Adequate Yearly Progress (AYP). In 2006 in Pennsylvania, meeting AYP meant that a school had at least a 95% participation rate in the PSSA, at least 45% of students scored at or above proficiency on the math assessment, and at least 54% of students scored at or above proficiency on the reading assessment within each subcategory.

In this paper, we use scores from the spring of 2006 6th grade math and reading PSSA tests. The academic standards adopted by the PDE in 1999 are the foundation upon which these tests are designed. In 2005, the PDE developed Assessment Anchor Content Standards (Assessment Anchors) to further clarify the material students should learn and would be tested on in each grade. As a result of these two adoptions, material on each of the tests is broken down first into reporting categories, which describe broad categories of content, then further into assessment anchors, and finally into eligible content, which specifies the type of question that

[5] Students are exempted from taking the tests if they are in their first year as a limited English proficiency (LEP) student. Students with severe cognitive disabilities may qualify to take an alternate assessment.
[6] Proficiency rates are reported for the entire school and nine subgroups (if they have at least 40 students): American Indian, Asian, Black, Hispanic, White, Multi-Racial, IEP, LEP, and Economically Disadvantaged.
[7] Pennsylvania defines four performance indices: Advanced, Proficient, Basic, and Below Basic.

may be asked on the PSSA[8] The reading test is broken into two reporting categories, "Comprehension and Reading Skills" and "Interpretation and Analysis of Fictional and Nonfictional Text." The math test is broken into five reporting categories: "Numbers and Operations," "Measurement," "Geometry," "Algebraic Concepts," and "Data Analysis and Probability." The further breakdown of reporting categories into assessment anchors and eligible content is presented in Section 8 below.

1.3 Overview of 4Sight and its Use in Pennsylvania

4Sight is a set of math and reading tests written by the Success for All Foundation of Maryland which they define as "a benchmark assessment tool that enables you to predict your students' reading – and in some states, math – achievement multiple times throughout the year." (Success for All 2009b) As of the 2008-09 school year, versions of 4Sight are available in 16 states (Success for All 2009c), with each state's tests tailored to assess the current set of academic standards on which that state bases its NCLB assessments. In Pennsylvania, 4Sight is available for grades 3-11 in both math and reading. These tests are designed to be given to students up to five times throughout the year, and predict student performance on the PSSA (Success for All 2009a).[9] An analysis of the 6th grade Pennsylvania 4Sight tests administered during school year 2005-6 performed by the authors revealed that each test contains between 28 and 36 questions, and each of the five versions of the math or reading test covers the same set of eligible content.

Neither the reading nor the math 4Sight tests covers all of Pennsylvania's eligible content, meaning that students should be learning and will potentially be tested on the PSSA on content not assessed by 4Sight. The details of 4Sight's alignment with Pennsylvania's 6th grade math and reading eligible content are presented in Appendix Tables A.1 and A.2. These tables list the reporting categories, assessment anchors, and eligible content covered by the PSSA. Eligible content which is italicized is not assessed by 4Sight. We find that the 4Sight math test covers 80%, and the 4Sight reading test covers 40% of the eligible content contained in Pennsylvania's 2006 academic standards.

---

[8] For a list of assessment anchors and eligible content by reporting category and grade, please see the appendix.

[9] The 4Sight Reading and Math Benchmarks 2008-2009 Technical Report for Pennsylvania provides statistical evidence that 4Sight allows "educators to use the estimated student proficiency levels and diagnostic subscale data with confidence to inform their instruction and professional development." (Success for All 2009a, p. 21) While the purpose of the Technical Report was to assess the accuracy of the predictive power the 4Sight exams provide for the PSSA, the purpose of our paper is to examine the effect of 4Sight on student performance on the PSSA.

The 2005-06 school-year was the pilot year for 4Sight in Pennsylvania. The use of 4Sight was determined at the school level, with 750 schools from 310 districts choosing to use 4Sight[10]. Schools chose to use 4Sight for one of several reasons, listed by the Pennsylvania Training & Technical Assistance Network (PaTTAN), an initiative of PDE:

> 4Sight has been used to assist districts in promoting change, addressing program needs, initiating data discussions, and fostering a data-driven culture. In addition, 4Sight has focused prevention and intervention efforts and provided a consistent reporting system for Pennsylvania districts involved in the Educational Assistance Program (EAP) Tutoring Initiative. (PaTTAN 2009)

Schools who chose to use 4Sight in 2005-06 paid $1,000 per building for up to 500 students to use the online version of the test, or about $3 per student per subject to use the paper version of the test.

Schools using 4Sight participated in training sessions organized by PaTTAN and designed to instruct teachers and administrators on the successful use of 4Sight. Topics covered in these sessions include general data analysis, using 4Sight data to prioritize concerns and determine root causes, and identifying targets to improve student achievement (PaTTAN 2008).[11] Teachers learned to interpret the results of the 4Sight exams in terms that allowed them to alter their teaching and improve student performance on certain tasks/eligible content. Neither PaTTAN nor Success for All provides a format specifically for feedback to the students, so while 4Sight does provide teachers with feedback regarding student performance, that feedback may never explicitly reach the student. In this sense, the benchmark assessment 4Sight fails to conform to the National Research Council's definition of a formative assessment.

1.4 Evaluation Methodology and Data Requirements

We are interested in evaluating the impact of 4Sight use on student performance, as measured by PSSA performance. In terms of statistical analysis, we are attempting to measure a "treatment effect:" the effect on a student of being "treated" by 4Sight. There is an extensive literature on the evaluation of social programs, most notably a literature evaluating the effectiveness of job training programs, motivated by LaLonde (1986) and discussed in detail in a handbook chapter by Heckman, LaLonde, and Smith (1999). In addition to the econometric hurdles of evaluating a treatment effect, we must evaluate this effect in the context of an education production model,

---

[10] Pennsylvania had approximately 3,000 public schools and 501 school districts in 2005-06, according to the National Center for Education Statistics (NCES) Common Core of Data (CCD).
[11]Thanks also to Marge McMackin for detailing the training process to the authors.

which, in the absence of a large amount of data, adds to the complexity of the econometric evaluation.

Data Requirements

Ideally, a model of student achievement would be evaluated using each student's complete history of educational inputs (Boardman, Davis, and Sanday 1977) and their endowment, or natural ability as a student, which is inherently unobservable. A complete history of educational inputs would include family inputs, such as the parents' educational attainment and intelligence, student inputs, such as the amount of time spent studying, teacher inputs, such as the teaching ability of each teacher the student has ever had, and school inputs, such as the academic support that students receive over time. Unfortunately, an exhaustive amount of information in each of these areas is never available to the econometrician, so we must make do with what information and proxies are available, and choose a model carefully to control for missing data as much as possible.

We can think of 4Sight as a treatment in the sense that some students receive it while others do not. The coefficient we are interested in is the expected effect of using 4Sight on a student's PSSA score. In order to evaluate this coefficient, it is desirable to observe each student's PSSA score, with and without having been treated, as demonstrated in equation (1).

$$\mathcal{E}\left[PSSA_{i,4Sight} - PSSA_{i,4Sight} \right] \tag{1}$$

Since instead each student is either tested or not tested, and we observe only one PSSA score (either under treatment or non-treatment), we must instead compare two groups of different students, who have been either treated or not treated. Instead of evaluating the coefficient in equation (1), we will be evaluating the coefficient in equation (2), which is equal to equation (1) if the assignment of treatment is random.

$$\mathcal{E}\left[PSSA \mid 4Sight, X \right] - \mathcal{E}\left[PSSA \mid \text{No } 4Sight, X \right] \tag{2}$$

Descriptive Statistical Analysis

The basic framework for our models is of education production. Economic theory tells us that student achievement can be thought of as a function of inputs from the student, the student's family, and the student's schools over time. We first examine the efficacy of 4Sight using OLS regression to estimate a linear version of the education production model of student achievement including one lagged-test score along with the covariates. The model is illustrated in equation (3), where $PSSA_t$ and $PSSA_{t-1}$ are a student's test scores in years t and $t - 1$ respectively, X_t includes a set of student characteristics at time t and S_t includes a set of school characteristics at time t.

$$(3) \qquad PSSA_t = \alpha PSSA_{t-1} + \beta X_t + \gamma S_t + \delta 4Sight + \varepsilon_t$$

The major econometric hurdle to be overcome in models of student achievement is the lack of data on each student's complete history of educational inputs and their endowment, which is difficult to observe. The model expressed in equation (3) overcomes the issue by using a lagged-test score as a proxy for non-contemporary inputs, such as school and family inputs prior to time t, and the student's endowment. Identification of the coefficients of interest assumes that the lagged-test score provides a sufficient statistic for unobserved non-contemporary and endowment inputs which decline geometrically with age.[12] Additionally, because our data is non-experimental, identification of a "treatment" effect on students using 4Sight requires that no endogeneity exists between unobservables not accounted for in the model and selection into the treatment group.

Tables A.1 and A.2 in the appendix demonstrate that 4Sight covers between 35-100% of the eligible content within a reporting category for any given grade. In order to examine the relationship between alignment and student outcomes, we use this variation in coverage to analyze the different effects of 4Sight between reporting categories. We use OLS to evaluate a value-added model of student achievement within each reporting category and then compare the coefficients on 4Sight with the coverage of the particular reporting category.

Correcting for Selection Bias

As stated above, we require selection into treatment to be random in order for equation (2) to identify the average treatment effect of 4Sight. In fact, schools choose whether or not their students use 4Sight, and these schools do not necessarily make the decision randomly. We might assume that schools take into consideration the costs of 4Sight and compare them to what there is to be gained: improvement in test scores. We might therefore expect that schools with more money and lower test scores would be more likely to use 4Sight than schools with less money and higher test scores.

[12] Todd and Wolpin (2003) provide a detailed discussion of the econometric assumptions imposed when using a single-year lagged test-score value-added model.

As is evident from the descriptive statistics in Tables 1-8, the distribution of covariates among students and schools using 4Sight and those not using 4Sight are different. This suggests the use of a matching model to correct for selection bias and identify a treatment-on-the-treated effect on test score outcomes. We use the approach first used by Rosenbaum and Rubin (1983) of matching on propensity scores. The propensity score, or the probability of a student receiving treatment given their characteristics, is defined in equation (4), where $4Sight_{it}$ is a dummy variable signifying whether student i received treatment in time t, $PSSA_{it}$ is a set of student i's test scores in time t-1 and X_{it} and S_{it}, are sets of contemporaneous student and school characteristics for student i in time t.

(4) $$\Pr\left(4Sight_{it} = 1 \mid PSSA_{i,t-1}, X_{it}, S_{it}\right)$$

Because selection is determined at the school level, and because S_{it} is composed of all individuals within a particular school, S_{it} is a sufficient statistic for determining student i's treatment. For this reason, we have aggregated all data to the school level for the purpose of calculating propensity scores. Matching thus occurs at the school level, estimating the average treatment on the treated (ATT) in terms of average test score. Estimation of the ATT is performed using the nearest-neighbor method.[13]

Dependent Variables

Data on student performance comes from the Data Recognition Corporation (DRC), which writes, administers, and scores the Pennsylvania System of School Assessment (PSSA) for the Pennsylvania Department of Education, under a signed confidentiality agreement with the Pennsylvania Department of Education. For the 2004-05 school-year, data includes detailed test score information for students in grades 5 and 11. For the 2005-06 school-year, data includes detailed test score information for students in grades 3-8 and 11. Using identifiers[14] from the data, we have created a 2-year data set matched at the individual-level, with a match-rate of 89.0%, limited to students in 5th grade in 2004-05 advancing to 6th grade in 2005-06.

PSSA scores are reported to students as scaled scores, which translate a raw score into a number greater than or equal to 700 according to a table created by PDE in association with psychometricians in any given test year. These scaled scores are normed to a school-level mean (1300) and standard deviation (100) based on raw school-level scores in the base year (1996).

[13] For a discussion on identifying ATT, see Heckman, Lalonde, and Smith (1999).

[14] Identifiers available were a state identification number, the student's name, and the student's birth date.

These scores can be appropriately interpreted at an interval-level, meaning that a 5-point difference means the same whether the base score is 1200 or 1600.[15] In addition, students see their PSSA performance broken down by reporting category. To evaluate student performance from year to year within each reporting category, we have constructed percentage correct scores for each student in each reporting category, calculated simply as the number of questions a student answered correctly divided by the number of questions asked in that category.

Explanatory Variables

The DRC test-score data includes identifying, socio-economic, and academic data for each student taking the test. We have used this data to create a set of dummy variables for each individual student including: gender (male = 0, female = 1), race categories (white, black, Hispanic, other), tutoring eligibility and tutoring status,[16] Title I status (indicates that a student is low-income or attending a school with a large percentage of low-income students), Title III status (indicates that the student is receiving instruction in English as a second language), IEP status (indicates that the student has a learning disability), and gifted status. We also know which school each student attended in each year, and have merged student-level data to school-level data according to the student's sixth-grade school. School-level data includes mean teacher experience,[17] percentage of teachers with master's degrees, mean teacher performance on standardized-tests (Praxis and National Teacher Examinations (NTE)) measured as percent-correct on the respective tests[18], student-teacher ratio, weapons violations per student, percent of students qualifying for Free and Reduced Lunch, and percent of white students. These data are also provided by PDE to this project as part of the master confidentiality agreement.

Data on the use of 4Sight is at the school-level by grade and subject (math or reading), and was available for a limited time from PDE.[19] We have merged this data to the individual-level dataset, so that a variable indicates whether a particular student used 4Sight math in 2005-

---

[15] More information is available in the *Technical Report for the Pennsylvania System of School Assessment: 2006 Reading and Mathematics Grades 4, 6, and 7* (Data Recognition Corporation 2007).

[16] Tutoring eligibility and status refer to sanctions required by NCLB Section 1116 Subsection (e) (1). Low-income students at schools which have failed AYP for at least two years in a row become eligible for private tutoring, which is paid for using federal Title I money. Not all students who are eligible choose to use this service; the tutoring status variable indicates the student did choose to receive this tutoring.

[17] Mean teacher experience is defined as the average number of years a teacher has been employed as a licensed professional in public K-12 in the state of Pennsylvania.

[18] See Strauss, Bowes, Marks and Plesko (2000) for a discussion and rationale for this transformation of teacher test scores. See Strauss and Sawyer (1986) for an earlier analysis of the effects of NTE on student achievement in North Carolina.

[19] We received this data from the PDE website in early 2008. It has since been removed.

06 and another variable indicates whether the student used 4Sight reading in 2005-06. Data regarding the coverage of 4Sight by reporting category comes from our analysis[20] of the 4Sight exams in conjunction with field discussions with area experts.

1.5 Empirical Results

Characteristics of the Data at the Student and School Levels

Descriptive statistics on students by 4Sight use, along with t- or F-values from a difference of means test comparing 4Sight users with 4Sight non-users, are provided in Tables 1 and 3. Table 1 shows statistics for students using 4Sight for math, and Table 3 shows statistics for students using 4Sight for reading. The correlation between students using 4Sight for math and those using 4Sight for reading is 0.8, meaning that most students using one are also using the other; as a result, the tables show similar statistics. Students using 4Sight are significantly more white, richer (as indicated by their Title I status), more likely to speak English as a first language (as indicated by their Title III status), more likely to be denoted as gifted, and less likely to be on an individualized education plan (an indicator of special education status) than students not using 4Sight. Students are equally as likely to qualify for tutoring under NCLB, but more likely to receive tutoring if they belong to the group using 4Sight. Differences in gender and homelessness are insignificant between the two groups. Students using 4Sight have teachers with significantly more experience and who are less likely to have a master's degree. Students using 4Sight are in significantly smaller classrooms as measured by student-teacher ratio, have significantly more peers on free or reduced lunch, and have significantly more white peers. These statistics indicate that students using 4Sight are more likely to belong to a school with slightly higher socio-economic status than their non-4Sight counterparts, but that students at their school are also more likely to be enrolled in programs such as Title I, NCLB tutoring, and free or reduced lunch. This is in accord with a model in which schools that are more pro-active in seeking helpful programs for their students are selecting into 4Sight use.

Descriptive statistics on schools by 4Sight use, along with t-values from a difference of means test comparing 4Sight users with 4Sight non-users are provided in Tables 5 and 7. As in the individual-level statistics, schools using 4Sight are significantly more white, richer (as

[20] The Pittsburgh Public Schools classified the questions on each 4Sight exam into the corresponding Reporting Categories, Assessment Anchors, and Eligible Content, and provided us with this information. The authors performed subsequent analysis of the coverage of 4Sight compared to the PSSA.

indicated by their Title I status), and have a higher percentage of students denoted as gifted, and a lower percentage of students on an individualized education plan than schools not using 4Sight. Schools using 4Sight have a higher percentage of students both qualifying for and receiving tutoring. Student-teacher ratios remain significantly lower, and teacher experience remains significantly higher for schools using 4Sight, but all other characteristics are not significantly different. Again, this is in accord with a model in which pro-active schools, via their teachers, administrators, or parents, self-select into treatment and use 4Sight.

Characteristics of the PSSA

Tables 2 and 4 show average student PSSA scores by 4Sight use. In both cases (4Sight math and reading), 4Sight users have math and reading scores that are significantly lower (by 9-11 points) than students not using 4Sight. One standard deviation for each of these tests is between 207 and 226 points, so the group means differ by approximately 5% of a standard deviation. Tables 6 and 8 show average school PSSA scores by 4Sight use. There is no statistically significant difference between either test for either 4Sight group at the school-level.

Individual Student Results: Scaled Scores

Individual-level OLS estimates of equation (3) were estimated[21] with and without the control variables, and are presented in Tables 10 and 11. Each table has six columns. The first regresses a student's math scaled score from 2006 on the student's prior math scaled score from 2005, the second column adds dummy variables for treatment with 4Sight math and 4Sight reading, and the third and fourth columns add individual and school characteristics (where white is the eliminated race category). The fifth and sixth columns include information on teachers' Praxis scores and NTE scores, respectively. These have been run as separate regressions because the data on teachers' test scores are incomplete, and their inclusion reduces the number of observations available for regression. In each case, the teacher test scores were also run in separate regressions (so that only one teacher test score is included in each regression) in order to correct for multicollinearity between these variables. Neither the signs nor the significance levels of the coefficients on these variables change much when separate regressions are performed, and results can be obtained from the authors by request.

Table 10 presents results for the PSSA math scaled scores. The first row of coefficients can be interpreted as elasticities of student performance in 2006 given performance in 2005.

[21] The stochastic specifications for all models are presented in Table 9.

Column (1) shows that in the absence of covariates, a 10% increase in 2005 scaled score implies an 8.5% increase in 2006 scaled score. This elasticity does not change in a statistically significant manner in column (2), when the use of 4Sight is added to the regression. The coefficient on 4Sight math is insignificant and near zero. Column (3) adds individual-level covariates to the regression, all of which are significant. With the addition of the individual-level covariates, the effect of 4Sight remains the same. Column (4) adds school-level covariates to the regression, all of which are significant. With the addition of the school-level covariates, the effect of 4Sight becomes positive and significant, but small; an effect of 0.12% is equivalent to a 1.7-point improvement at the mean scaled score of 1406. The coefficients on the covariates all take the expected signs, and are in line with estimates from previous studies (Krueger 1999; Hanushek 1986). Column (5) adds teachers' average Praxis scores, with a significant positive coefficient on teachers' writing scores, and significant negative coefficient on teachers' math scores. Column (6) adds teachers' average NTE scores, with significant, positive coefficients on professional and common knowledge scores, and a significant, negative coefficient on general knowledge scores.[22]

Table 11 presents results for the PSSA reading scaled scores. Again, the first row of coefficients can be interpreted as elasticities of student reading performance in 2006 given performance in 2005. Column (1) shows that in the absence of covariates, a 10% increase in 2005 scaled score implies a 7.3% increase in 2006 scaled score, which is lower than in the case of the math scaled scores. This elasticity does not change in column (2), when the use of 4Sight is added to the regression. The coefficient on 4Sight is negative and significant; a 0.26% change is equivalent to a 3.5-point decrease at the mean scaled score of 1343. This coefficient remains negative and significant in all specifications of the model. Column (3) adds individual-level covariates to the regression, all of which are significant. The coefficients on the covariates all take the expected signs. Column (4) adds school-level covariates to the regression, all of which are significant. The coefficients on the covariates all take the expected signs. Column (5) adds teachers' average Praxis scores, with a significant positive coefficient on teachers' writing scores, and significant negative coefficient on teachers' math scores. Column (6) adds teachers'

[22] See Strauss and Vogt (2007) for somewhat different, district-level results that take into account teacher selection effects.

average NTE scores, with significant, positive coefficients on professional and common knowledge scores, and an insignificant coefficient on general knowledge scores.

Tables 12 and 13 present coefficients from quartile regressions with specifications analogous to those in columns (1) through (4) in the OLS regressions. The quartile regressions examine the impact of the covariates on students at the 25th, 50th, and 75th percentiles of performance on the 2006 6th grade PSSA. Table 12 presents results for the PSSA math. Column (1) shows what can again be interpreted as an elasticity of student performance in 2006 given performance in 2005 in the absence of the covariates. These coefficients show that elasticity is nearly invariant across the quartiles of performance, with performance in 2005 predicting 84-85% of performance in 2006. Column (2) introduces a dummy variable for the use of 4Sight math, and across all quartiles, there is a very small and insignificant positive or zero effect of 4Sight on PSSA score. Column (3) introduces individual-level covariates to the regression, all of which are significant, with the exception of tutoring eligibility at the median, and gender at the 25th percentile. Race, homelessness, and IEP-status have increasing, negative effects moving from the top of the distribution downward. Title I status has a fairly constant negative effect, between 1-2%, across all quartiles, and gifted-status has a fairly constant positive effect, around 5%, across all quartiles. Title III status has a small, positive effect on performance which increases from the top of the distribution downward. Column (4) introduces school-level covariates to the regression, all of which are significant with the exception of the white-student ratio (at all quartiles) and the student-teacher ratio at the median. These coefficients show that mean teacher experience, free and reduced lunch ratio, weapons violations per student, and the ratio of teachers with a Master's degree all have the expected signs, with larger effects on the lower quartiles of performance than the higher. Student-teacher ratio shows a negative effect on the 25th percentile of performance, and a positive effect on the 75th percentile of performance.

Table 13 presents quartile results for PSSA reading. Column (1) shows that, unlike for math scores, the prior year's performance on PSSA reading has a differing and decreasing effect across quartiles. Column (2) adds a dummy variable for the use of 4Sight reading. The coefficients are small, negative, and significant for the 25th and 50th percentiles of performance. Column (3) introduces individual-level covariates to the regression, all of which are significant, with the exception of homelessness, which is only significant at the bottom quartile, and Title III status, which is not significant in any quartile. Eligibility for tutoring has a negative effect across quartiles, larger in magnitude at the bottom of the distribution; receiving tutoring has a negative and constant effect of about 1% across all quartiles. Gender has a constant effect across quartiles, with females performing a little bit less than 1% better than males. Race, homelessness, Title I status, and IEP status have the expected signs, and show a larger effect toward the bottom of the distribution of performance, as in the math results. Gifted status remains constant across all quartiles, with these students performing about 4% better than their peers. Column (4) adds school-level covariates. Mean teacher experience is insignificant and near zero across the distribution, while the remaining school-level variables all have the expected signs and show larger effects towards the bottom of the distribution.

Individual Student Results: Probit Regressions

Tables 10-13 presented the effects of 4Sight use and other covariates on the log of student scores on the PSSA, however, the policy-relevant outcome in terms of NCLB is whether or not a student performs at or above proficiency on the test. For Tables 14 and 15, the dependent variable is categorical, taking a value of one if the student performed at or above proficiency, as defined by the PDE, on the 2006 PSSA, and zero otherwise. Probit regressions were performed, and marginal effects are reported. Table 14 presents the results for PSSA math. In column (1), we see that, in the absence of covariates, performing at or above proficiency in 2005 meant the student had a 67% chance of performing at or above proficiency in 2006. Column (2) adds a dummy variable for use of 4Sight math, which has a small positive but insignificant effect on student proficiency. Column (3) includes dummy variables for whether the student was eligible and/or received NCLB-mandated tutoring. Even after controlling for eligibility, students had a 9% lower chance of performing at proficiency if they did receive tutoring. This is a surprising result, given that tutoring is a federally-mandated program intended to improve a student's chance of performing well on the exam. Column (4) adds student-level covariates, all of which are significant with the exception of tutoring eligibility, homelessness, and Title III status. These coefficients all take the expected signs. Race and gifted/IEP status all have effects greater than 10% on student proficiency. Column (5) adds school-level covariates. Student-teacher ratio and white student ratio have insignificant effects on proficiency level. Mean teacher experience and the ratio of teachers with Master's degrees both have positive and significant effects on student proficiency. Weapons violations per student and the free-and-reduced lunch ratio both have large, negative, and significant effect on student proficiency. 4Sight math does not have a significant effect in any specification.

Table 14 presents the results for PSSA reading. In column (1), we see that, in the absence of covariates, performing at or above proficiency in 2005 meant the student had a 65% chance of performing at or above proficiency in 2006. Column (2) adds a dummy variable for use of 4Sight reading, which has a small negative but insignificant effect on student proficiency. Column (3) includes dummy variables for whether the student was eligible and/or received NCLB-mandated tutoring. Again, even after controlling for eligibility, students had a 6% lower chance of performing at proficiency if they did receive tutoring. Column (4) adds student-level covariates, tutoring eligibility, receipt of tutoring, and homelessness have insignificant impacts on student proficiency level. The remaining coefficients take the expected signs, with females performing at proficiency 3% more often than males, black students about 2% less often and Hispanic student about 13% less often than white students. Title III status has a negative effect of about 2% on proficiency, and gifted/IEP status all have effects of the expected sign around 25% on student proficiency. Column (5) adds school-level covariates. Mean teacher experience, student-teacher ratio, and white student ratio have insignificant effects on proficiency level. The ratio of teachers with Master's degrees has a positive and significant effects on student proficiency. Weapons violations per student and the free-and-reduced lunch ratio both have large, negative, and significant effect on student proficiency. 4Sight math has a negative effect of 1-2% on proficiency level across all specifications, though the coefficient is only significant in column (4).

Individual Student Results: Reporting Categories

Tables 16-19 present OLS regressions of student performance within reporting categories which are measured as percent correct. Tables 16-17 are simple regressions of student performance in 2005 on performance in 2006 within each reporting category. Table 16 shows that a student's math performance within a reporting category does predict their performance the following year; the coefficients are all positive and significant, ranging from 0.90 in Category A (Numbers and Operations) to 0.56 in Category B (Measurement). Table 17 shows that a student's reading performance within a reporting category predicts their performance the following year. The coefficients are positive and significant.

Tables 18-19 add covariates to the regressions performed in Tables 16-17. Table 18 presents results for PSSA math by reporting category. The coefficient on 4Sight is positive and significant in category A, negative and significant in category E, and insignificant and near zero in categories B, C, and D. The coefficients on the covariates are significant and have the expected signs, with the exception of gender, whose sign varies between reporting categories. The coefficient on black varies in magnitude between reporting categories. Table 19 presents results for PSSA reading by reporting category. The coefficients on 4Sight are negative, significant, and larger in category A than in category B. The coefficients on the covariates have the expected signs.

School-Level Results: OLS

School-level OLS estimates of equation (1) were estimated with and without the control variables, and are presented in Tables 20 and 21. These regressions are identical to those run at the individual-level (Tables 10 and 11), but use mean aggregated data in place of individual data.

Table 20 presents results for the PSSA math scaled scores. The first row of coefficients can be interpreted as elasticities of student performance in 2006 given performance in 2005. Column (1) shows that in the absence of covariates, a 10% increase in 2005 scaled score implies a 9.7% increase in 2006 scaled score. This elasticity is noticeably larger than the elasticities from the individual-level analysis. The coefficients on 4Sight are small and insignificant in all specifications of the model. Column (3) adds individual-level covariates to the regression, all of which have the expected signs. Column (4) adds teachers' average Praxis scores, with a significant positive coefficient on teachers' writing scores, and significant negative coefficient on teachers' math scores. Column (5) adds teachers' average NTE scores, with insignificant coefficients on professional and common knowledge scores, and a significant, negative coefficient on general knowledge scores.

Table 21 presents results for the PSSA reading scaled scores. Again, the first row of coefficients can be interpreted as elasticities of student performance in 2006 given performance in 2005. Column (1) shows that in the absence of covariates, a 10% increase in 2005 scaled score implies an 8.7% increase in 2006 scaled score, which is again lower than in the case of the math scaled scores at the school-level, and higher than the reading scaled scores at the individual-level. The coefficients on 4Sight are small and insignificant in all specifications of the model. Column (3) adds individual-level covariates to the regression, all of which have the expected signs. Column (4) adds teachers' average Praxis scores, with a significant positive coefficient on teachers' writing scores. Column (5) adds teachers' average NTE scores, with insignificant coefficients on professional and common knowledge scores, and a significant, negative coefficient on general knowledge scores.

School-Level Results: Matching Models

Propensity scores were estimated using a probit model regressing 4Sight use on the school-level covariates that were significantly different in Tables 5 and 7, and both the mean math and mean reading scaled scores from 2005. Histograms of propensity scores for schools using and not using 4Sight are presented in Figures 1 and 2. A two-sample Kolmogorov-Smirnov test for distributional equality was performed to compare the distributions of propensity scores for schools using and not using 4Sight, and for both reading and math these distributions were found to be statistically different at the 1%-level. This is further evidence that there are in fact differences between the treated and untreated groups. There is a large area of overlap over the support for treated and non-treated districts for both 4Sight tests. Treated schools were matched to their nearest-neighbor based on propensity score to calculate the ATT and the standard errors were evaluated using the population variance estimator proposed in Abadie and Imbens (2006). Estimates of the ATT are presented in Tables 22 and 23. The estimates are negative. The estimate of -19.013 for reading is significant at the 5%-level.

Summary and Discussion

The coefficients on 4Sight vary in level of significance and in sign across the many specifications. At the individual level, 4Sight has a small positive effect (one- to two-tenths of a percentage point) in some specifications of the math model, and has a negative effect of slightly larger magnitude in all specifications of the reading model. There does not appear to be a large difference in the effect of 4Sight across the distribution of student performance. Similarly, the probit regressions show that 4Sight does not have a significant effect on student proficiency level in math, and has a small negative effect on student proficiency level in reading. When broken down by reporting category, 4Sight math seems to have a positive effect in category A (Numbers and Operations), and a slightly smaller negative effect in category E (Data and Probability). In reading, 4Sight has a negative effect on both reporting categories. There is no clear pattern between the coefficients on 4Sight use and the percentage of eligible content that is covered by

4Sight within each reporting category.[23] This evidence suggests that in its first year 4Sight did not improve student performance on the PSSA and in fact negatively affected reading scores.

The descriptive statistics show that the group of schools using 4Sight in 2005-06 is rich and white compared to schools not using 4Sight. In addition, these schools have significantly smaller class sizes. This suggests that the effect of 4Sight use that we have measured, the "treatment on the treated," is not necessarily accurate in terms of inference. In other words, we cannot predict the effect of 4Sight on schools significantly different than those currently using 4Sight.

Some caution should be used in interpreting these results. OLS and matching estimators do a good job of controlling for observed differences between treated and untreated students and schools, but there is reason to suspect that some endogeneity remains in the model. If there is an unobserved factor that both increases the likelihood that a school uses 4Sight and affects PSSA scores, this can cause our estimates of the effect of 4Sight to be biased. The descriptive statistics suggest that relatively wealthy schools that are more active in seeking additional help for their students are using 4Sight than those not using 4Sight; if we believe that an unobserved factor common to these schools but not the others contributes to the school's decision to use 4Sight, then we might expect this factor also to cause treated students to have higher PSSA scores regardless of 4Sight treatment. If this is the case, then our estimates of the 4Sight treatment effect are upwardly biased, and 4Sight actually has more of a negative effect than the coefficients convey.

An interesting outcome of the reporting category analysis is the emergence of differences among demographic groups within certain reporting categories. Specifically, girls perform significantly better than boys on reporting categories A (Numbers and Operations) and C (Geometry) and significantly worse than boys on reporting categories B (Measurement) and E (Data Analysis and Probability). Also, black students perform significantly worse than white students in all math reporting categories, but the coefficient has twice the magnitude in reporting category B (Measurement). The quartile regressions also reflect several interesting findings that

[23] A scatter plot comparing the percentage of coverage to the 4Sight coefficient within math reporting categories is presented in the appendix. No clear pattern emerges in this figure.

the authors have not seen reported in the literature.[24] Minority status and poverty seem to have larger effects on students towards the bottom of the distribution. Also, the elasticity of reading test performance from year to year is much higher at lower percentiles of the distribution than higher. Finally, probit regressions show that scoring at proficiency last year has surprisingly low predictive power for scoring at proficiency level this year, and in addition, reveal that students receiving tutoring under NCLB are significantly and between four and nine percent less likely to pass their exams, even after controlling for all covariates, including eligibility for tutoring. This is a surprising negative result, implying that this publicly-funded tutoring is actually hurting those students who receive it.

1.6 Conclusions, Implications, and Future Research

Our analysis of the alignment of the 6th grade Pennsylvania 4Sight exams has revealed coverage of 40% of the eligible content on the reading exam and 80% of the eligible content on the math exam. The failure of 4Sight to have a significant or positive effect on PSSA scores might be caused in part by its incomplete coverage. We suggest that the coverage of 4Sight be expanded to become fully aligned with the Pennsylvania state standards in math and reading. This is feasible at the 6th grade level without lengthening the 4Sight exams. Our analysis of the current training for and usage of 4Sight show that the assessment provides direct feedback to teachers but not to students. We suggest that Success for All and the PDE include additional tools and training with 4Sight to allow both teachers and students to receive formal feedback from each test so that it becomes a true formative assessment as defined by the National Research Council.

Our estimates of OLS and matching models of the education production function reveal that 4Sight has a small and indeterminate effect on student performance on the math PSSA in its pilot year and a small and negative effect on student performance on the reading PSSA.[25] Despite this evidence, careful and thorough statistical analysis of PSSA performance with regard to 4Sight should be performed to establish a reliable pattern of the effects of the intervention. It is clear from the initial results, however, that widespread use of 4Sight should be put off until it

[24] Sudies such as Levin (2001) and Eide and Showalter (1998) have used quantile regression to analyze education production, but have not reported coefficients for individual student characteristics such as gender, race, or enrollment in various educational programs.

[25] These findings, in effect an evaluation of an intervention intended to boost student test scores, are not dissimilar to results from Bifulco, Duncombe, and Yinger (2005), in which the authors evaluated programs of whole-school reform, including one offered by Success for All, and found that these programs did not effect student reading outcomes.

can be improved as an effective tool for improving student performance. In light of these results, policy-makers at the school, district, and state levels should be wary when committing to interventions intended to improve student test scores in response to NCLB, especially when choosing to use these interventions on a large scale.

In addition to our findings regarding school-level interventions designed to help raise student performance on standardized tests, our analysis joins a rich literature which provides estimates of the education production function. The coefficients on covariates in the individual and school-level OLS models are within the range of coefficients commonly found in the literature. In addition, we have provided coefficients on teacher test scores; Wayne and Youngs (2003) survey the studies including teacher test scores in education production functions, and our estimates on the National Teacher Examination (NTE) and Praxis tests confirm a finding summarized in their paper: "test scores matter, if college ratings have not already been taken into account" (Wayne and Youngs 2003, p.100). Perhaps the most surprising finding of this paper is that while 4Sight has a little or no effect on student outcomes, students receiving tutoring under NCLB are significantly and between four and nine percent less likely to pass their exams, even after controlling for all covariates, including eligibility for tutoring. This implies that the publicly-funded tutoring is actually hurting those students who receive it.

1.7 References

Abadie, Alberto, David Drukker, Jane Leber Herr, and Guido W. Imbens. 2001. Implementing matching estimators for average treatment effects in Stata. *The Stata Journal* 1(1): 1-18.

Abadie, Alberto, and Guido W. Imbens. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica* 74(1): 235-267.

Abadie, Alberto, and Guido W. Imbens. 2008. On the failure of the bootstrap for matching estimators. *Econometrica* 76(6): 1537-1557.

Bifulco, Robert, William Duncombe, and John Yinger. 2005. Does whole-school reform boost student performance? The case of New York City. *Journal of Policy Analysis and Management* 24 (1): 47-72.

Boardman, Anthony E., Otto A. Davis, and Peggy R. Sanday. 1977. A simultaneous equations model of the educational process. *Journal of Public Economics* 7: 23-49.

Boardman, Anthony E., and Richard J. Murnane. 1979. Using Panel Data to Improve Estimates of the Determinants of Educational Achievement. *Sociology of Education* 52(2): 113-121.

Cavanagh, Sean, and David J. Hoff. 2008. Schools found likely to miss NCLB targets. *Education Week* 28 (6): 9.

Data Recognition Corporation. 2007. Technical report for the Pennsylvania System of School Assessment: 2006 Reading and mathematics grades 4, 6, and 7. Available http://www.pde.state.pa.us/a_and_t/lib/a_and_t/2006_ReadingMathGr4_6_7_Tech_Report.pdf. Accessed 31 July 2008.
http://www.easybib.com/cite/edit/266433
Data Recognition Corporation. 2005. Technical report for the Pennsylvania System of School Assessment: 2005 reading and mathematics. Available http://www.pde.state.pa.us/a_and_t/lib/a_and_t/ 2005_PSSA_Reading_and_Math_Technical_Report.pdf. Accessed 31 July 2008.

Eide, Eric and Mark H. Showalter. 1998. The effect of school quality on student performance: A quantile regression approach. *Economic Letters* 58: 345-350.

Hanushek, Eric A. 1986. The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature* 24(3): 1141-1177.

Heckman, James J. and Lalonde, Robert J. & Smith, Jeffrey A. 1999. The economics and econometrics of active labor market programs. In *Handbook of labor economics* 1(3), edited by O. Ashenfelter and D. Card, pp. 1865-2097. Amsterdam: Elsevier.

Hoover Institution, Stanford University. 2006. Facts on policy: School funding shift. Available http://www.hoover.org/research/factsonpolicy/facts/4249156.html. Accessed 3 March 2009.

Krueger, Alan B. 1999. Experimental estimates of education production functions. *The Quarterly Journal of Economics* 114(2): 497-532.

LaLonde, Robert J. 1986. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 76 (4): 604-620.

Levin, Jesse. 2001. For whom the reductions count: A quantile regression analysis of class size and peer effects on scholastic achievement. *Empirical Economics* 26: 221-246.

National Research Council. 2000. *How people learn: Brain, mind, experience, and school.* 5th edition. Washington, D.C.: National Academy Press.

PaTTAN. 2008. Training Events. Available http://www.pattan.k12.pa.us/TrainingEvents.aspx?ContentLocation=/teachlead/AssessingtoLearn.aspx. Accessed 27 August 2009.

PaTTAN. 2009. Assessing to Learn: PA Benchmark Initiative. Available http://www.pattan.net/teachlead/AssessingtoLearn.aspx. Accessed 27 September 2009.

The Pew Research Center for the People and the Press. 2007. Young women propel Clinton's lead in '08 test: A year ahead, Republicans face tough political terrain. Press release, 31 October. Available http://people-press.org/report/?pageid=1205. Accessed 27 August 2008.

Rosenbaum, Paul R. and Donald B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41-55.

Strauss, Robert P and Elizabeth A. Sawyer. 1986. Some new evidence on teacher and student competency. *Economics of Education Review* 5(1): 41-48.

Strauss, Robert P., Lori Bowes, Mindy Marks, and Mark Plesko. 2000. Improving teacher preparation and selection: Lessons from the Pennsylvania experience, *Economics of Education Review* 19(4): 387-415.

Strauss, Robert P. and William B. Vogt. 2007. Should teachers know, or know how to teach? Unpublished manuscript, Heinz College, Carnegie Mellon.

Success for All Foundation. 2009a. 4Sight Reading and Math Benchmarks 2008-2009 Technical Report for Pennsylvania. Unpublished manuscript.

Success for All Foundation. 2009b. Elementary 4Sight Benchmarks. Available http://www.successforall.org/elementary/4sight.htm. Accessed 27 September 2009.

Success for All Foundation. 2009c. Making AYP: Active Benchmarks. Available http://www.successforall.org/ayp/benchmarks.htm. Accessed 27 September 2009.

Todd, Petra E. and Kenneth I. Wolpin. 2003. Towards a unified approach for modeling the production function for cognitive achievement. *Economic Journal* 113(485): 3-33.

Viadero, Debra. 2007. No easy answers about NCLB's effect on 'poverty gap.' *Education Week* 27(12): 12.

Wayne, Andrew J. and Peter Youngs. 2003. Teacher characteristics and student achievement gains: A review. *Review of Educational Research* 73(1): 89-122.

1.8 Tables and Figures

Table 1.1: Student-Level Descriptive Statistics by 4Sight Math Use

		Do Not Use 4Sight Math	Use 4Sight Math	Difference of Means Test[1]
Student Characteristics	**Female**	.491 (.500)	.496 (.500)	86.123***
	White	.745 (.436)	.782 (.413)	
	Black	.168 (.374)	.126 (.332)	
	Hispanic	.055 (.228)	.072 (.259)	
	Other	.028 (.166)	.015 (.120)	
	Eligible for Tutoring	.025 (.155)	.023 (.150)	
	Received Tutoring	.006 (.079)	.009 (.094)	
	Homeless	.001 (.031)	.001 (.029)	
	Title I	.258 (.437)	.209 (.406)	
	Title III	.075 (.263)	.093 (.291)	
	Gifted	.063 (.243)	.048 (.214)	
	IEP	.146 (.353)	.159 (.365)	
Student's School Characteristics	**Mean Teacher Experience**	13.7 (3.5)	14.1 (3.3)	-16.6***
	Student-Teacher Ratio	15.7 (3.3)	15.2 (2.0)	25.4***
	Students on Free and Reduced Lunch	.317 (.272)	.353 (.237)	-20.0***
	Percentage of White Students	.742 (.326)	.780 (.261)	-18.2***
	Weapons Violations Per Student	.002 (.003)	.002 (.003)	3.9***
	Teachers With Master's Degrees	.4421	.415 (.155)	24.1***
	Average Praxis I: Reading Percent Correct†	.757 (.055)	.754 (.055)	9.6***
	Average Praxis I: Writing Percent Correct††	.677 (.044)	.676 (.044)	3.7***
	Average Praxis I: Math Percent Correct†††	.771 (.070)	.768 (.073)	6.4***
	Average NTE: Common Knowledge Percent Correct‡	.652 (.045)	.650 (.041)	6.3***
	Average NTE: General Knowledge Percent Correct‡‡	.625 (.049)	.627 (.042)	-6.8***
	Average NTE: Professional Knowledge Percent Correct‡‡‡	.654 (.060)	.660 (.051)	-16.6***
	N	*89,153*	*29,367*	

[1]For the set of binary variables, an F-statistic reflects the Hotelling generalized means test. For each continuous variable, a t-statistic reflects a simple difference in means test. †*N* is 86,468 (non-users) and 27,744 (users). ††*N* is 86,386 (non-users) and 27,744 (users). †††*N* is 86,541(non-users) and 27,982 (users). ‡*N* is 88,475 (non-users) and 29,046 (users). ‡‡*N* is 88,595 (non-users) and 28,982 (users). ‡‡‡*N* is 84,089 (non-users) and 27,642 (users).
*Statistically different at the 10% level. **Statistically different at the 5% level. ***Statistically different at the 1% level.

Table 1.2: Mean Student-Level 2005-06 PSSA Scaled Scores by 4Sight Math Use

	All Students	**Do Not Use 4Sight Math**	**Use 4Sight Math**	**Difference of Means Test: t-value**
6th Grade Math	1406 (226)	1411 (226)	1402 (223)	6.3***
6th Grade Reading	1341 (208)	1346 (207)	1335 (207)	7.6***
N	*119,778*	*89,153*	*29,367*	

***Statistically different at the 1% level.

Table 1.3: Student Descriptive Statistics by 4Sight Reading Use

		Do Not Use 4Sight Reading	Use 4Sight Reading	Difference of Means Test[1]
Student Characteristics	**Female**	.491 (.500)	.496 (.500)	94.796***
	White	.744 (.436)	.784 (.411)	
	Black	.168 (.374)	.126 (.332)	
	Hispanic	.055 (.228)	.071 (.257)	
	Other	.028 (.166)	.015 (.121)	
	Eligible for Tutoring	.024 (.152)	.026 (.158)	
	Received Tutoring	.006 (.077)	.010 (.100)	
	Homeless	.001 (.032)	.001 (.028)	
	Title I	.260 (.439)	.203 (.402)	
	Title III	.075 (.264)	.091 (.288)	
	Gifted	.064 (.244)	.047 (.211)	
	IEP	.147 (.353)	.157 (.364)	
Student's School Characteristics	**Mean Teacher Experience**	13.7 (3.5)	14.0 (3.3)	-15.1***
	Student-Teacher Ratio	15.7 (3.3)	15.2 (1.9)	27.0***
	Students on Free and Reduced Lunch	.317 (.273)	.354 (.233)	-21.0***
	Percentage of White Students	.742 (.326)	.782 (.260)	-19.3***
	Weapons Violations Per Student	.002 (.003)	.002 (.003)	5.3***
	Teachers With Master's Degrees	.444 (.171)	.409 (.156)	31.8***
	Average Praxis I: Reading Percent Correct†	.757 (.055)	.753 (.055)	8.9***
	Average Praxis I: Writing Percent Correct††	.677 (.044)	.676 (.043)	2.5**
	Average Praxis I: Math Percent Correct†††	.771 (.070)	.769 (.072)	3.8***
	Average NTE: Common Knowledge Percent Correct‡	.652 (.045)	.651 (.041)	4.2***
	Average NTE: General Knowledge Percent Correct‡‡	.625 (.049)	.627 (.042)	-7.3***
	Average NTE: Professional Knowledge Percent Correct‡‡‡	.655 (.060)	.660 (.052)	-12.9***
	N	*88,626*	*29,894*	

[1]For the set of binary variables, an F-statistic reflects the Hotelling generalized means test. For each continuous variable, a t-statistic reflects a simple difference in means test. †*N* is 85,941 (non-users) and 27,744 (users). ††*N* is 85,859 (non-users) and 27,744 (users). †††*N* is 86,014 (non-users) and 27,982 (users). ‡*N* is 87,948 (non-users) and 29,046 (users). ‡‡*N* is 88,068 (non-users) and 28,982 (users). ‡‡‡*N* is 83,560 (non-users) and 27,642 (users). *Statistically different at the 10% level. **Statistically different at the 5% level. ***Statistically different at the 1% level.

Table 1.4: Mean Student-Level 2005-06 PSSA Scaled Scores by 4Sight Reading Use

	All Students	Do Not Use 4Sight Math	Use 4Sight Math	Difference of Means Test: t-value
6th Grade Math	1409 (225)	1411 (227)	1402 (222)	6.2***
6th Grade Reading	1343 (207)	1346 (207)	1335 (206)	7.4***
N	*118,520*	*88,626*	*29,894*	

***Statistically different at the 1% level.

Table 1.5: School-Level Descriptive Statistics by 4Sight Math Use

	Do Not Use 4Sight Math	Use 4Sight Math	Difference of Means Test: t-value
Mean % Female	.487 (.093)	.494 (.112)	-1.1
Mean % White	.684 (.381)	.767 (.308)	-3.3***
Mean % Black	.234 (.346)	.160 (.263)	3.3***
Mean % Hispanic	.050 (.129)	.054 (.112)	-0.4
Mean % Other	.023 (.047)	.014 (.027)	3.0***
Mean % Eligible for Tutoring	.022 (.095)	.040 (.139)	-2.4**
Mean % Received Tutoring	.006 (.039)	.016 (.080)	-2.7***
Mean % Homeless	.001 (.008)	.001 (.006)	0.6
Mean % Title I	.354 (.446)	.263 (.396)	3.0***
Mean % Title III	.088 (.251)	.078 (.240)	0.6
Mean % Gifted	.051 (.058)	.041 (.049)	2.6***
Mean % IEP	.165 (.111)	.179 (.104)	-2.2**
Mean Teacher Experience	13.5 (4.19)	14.4 (3.69)	-3.7***
Student-Teacher Ratio	15.7 (4.17)	15.0 (2.32)	2.7***
Students on Free and Reduced Lunch	.383 (.292)	.389 (.249)	-0.4
Weapons Violations Per Student	.002 (.007)	.002 (.003)	1.1
Teachers With Master's Degrees	.409 (.177)	.412 (.163)	-0.4
Average Praxis I: Reading Percent Correct†	.753 (.061)	.747 (.063)	1.3
Average Praxis I: Writing Percent Correct††	.674 (.047)	.671 (.051)	0.8
Average Praxis I: Math Percent Correct†††	.761 (.077)	.759 (.083)	0.2
Average NTE: Common Knowledge Percent Correct‡	.645 (.053)	.647 (.048)	-0.6
Average NTE: General Knowledge Percent Correct‡‡	.615 (.057)	.620 (.048)	-1.3
Average NTE: Professional Knowledge Percent Correct‡‡‡	.652 (.068)	.659 (.060)	-1.4
N	*880*	*279*	

†*N* is 822 (non-users) and 248 (users). ††*N* is 821 (non-users) and 248 (users). †††*N* is 823(non-users) and 251 (users). ‡*N* is 858 (non-users) and 270 (users). ‡‡*N* is 861 (non-users) and 268 (users). ‡‡‡*N* is 767 (non-users) and 246 (users). *Statistically different at the 10% level. **Statistically different at the 5% level. ***Statistically different at the 1% level.

Table 1.6: Mean School--Level 2005-06 PSSA Scaled Scores by 4Sight Math Use

	All Students	Do Not Use 4Sight Math	Use 4Sight Math	Difference of Means Test: t-value
6^{th} Grade Math	1387 (122)	1388 (124)	1385 (113)	0.3
6^{th} Grade Reading	1320 (112)	1320 (115)	1322 (100)	-0.2
N	*1159*	*880*	*279*	

***Statistically different at the 1% level.

Table 1.7: School-Level Descriptive Statistics by 4Sight Reading Use

	Do Not Use 4Sight Reading	Use 4Sight Reading	Difference of Means Test: t-value
% Female	.487 (.091)	.493 (.112)	-0.9
% White	.682 (.382)	.773 (.305)	-3.6***
% Black	.235 (.346)	.155 (.259)	3.6***
% Hispanic	.050 (.129)	.052 (.111)	-0.2
% Other	.023 (.046)	.014 (.027)	3.1***
% Eligible for Tutoring	.021 (.092)	.044 (.143)	-3.1***
% Received Tutoring	.006 (.038)	.018 (.081)	-3.5***
% Homeless	.001 (.008)	.001 (.006)	0.6
% Title I	.355 (.447)	.258 (.391)	3.3***
% Title III	.089 (.252)	.076 (.237)	0.7
% Gifted	.051 (.058)	.040 (.049)	3.0***
% IEP	.164 (.104)	.176 (.105)	-1.7*
Mean Teacher Experience	13.5 (4.18)	14.5 (3.69)	-3.7***
Student-Teacher Ratio	15.8 (4.16)	15.0 (2.29)	2.9***
Students on Free and Reduced Lunch	.382 (.294)	.391 (.245)	-0.5
Weapons Violations Per Student	.002 (.008)	.002 (.003)	1.2
Teachers With Master's Degrees	.410 (.177)	.407 (.163)	0.3
Average Praxis I: Reading Percent Correct†	.752 (.062)	.747 (.063)	1.2
Average Praxis I: Writing Percent Correct††	.674 (.048)	.672 (.050)	0.5
Average Praxis I: Math Percent Correct†††	.760 (.077)	.761 (.082)	-0.1
Average NTE: Common Knowledge Percent Correct‡	.644 (.053)	.647 (.047)	-0.8
Average NTE: General Knowledge Percent Correct‡‡	.615 (.057)	.620 (.049)	-1.4
Average NTE: Professional Knowledge Percent Correct‡‡‡	.652 (.068)	.658 (.062)	-1.2
N	*874*	*285*	

†*N* is 816 (non-users) and 254 (users). ††*N* is 815 (non-users) and 254 (users). †††*N* is 817 (non-users) and 257 (users). ‡*N* is 852 (non-users) and 276 (users). ‡‡*N* is 855 (non-users) and 274 (users). ‡‡‡*N* is 763 (non-users) and 250 (users). *Statistically different at the 10% level. **Statistically different at the 5% level. ***Statistically different at the 1% level.

Table 1.8: Mean School-Level 2005-06 PSSA Scaled Scores by 4Sight Reading Use

	All Students	Do Not Use 4Sight Math	Use 4Sight Math	Difference of Means Test: t-value
6^{th} Grade Math	1387 (122)	1388 (125)	1386 (112)	0.3
6^{th} Grade Reading	1320 (112)	1320 (115)	1322 (99)	-0.3
N	*1159*	*874*	*285*	

***Statistically different at the 1% level.

Table 1.9: Stochastic Specifications

Individual-Level OLS

$$\begin{aligned}\log\left(PSSA06\right)_{is} = {} & \beta_0 + \beta_1 \log\left(PSSA05\right)_{is} \\ & (2) + \beta_2 4Sight_s \\ & (3) + \beta_3 TutorElig_{is} + \beta_4 Tutor_{is} + \beta_5 Female_{is} + \beta_6 RaceDummies_{is} + \beta_7 Homeless_{is} + \beta_8 TitleI_{is} + \beta_9 TitleIII_{is} + \beta_{10} Gifted_{is} + \beta_{11} IEP_{is} \\ & (4) + \beta_{12} TeachExp_s + \beta_{13} StuTeachRat_s + \beta_{14} FRL_s + \beta_{15} WhiteStuRat_s + \beta_{16} Weapons / Stu_s + \beta_{17} Masters_s \\ & (5) + \beta_{18} Praxis_s \\ & (6) + \beta_{29} NTE_s \\ & + \varepsilon_{is}\end{aligned}$$

Individual-Level Quantile Regressions (25%, 50%, 75%)

$$\begin{aligned}\log\left(PSSA06\right)_{is} = {} & \beta_0 + \beta_1 \log\left(PSSA05\right)_{is} \\ & (2) + \beta_2 4Sight_s \\ & (3) + \beta_3 TutorElig_{is} + \beta_4 Tutor_{is} + \beta_5 Female_{is} + \beta_6 RaceDummies_{is} + \beta_7 Homeless_{is} + \beta_8 TitleI_{is} + \beta_9 TitleIII_{is} + \beta_{10} Gifted_{is} + \beta_{11} IEP_{is} \\ & (4) + \beta_{12} TeachExp_s + \beta_{13} StuTeachRat_s + \beta_{14} FRL_s + \beta_{15} WhiteStuRat_s + \beta_{16} Weapons / Stu_s + \beta_{17} Masters_s \\ & + \varepsilon_{is}\end{aligned}$$

Individual-Level Probit: Proficiency

$$\begin{aligned}proficient06_{is} = {} & \beta_0 + \beta_1 proficient05_{is} \\ & (2) + \beta_2 4Sight_s \\ & (3) + \beta_3 TutorElig_{is} + \beta_4 Tutor_{is} + \beta_5 Female_{is} + \beta_6 RaceDummies_{is} + \beta_7 Homeless_{is} + \beta_8 TitleI_{is} + \beta_9 TitleIII_{is} + \beta_{10} Gifted_{is} + \beta_{11} IEP_{is} \\ & (4) + \beta_{12} TeachExp_s + \beta_{13} StuTeachRat_s + \beta_{14} FRL_s + \beta_{15} WhiteStuRat_s + \beta_{16} Weapons / Stu_s + \beta_{17} Masters_s \\ & + \varepsilon_{is}\end{aligned}$$

Individual-Level Tobit: Reporting Category

$$\begin{aligned}Category\%Correct_{is} = {} & \beta_0 + \beta_1 Category\%Correct_{is} \\ & (2) + \beta_2 4Sight_s + \beta_3 TutorElig_{is} + \beta_4 Tutor_{is} + \beta_5 Female_{is} + \beta_6 RaceDummies_{is} + \beta_7 Homeless_{is} + \beta_8 TitleI_{is} + \beta_9 TitleIII_{is} + \beta_{10} Gifted_{is} + \beta_{11} IEP_{is} \\ & + \beta_{12} TeachExp_s + \beta_{13} StuTeachRat_s + \beta_{14} FRL_s + \beta_{15} WhiteStuRat_s + \beta_{16} Weapons / Stu_s + \beta_{17} Masters_s \\ & + \varepsilon_{is}\end{aligned}$$

Table 1.9 (continued): Stochastic Specifications

School-Level OLS
$\log\left(meanPSSA06\right)_s = \beta_0 + \beta_1 \log\left(meanPSSA05\right)_s$ $(2) + \beta_2 4Sight_s$ $(3) + \beta_3 \%TutorElig_s + \beta_4 \%Tutor_s + \beta_5 \%Female_s + \beta_6 \%Race_s + \beta_7 \%Homeless_s + \beta_8 \%TitleI_s + \beta_9 \%TitleIII_s + \beta_{10} \%Gifted_s + \beta_{11} \%IEP_s$ $+\beta_{12} TeachExp_s + \beta_{13} StuTeachRat_s + \beta_{14} FRL_s + \beta_{15} Weapons / Stu_s + \beta_{16} Masters$ $(5) + \beta_{17} Praxis_s$ $(6) + \beta_{18} NTE_s$ $+\varepsilon_s$
Propensity Scores for Matching Models: Probit
$4Sight_s = \beta_0 + \beta_1 PSSA05_s + \beta_2 tutor_s + \beta_3 tutorelig_s + \beta_4 \%white_s + \beta_5 \%black_s + \beta_6 \%TitleI_s + \beta_7 \%GIEP_s + \beta_8 \%IEP_s + \beta_9 TeachExp_s + \beta_{10} StuTeachRat_s + \varepsilon_s$

Table 1.10: Individual-Level OLS Results: Math

	(1)	(2)	(3)	(4)	(4a) Elasticities	(5)	(6)
	Log of 2006 6th Grade Math Scaled Score						
Log of 2005 5th Grade Math Scaled Score	0.8473 (0.0040)***	0.8473 (0.0039)***	0.7415 (0.0046)***	0.7361 (0.0045)***		0.7344 (0.0046)***	0.7355 (0.0047)***
4Sight Math		0.0001 (0.0031)	-0.0008 (0.0028)	0.0012 (0.0028)		0.0008 (0.0028)	0.0016 (0.0028)
Eligible for Tutoring			-0.0077 (0.006)	-0.0067 (0.0058)		-0.0074 (0.0062)	-0.0062 (0.0062)
Receiving Tutoring			-0.0105 (0.0072)	-0.0145 (0.0069)**		-0.0142 (0.0073)**	-0.0152 (0.0073)**
Female			-0.0013 (0.0006)**	-0.0015 (0.0006)***		-0.0015 (0.0006)**	-0.0015 (0.0006)***
Black			-0.0327 (0.0021)***	-0.0245 (0.0015)***		-0.0249 (0.0016)***	-0.024 (0.0016)***
Hispanic			-0.0226 (0.0029)***	-0.0147 (0.0026)***		-0.0148 (0.0026)***	-0.0156 (0.0026)***
Other			0.0165 (0.0024)***	0.0172 (0.0022)***		0.0169 (0.0021)***	0.0171 (0.0022)***
Homeless			-0.0227 (0.0098)**	-0.0177 (0.0095)*		-0.0182 (0.0095)*	-0.0135 (0.0103)
Title I			-0.017 (0.0024)***	-0.0027 (0.0031)		-0.0019 (0.0032)	-0.0033 (0.0032)
Title III			0.0037 (0.0041)	0.0038 (0.004)		0.0033 (0.0042)	0.004 (0.0042)
Gifted			0.054 (0.0017)***	0.053 (0.0018)***		0.0534 (0.0018)***	0.0533 (0.0018)***
IEP			-0.047 (0.0015)***	-0.0476 (0.0015)***		-0.048 (0.0015)***	-0.0479 (0.0016)***
Mean Teacher Experience				0.0005 (0.0004)	0.0073	0.0005 (0.0004)	0.0003 (0.0004)
Student-Teacher Ratio				-0.0001 (0.0004)	-0.0013	0 (0.0004)	0.0002 (0.0004)
Free and Reduced Lunch Ratio				-0.0321 (0.0059)***	-0.0105	-0.0322 (0.0060)***	-0.0324 (0.0061)***
White Student Ratio				0.0022 (0.006)	0.0017	0.0036 (0.0062)	0.0015 (0.0063)
Weapons Violations Per Student				-0.8781 (0.3727)**	-0.0018	-0.73 (0.3950)*	-0.9287 (0.4260)**
Ratio of Teachers with Master's Degree				0.0153 (0.0066)**	0.0067	0.0123 (0.0071)*	0.0179 (0.0070)**
Mean Praxis Reading Percent Correct						-0.0057 (0.0258)	
Mean Praxis Writing Percent Correct						0.0791 (0.0324)**	
Mean Praxis Math Percent Correct						-0.0324 (0.0190)*	
Mean NTE Common Knowledge Percent Correct							0.0354 (0.0414)
Mean NTE General Knowledge Percent Correct							-0.0532 (0.0349)
Mean NTE Professional Knowledge Percent Correct							0.022 (0.0224)
Constant	1.0938 (0.0285)***	1.0937 (0.0283)***	1.8754 (0.0335)***	1.9075 (0.0338)***		1.8934 (0.0386)***	1.9084 (0.0393)***
Observations	118,391	118,391	118,391	118,391		113,878	111,626
Adjusted R-squared	0.6867	0.6867	0.7091	0.7117		0.7119	0.7120

Robust, clustered (by school) standard errors in parentheses * significant at 10%; ** significant at 5%; * significant at 1%**

Table 1.11: Individual-Level OLS Results: Reading

	(1)	(2)	(3)	(4)	(4a) Elasticities	(5)	(6)
	Log of 2006 6th Grade Reading Scaled Score						
Log of 2005 5th Grade Reading Scaled Score	0.7291 (0.0032)***	0.7290 (0.0032)***	0.6275 (0.0033)***	0.6204 (0.0032)***		0.6201 (0.0033)***	0.6185 (0.0033)***
4Sight Reading		-0.0027 (0.0022)	-0.0039 (0.0020)*	-0.0017 (0.002)		-0.0017 (0.0021)	-0.0016 (0.0021)
Eligible for Tutoring			-0.0045 (0.0053)	-0.0037 (0.0046)		-0.0024 (0.0047)	-0.0025 (0.0047)
Receiving Tutoring			-0.011 (0.0072)	-0.015 (0.0062)**		-0.0165 (0.0062)***	-0.015 (0.0063)**
Female			0.0095 (0.0006)***	0.0095 (0.0006)***		0.0094 (0.0006)***	0.0095 (0.0006)***
Black			-0.0309 (0.0020)***	-0.0229 (0.0016)***		-0.0228 (0.0016)***	-0.0229 (0.0016)***
Hispanic			-0.0229 (0.0030)***	-0.0157 (0.0027)***		-0.0154 (0.0027)***	-0.0169 (0.0027)***
Other			0.0074 (0.0020)***	0.0079 (0.0020)***		0.0077 (0.0020)***	0.0068 (0.0020)***
Homeless			-0.0249 (0.0119)**	-0.0199 (0.0114)*		-0.0203 (0.0115)***	-0.0085 (0.0108)
Title I			-0.0185 (0.0021)***	-0.0052 (0.0025)**		-0.0044 (0.0026)*	-0.0061 (0.0026)**
Title III			0.0009 (0.0027)	0.0009 (0.0027)		-0.0001 (0.0027)	0.0007 (0.0028)
Gifted			0.0473 (0.0013)***	0.0465 (0.0013)***		0.0467 (0.0013)***	0.0465 (0.0013)***
IEP			-0.0534 (0.0015)***	-0.0547 (0.0015)***		-0.0548 (0.0015)***	-0.0551 (0.0015)***
Mean Teacher Experience				0.0001 (0.0003)	0.0017	0.0003 (0.0003)	0.0002 (0.0003)
Student-Teacher Ratio				-0.0004 (0.0003)	-0.0064	-0.0003 (0.0003)	-0.0001 (0.0003)
Free and Reduced Lunch Ratio				-0.0236 (0.0048)***	-0.0077	-0.0234 (0.0051)***	-0.0235 (0.0051)***
White Student Ratio				0.0047 (0.0047)	0.0035	0.0046 (0.0049)	0.0014 (0.0049)
Weapons Violations Per Student				-1.2875 (0.3053)***	-0.0027	-1.2953 (0.3267)***	-1.4479 (0.3377)***
Ratio of Teachers with Master's Degree				0.0241 (0.0049)***	0.0105	0.0227 (0.0051)***	0.0228 (0.0051)***
Mean Praxis Reading Percent Correct						-0.0074 (0.0173)	
Mean Praxis Writing Percent Correct						0.0704 (0.0217)***	
Mean Praxis Math Percent Correct						-0.0127 (0.0139)	
Mean NTE Common Knowledge Percent Correct							0.0303 (0.0299)
Mean NTE General Knowledge Percent Correct							-0.0015 (0.0264)
Mean NTE Professional Knowledge Percent Correct							0.0207 (0.0165)
Constant	1.9508 (0.0232)***	1.9523 (0.0233)***	2.6928 (0.0239)***	2.7401 (0.0249)***		2.7064 (0.0286)***	2.7186 (0.0300)***
Observations	117,974	117,974	117,974	117,974		113,475	111,225
Adjusted R-squared	0.6691	0.6692	0.6916	0.6942		0.6944	0.6945

Robust, clustered (by school) standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.12: Individual-Level Quartile Regressions: Math

	(1)	(2)	(3)	(3a) Elasticities	(1)	(2)	(3)	(3a) Elasticities	(1)	(2)	(3)	(3a) Elasticities
	Log of 2006 6th Grade Math Scaled Score											
	25% Quantile				50% Quantile				75% Quantile			
Log of 2005 5th Grade Math Scaled Score[2,3]	0.8410 (0.0033)***	0.7362 (0.0025)***	0.731 (0.0024)***		0.8384 (0.0018)***	0.7362 (0.0025)***	0.7448 (0.0026)***		0.8456 (0.0032)***	0.7488 (0.0018)***	0.7429 (0.0033)***	
4Sight Math	0.0006 (0.0008)	-0.0005 (0.0007)	0.0016 (0.0007)*		0 (0.0007)	-0.0006 (0.0009)	0.0011 (0.0007)		0 (0.0004)	-0.0008 (0.0101)	0.0011 (0.0007)**	
Eligible for Tutoring[2]		-0.0101 (0.0021)***	-0.0073 (0.0022)***			-0.0033 (0.0021)	-0.0085 (0.0024)			-0.0068 (0.0017)***	-0.0028 (0.0029)***	
Receiving Tutoring[2]		-0.0125 (0.0041)***	-0.009 (0.0052)*			-0.0179 (0.0051)***	-0.0171 (0.0047)***			-0.0076 (0.0049)	-0.0171 (0.0047)***	
Female		-0.0007 (0.0007)	-0.0017 (0.0006)**			-0.0019 (0.0007)***	-0.0016 (0.0007)**			-0.0022 (0.0010)***	-0.0013 (0.0006)***	
Black[2]		-0.0337 (0.0012)***	-0.022 (0.0015)***			-0.0315 (0.0008)***	-0.0253 (0.0010)***			-0.0282 (0.0010)***	-0.0239 (0.0016)***	
Hispanic[2,3]		-0.0269 (0.0023)***	-0.0186 (0.0014)***			-0.0212 (0.0014)***	-0.0144 (0.0015)***			-0.0175 (0.0016)***	-0.0113 (0.0013)***	
Other		0.0125 (0.0018)***	0.0147 (0.0013)***			0.0156 (0.0015)***	0.0154 (0.0014)***			0.0155 (0.0022)***	0.0152 (0.0016)***	
Homeless[2]		-0.0418 (0.0093)***	-0.011 (0.0118)*			-0.0273 (0.0124)**	-0.0219 (0.0133)**			-0.0081 (0.0123)	-0.0282 (0.0071)	
Title I[2]		-0.0190 (0.0012)***	-0.0041 (0.0011)***			-0.0154 (0.0008)***	-0.0029 (0.0009)***			-0.0153 (0.0007)***	-0.0035 (0.0014)***	
Title III		0.0052 (0.0014)***	0.0046 (0.0009)***			0.0042 (0.0009)***	0.0049 (0.0011)***			0.0035 (0.0018)*	0.0035 (0.0012)***	
Gifted[2,3]		0.0476 (0.0015)***	0.0471 (0.0014)***			0.0466 (0.0007)***	0.0551 (0.0013)***			0.0559 (0.0015)***	0.0463 (0.0014)***	
IEP[2,3]		-0.0548 (0.0013)***	-0.0559 (0.0010)***			-0.0409 (0.0008)***	-0.0334 (0.0011)***			-0.0329 (0.0015)***	-0.0408 (0.0013)***	
Mean Teacher Experience			0.0006 (0.0001)***	0.0081			0.0005 (0.0001)***	0.0063			0.0004 (0.0001)***	0.0061
Student-Teacher Ratio[3]			-0.0004 (0.0001)***	-0.0069			0.0001 (0.0001)	0.0017			0.0006 (0.0001)***	0.0091
Free and Reduced Lunch Ratio[3]			-0.0369 (0.0014)***	-0.0120			-0.0285 (0.0020)***	-0.0093			-0.0243 (0.0022)***	-0.0079
White Student Ratio			-0.0001 (0.002)	-0.0000			0.0029 (0.0021)	0.0022			0.0022 (0.0026)	0.0017
Weapons Violations Per Student			-1.0072 (0.1632)***	-0.0021			-0.7455 (0.1705)***	-0.0016			-0.8002 (0.0986)***	-0.0017
Ratio of Teachers with Master's Degree			0.0165 (0.0024)***	0.0072			0.0137 (0.0025)***	0.0060			0.0129 (0.0025)***	0.0056
Constant	1.0826 (0.0132)***	1.8230 (0.0287)***	1.8548 (0.0251)***		1.1601 (0.0237)***	1.8821 (0.0130)***	1.886 (0.0182)***		1.1641 (0.0232)***	1.8612 (0.0177)***	1.8993 (0.0183)***	
Observations	118,391	118,391	118,391		118,391	118,391	118,391		118,391	118,391	118,391	

* significant at 10%; ** significant at 5%; *** significant at 1%; [1,2,3] coefficients for the three quartiles are statistically different at the 5% level in columns (1), (2) or (3) respectively

Table 1.13: Individual-Level Quartile Regressions: Reading

	(1)	(2)	(3)	(3a) Elasticities	(1)	(2)	(3)	(3a) Elasticities	(1)	(2)	(3)	(3a) Elasticities
	Log of 2006 6th Grade Reading Scaled Score											
	25% Quantile				50% Quantile				75% Quantile			
Log of 2005 5th Grade Reading Scaled Score[1,2,3]	0.7672 (0.0020)***	0.6522 (0.0028)***	0.6449 (0.0033)***		0.7170 (0.0023)***	0.6277 (0.0032)***	0.6190 (0.0028)***		0.6641 (0.0031)***	0.5979 (0.0026)***	0.5915 (0.0030)***	
4Sight Reading	-0.0019 (0.0008)**	-0.0047 (0.0009)***	-0.0022 (0.0008)**		-0.0015 (0.0004)***	-0.0033 (0.0008)***	-0.0008 (0.0007)		-0.0011 (0.0008)	-0.0022 (0.0011)**	-0.0009 (0.0008)	
Eligible for Tutoring		-0.0056 (0.0021)***	-0.0042 (0.0025)*			-0.0042 (0.0022)*	-0.0053 (0.0018)*			-0.0034 (0.0023)	-0.0031 (0.0035)	
Receiving Tutoring		-0.0148 (0.0048)***	-0.0124 (0.0035)***			-0.0096 (0.0026)***	-0.0162 (0.0041)***			-0.0154 (0.0039)***	-0.0162 (0.0075)**	
Female[3]		0.0073 (0.0008)***	0.007 (0.0007)***			0.0070 (0.0005)***	0.0075 (0.0006)***			0.0070 (0.0009)***	0.0084 (0.0008)***	
Black[2]		-0.0346 (0.0015)***	-0.0194 (0.0018)***			-0.0273 (0.0010)***	-0.0212 (0.0012)***			-0.0239 (0.0011)***	-0.0242 (0.0011)***	
Hispanic[2]		-0.0226 (0.0017)***	-0.0149 (0.0018)***			-0.0200 (0.0016)***	-0.0142 (0.0015)***			-0.0188 (0.0022)***	-0.0143 (0.0016)***	
Other		0.0094 (0.0016)***	0.0103 (0.0015)***			0.0066 (0.0018)***	0.0063 (0.0021)***			0.0071 (0.0023)***	0.0061 (0.0018)***	
Homeless		-0.0350 (0.0165)**	-0.0333 (0.0150)**			-0.0184 (0.0162)	-0.0009 (0.0181)			-0.0073 (0.0141)	-0.0082 (0.0136)	
Title I[2,3]		-0.0225 (0.0006)***	-0.0078 (0.0008)***			-0.0172 (0.0009)***	-0.0036 (0.0012)***			-0.0130 (0.0012)***	-0.005 (0.0014)***	
Title III		0.0007 (0.0014)	0.0011 (0.0014)			0.0003 (0.0012)	0.0022 (0.0019)			0.0014 (0.0011)	0.001 (0.0013)	
Gifted		0.0435 (0.0019)***	0.0428 (0.0016)***			0.0444 (0.0014)***	0.0427 (0.0017)***			0.0430 (0.0010)***	0.0433 (0.0012)***	
IEP[2,3]		-0.0616 (0.0017)***	-0.0417 (0.0010)***			-0.0488 (0.0013)***	-0.0634 (0.0015)***			-0.0402 (0.0011)***	-0.05 (0.0012)***	
Mean Teacher Experience			0 (0.0001)	0.0016			0 (0.0001)	0.0007			0.0001 (0.0001)	-0.0003
Student-Teacher Ratio[3]			-0.0002 (0.0001)**	-0.0069			-0.0001 (0.0001)	-0.0034			-0.0004 (0.0001)***	-0.0015
Free and Reduced Lunch Ratio[3]			-0.022 (0.0020)***	-0.0083			-0.0172 (0.0020)***	-0.0072			-0.0256 (0.0029)***	-0.0056
White Student Ratio			0.0045 (0.0017)***	0.0055			0.0014 (0.0023)	0.0034			0.0073 (0.0019)***	0.0011
Weapons Violations Per Student[3]			-1.2103 (0.1338)***	-0.0032			-1.1422 (0.1089)***	-0.0025			-1.5114 (0.1676)***	-0.0024
Ratio of Teachers with Master's Degree[3]			0.0195 (0.0019)***	0.0111			0.0205 (0.0021)***	0.0085			0.0255 (0.0028)***	0.0089
Constant	1.6239 (0.0145)***	2.9618 (0.0173)***	2.7542 (0.0255)***		2.0434 (0.0168)***	2.4667 (0.0201)***	3.0035 (0.0228)***		2.4793 (0.0226)***	2.4793 (0.0135)***	2.5120 (0.0200)***	
Observations	117,974	117,974	117,974		117,974	117,974	117,974		117,974	117,974	117,974	

* significant at 10%; ** significant at 5%; *** significant at 1%; [1,2,3] coefficients for the three quartiles are statistically different at the 5% level in columns (1), (2) or (3) respectively

Table 1.14: Individual-Level Probit Results†: Math

	(1)	(2)	(3)	(4)	(5)
	At or Above Proficiency: 2006 6th Grade Math				
At or Above Proficiency: 2005 5th Grade Math	0.6706 (0.0043)***	0.6706 (0.0043)***	0.6689 (0.0044)***	0.5260 (0.0052)***	0.5195 (0.0052)***
4Sight Math		0.0020 (0.0077)		-0.0022 (0.0063)	0.0018 (0.0061)
Eligible for Tutoring			-0.0133 (0.0196)	-0.0154 (0.0160)	-0.0126 (0.0165)
Receiving Tutoring			-0.09 (0.0261)***	-0.0559 (0.0199)***	-0.0685 (0.0203)***
Female				-0.0108 (0.0020)***	-0.0114 (0.0020)***
Black				-0.1213 (0.0066)***	-0.0969 (0.0054)***
Hispanic				-0.0804 (0.0081)***	-0.057 (0.0076)***
Other				0.033 (0.0069)***	0.0359 (0.0066)***
Homeless				-0.0851 (0.0418)**	-0.0736 (0.0397)*
Title I				-0.0627 (0.0064)***	-0.0248 (0.0075)***
Title III				-0.0084 (0.0099)	-0.0093 (0.0100)
Gifted				0.1794 (0.0058)***	0.176 (0.0057)***
IEP				-0.1957 (0.0056)***	-0.1978 (0.0054)***
Mean Teacher Experience					0.0017 (0.0009)**
Student-Teacher Ratio					-0.0015 (0.0009)*
Free and Reduced Lunch Ratio					-0.0814 (0.0131)***
White Student Ratio					-0.0087 (0.0139)
Weapons Violations Per Student					-3.1355 (0.9268)***
Ratio of Teachers with Master's Degree					0.0568 (0.0156)***
Observations	118,569	118,569	118,569	118,569	118,569
Pseudo R-squared	0.3577	0.3577	0.3583	0.4172	0.4218

†Average marginal effects are reported;
clustered (by school) standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.15: Individual-Level Probit Results†: Reading

	(1)	(2)	(3)	(4)	(5)
	At or Above Proficiency: 2006 6th Grade Reading				
At or Above Proficiency: 2005 5th Grade Reading	0.6511 (0.0038)***	0.651 (0.0038)***	0.6499 (0.0039)***	0.4993 (0.0044)***	0.4909 (0.0043)***
4Sight Reading		-0.0083 (0.0065)		-0.0138 (0.0054)**	-0.0081 (0.0051)
Eligible for Tutoring			-0.0104 (0.0175)	-0.0174 (0.0131)	-0.0145 (0.0128)
Receiving Tutoring			-0.0574 (0.0264)**	-0.0306 (0.0239)	-0.0428 (0.0203)**
Female				0.0244 (0.0020)***	0.0241 (0.0020)***
Black				-0.1036 (0.0057)***	-0.0774 (0.0046)***
Hispanic				-0.0761 (0.0074)***	-0.0525 (0.0068)***
Other				0.0132 (0.0065)**	0.0169 (0.0064)***
Homeless				-0.0413 (0.0356)	-0.0286 (0.0336)
Title I				-0.0748 (0.0058)***	-0.0325 (0.0066)***
Title III				-0.0085 (0.0080)	-0.0091 (0.0075)
Gifted				0.1859 (0.0051)***	0.1833 (0.0052)***
IEP				-0.1989 (0.0049)***	-0.2021 (0.0047)***
Mean Teacher Experience					0.0005 (0.0007)
Student-Teacher Ratio					-0.0013 (0.0007)*
Free and Reduced Lunch Ratio					-0.0792 (0.0125)***
White Student Ratio					0.0072 (0.0122)
Weapons Violations Per Student					-2.6844 (0.8266)***
Ratio of Teachers with Master's Degree					0.0710 (0.0126)***
Observations	118,148	118,148	118,148	118,148	118,148
Pseudo R-squared	0.3531	0.3532	0.3534	0.4070	0.4113

†Average marginal effects are reported;
clustered (by school) standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.16: Individual-Levle Tobit Regression Results[†]: Math Reporting Category without Controls

	(1)	(2)	(3)	(4)	(5)
	Reporting Category A	Reporting Category B	Reporting Category C	Reporting Category D	Reporting Category E
Reporting Category A	0.902 (0.0022)***				
Reporting Category B		0.5611 (0.0026)***			
Reporting Category			0.6562 (0.0027)***		
Reporting Category D				0.6194 (0.0026)***	
Reporting Category E					0.6468 (0.0024)***
Constant	0.0033 (0.0016)**	0.3043 (0.0018)***	0.1779 (0.0020)***	0.1472 (0.0020)***	0.2041 (0.0019)***
Observations	119132	119132	119132	119132	119132
McFadden's Adjusted R^2	19.7888	0.9418	1.7412	-18.9595	3.3044

Robust standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
[†]Tobit regression bounded at 0 and 1

Table 1.17: Individual-Level Tobit Regression Results†: Reading Reporting Category without Controls

	(1) 2006 6^{th} Grade Reading Percentage Correct: Reporting Category A	(2) 2006 6^{th} Grade Reading Percentage Correct: Reporting Category B
2005 5^{th} Grade Reading Percentage Correct: Reporting Category A	0.8099 (0.0018)***	
2005 5^{th} Grade Reading Percentage Correct: Reporting Category B		0.6102 (0.0019)***
Constant	0.1237 (0.0013)***	0.1495 (0.0014)***
Observations	118,707	118,707
McFadden's Adjusted R^2	-1.8515	-1.1738

Robust standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
†Tobit regression bounded at 0 and 1

Table 1.18: Individual-Level Tobit Regression Results[†]: Math Reporting Category with Controls

	(1) Reporting Category A	(2) Reporting Category B	(3) Reporting Category C	(4) Reporting Category D	(5) Reporting Category E
Reporting Category A	0.7769 (0.0025)***				
Reporting Category B		0.4038 (0.0027)***			
Reporting Category C			0.5021 (0.0029)***		
Reporting Category D				0.4769 (0.0029)***	
Reporting Category E					0.4824 (0.0026)***
4Sight Math	0.0038 (0.0010)***	0.0021 (0.0014)	-0.0001 (0.0013)	0.0013 (0.0012)	-0.0089 (0.0012)***
Eligible for Tutoring	-0.0151 (0.0032)***	-0.0419 (0.0045)***	-0.0337 (0.0043)***	-0.0297 (0.0039)***	-0.0353 (0.0038)***
Receiving Tutoring	-0.0129 (0.0059)**	-0.0318 (0.0082)***	-0.0238 (0.0079)***	-0.0141 (0.0073)*	-0.0218 (0.0070)***
Female	0.0039 (0.0008)***	-0.0479 (0.0012)***	0.0047 (0.0011)***	-0.0014 (0.001)	-0.0058 (0.0010)***
Black	-0.0417 (0.0017)***	-0.0696 (0.0024)***	-0.0584 (0.0023)***	-0.0427 (0.0021)***	-0.0578 (0.0020)***
Hispanic	-0.0208 (0.0020)***	-0.0453 (0.0029)***	-0.0483 (0.0027)***	-0.0182 (0.0025)***	-0.0487 (0.0024)***
Other	0.0301 (0.0027)***	0.0535 (0.0039)***	0.0219 (0.0037)***	0.0469 (0.0034)***	0.0076 (0.0033)**
Homeless	-0.0233 (0.0132)*	-0.0329 (0.0185)*	-0.0131 (0.0177)	-0.0279 (0.0163)*	-0.0244 (0.0157)
Title I	-0.0106 (0.0015)***	-0.0325 (0.0022)***	-0.0078 (0.0021)***	-0.0187 (0.0019)***	-0.0211 (0.0018)***
Title III	0.0085 (0.0016)***	-0.0112 (0.0022)***	-0.0059 (0.0021)***	-0.0032 (0.0019)*	-0.0029 (0.0019)
Gifted	0.0978 (0.0019)***	0.1693 (0.0027)***	0.1267 (0.0025)***	0.1258 (0.0023)***	0.1397 (0.0023)***
IEP	-0.0621 (0.0013)***	-0.1334 (0.0017)***	-0.1245 (0.0017)***	-0.0982 (0.0016)***	-0.1067 (0.0015)***
Mean Teacher Experience	0.0004 (0.0001)**	0.0006 (0.0002)***	0.0017 (0.0002)***	0.001 (0.0002)***	0.0004 (0.0002)**
Student-Teacher Ratio	-0.0003 (0.0001)*	-0.001 (0.0002)***	-0.0008 (0.0002)***	0.0003 (0.0002)	-0.001 (0.0002)***
Free and Reduced Lunch Ratio	-0.0372 (0.0024)***	-0.0646 (0.0034)***	-0.0476 (0.0033)***	-0.0499 (0.0030)***	-0.0563 (0.0029)***
White Student Ratio	0.0045 (0.0029)	-0.0076 (0.0041)*	-0.0003 (0.0039)	-0.0031 (0.0036)	-0.0065 (0.0035)*
Weapons Violations Per Student	-1.3403 (0.1544)***	-1.8192 (0.2170)***	-2.3229 (0.2081)***	-1.2917 (0.1915)***	-1.3112 (0.1839)***
Ratio of Teachers with Master's Degree	0.0206 (0.0028)***	0.0353 (0.0039)***	0.0313 (0.0037)***	0.0406 (0.0034)***	0.0428 (0.0033)***
Constant	0.1061 (0.0046)***	0.4835 (0.0061)***	0.3113 (0.0060)***	0.2551 (0.0056)***	0.3797 (0.0053)***
Observations	118569	118569	118569	118569	118569
McFadden's Adjusted R^2	20.0500	1.3807	2.1677	-30.7020	4.0961

Standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
[†]Tobit regression bounded at 0 and 1

Table 1.19: Individual-Level Tobit Regression Results:[†]
Reading Reporting Category with Controls

	(1)	(2)
	2006 6th Grade Reading Percentage Correct: Reporting Category A	2006 6th Grade Reading Percentage Correct: Reporting Category B
2005 5th Grade Reading Percentage Correct: Reporting Category A	0.6899 (0.0022)***	
2005 5th Grade Reading Percentage Correct: Reporting Category B		0.4615 (0.0022)***
4Sight Reading	-0.0032 (0.0008)***	-0.0016 (0.0009)*
Eligible for Tutoring	-0.0044 (0.0025)*	-0.0243 (0.0028)***
Receiving Tutoring	-0.0162 (0.0046)***	-0.0127 (0.0052)**
Female	0.0021 (0.0006)***	0.0226 (0.0007)***
Black	-0.0238 (0.0013)***	-0.041 (0.0015)***
Hispanic	-0.0223 (0.0016)***	-0.0228 (0.0018)***
Other	-0.0005 (0.0021)	0.0221 (0.0024)***
Homeless	-0.0205 (0.0106)*	-0.0204 (0.0122)*
Title I	-0.0072 (0.0012)***	-0.0191 (0.0014)***
Title III	-0.0006 (0.0012)	-0.0005 (0.0014)
Gifted	0.0449 (0.0014)***	0.1017 (0.0016)***
IEP	-0.06 (0.0010)***	-0.0876 (0.0011)***
Mean Teacher Experience	0.0003 (0.0001)***	0 (0.0001)
Student-Teacher Ratio	-0.0002 (0.0001)**	-0.0009 (0.0001)***
Free and Reduced Lunch Ratio	-0.0262 (0.0019)***	-0.0365 (0.0022)***
White Student Ratio	0.0133 (0.0023)***	-0.0092 (0.0026)***
Weapons Violations Per Student	-1.3463 (0.1197)***	-1.5752 (0.1378)***
Ratio of Teachers with Master's Degree	0.0182 (0.0021)***	0.0444 (0.0024)***
Constant	0.212 (0.0036)***	0.282 (0.0040)***
Observations	118148	118148
	-1.9846	-1.4247

Standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
[†]Tobit regression bounded at 0 and 1

Table 1.20: School-Level OLS Results: Math

	(1)	(2)	(3)	(3a) Elasticities	(4)	(5)
	Log of Mean 2006 6th Grade Math Scaled Score					
Log of Mean 2005 5th Grade Math Scaled Score	0.9704 (0.0152)***	0.9705 (0.0152)***	0.7215 (0.0219)***		0.7285 (0.0230)***	0.7125 (0.0233)***
4Sight Math		0.001 (0.0029)	-0.002 (0.0027)		-0.0019 (0.0028)	-0.0021 (0.0028)
% of Students Eligible for Tutoring			0.0121 (0.014)	0.0003	0.0084 (0.0148)	0.0085 (0.0144)
% of Students Receiving Tutoring			-0.0501 (0.0288)*	-0.0004	-0.0485 (0.0294)*	-0.0454 (0.0295)
% Female			0.0506 (0.0123)***	0.0248	0.0507 (0.0132)***	0.0572 (0.0138)***
% Black			-0.0468 (0.0064)***	-0.0100	-0.0488 (0.0067)***	-0.0446 (0.0069)***
% Hispanic			-0.0459 (0.0108)***	-0.0024	-0.0471 (0.0110)***	-0.0477 (0.0117)***
% Other			0.0501 (0.0275)*	0.0010	0.046 (0.0274)*	0.0558 (0.0280)**
% Homeless			0.0094 (0.1624)	0	-0.0100 (0.1600)	0.1714 (0.2078)
% Title I			0.0032 (0.0045)	0.0011	0.0074 (0.0048)	0.0057 (0.0048)
% Title III			0.0020 (0.0046)	0.0002	0.0018 (0.0050)	0.0036 (0.0050)
% Gifted			0.0784 (0.0232)***	0.0038	0.0753 (0.0239)***	0.0724 (0.0249)***
% IEP			-0.0372 (0.0128)***	-0.0062	-0.0544 (0.0140)***	-0.0345 (0.0141)**
Mean Teacher Experience			0.0006 (0.0003)*	0.0088	0.0008 (0.0004)**	0.0005 (0.0004)
Student-Teacher Ratio			-0.0003 (0.0003)	-0.0042	-0.0002 (0.0003)	0 (0.0004)
Free and Reduced Lunch Ratio			-0.0212 (0.0059)***	-0.0081	-0.0195 (0.0061)***	-0.0239 (0.0065)***
Weapons Violations Per Student			-0.6036 (0.1799)***	-0.0013	-0.8745 (0.4042)**	-1.2814 (0.4305)***
Ratio of Teachers with Master's Degree			0.0287 (0.0074)***	0.0118	0.0239 (0.0079)***	0.0337 (0.0079)***
Mean Praxis Reading Percentile					0.0101 (0.0247)	
Mean Praxis Writing Percentile					0.0881 (0.0316)***	
Mean Praxis Math Percentile					-0.0396 (0.0192)**	
Mean NTE Common Knowledge Percent Correct						0.0652 (0.0402)
Mean NTE General Knowledge Percent Correct						-0.0596 (0.0366)
Mean NTE Professional Knowledge Percent Correct						0.0036 (0.0210)
Constant	0.2027 (0.1099)*	0.202 (0.1099)*	1.9872 (0.1606)***		1.9013 (0.1677)***	2.0367 (0.1716)***
Observations	1159	1159	1148		1057	1002
Adjusted R-squared	0.7794	0.7793	0.8198		0.8258	0.8192

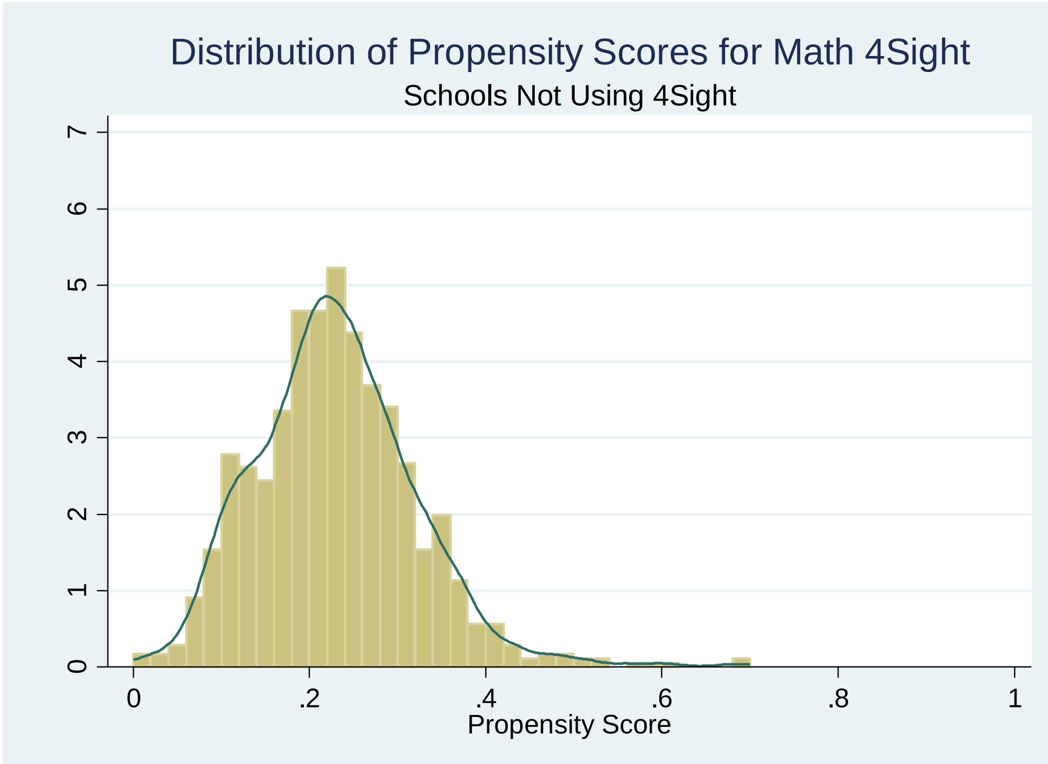
Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%

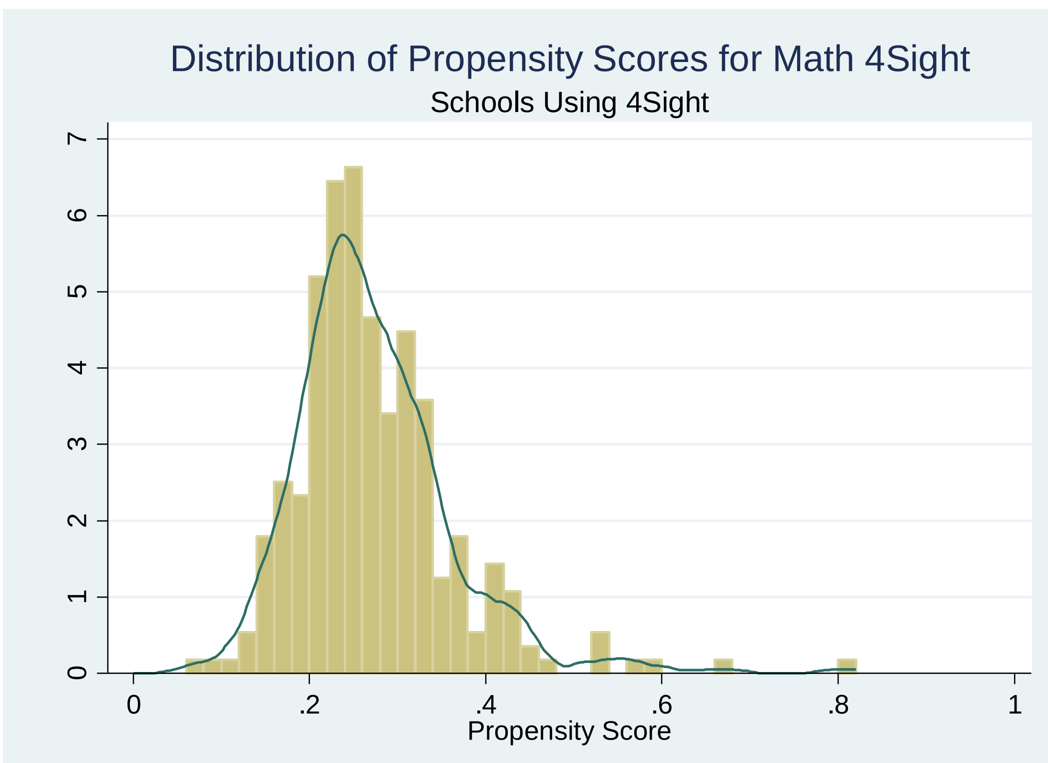
Table 1.21: School-Level OLS Results: Reading

	(1)	(2)	(3)	(3a) Elasticities	(4)	(5)
	Log of Mean 2006 6th Grade Reading Scaled Score					
Log of Mean 2005 5th Grade Reading Scaled Score	0.8661 (0.0105)***	0.8661 (0.0105)***	0.6893 (0.0189)***		0.7006 (0.0202)***	0.7045 (0.0196)***
4Sight Reading		-0.0001 (0.0023)	-0.0006 (0.0023)		-0.0018 (0.0023)	-0.0007 (0.0023)
% of Students Eligible for Tutoring			0.0184 (0.0118)	0.0005	0.0248 (0.0121)**	0.0291 (0.0115)**
% of Students Receiving Tutoring			-0.021 (0.0243)	-0.0002	-0.0258 (0.0241)	-0.0377 (0.0235)
% Female			0.0111 (0.0103)	0.0054	0.0405 (0.0108)***	0.0303 (0.0110)***
% Black			-0.0393 (0.0055)***	-0.0084	-0.0369 (0.0057)***	-0.0358 (0.0056)***
% Hispanic			-0.0336 (0.0093)***	-0.0017	-0.0318 (0.0092)***	-0.0326 (0.0095)***
% Other			0.0546 (0.0231)**	0.0011	0.053 (0.0223)**	0.0579 (0.0223)***
% Homeless			-0.171 (0.1369)	-0.0002	-0.1766 (0.131)	0.124 (0.1657)
% Title I			0.0075 (0.0038)*	0.0025	0.0065 (0.0039)	0.0059 (0.0038)
% Title III			0.003 (0.0039)	0.0003	0.002 (0.0041)	0.0031 (0.004)
% Gifted			0.0487 (0.0197)**	0.0024	0.0547 (0.0199)***	0.057 (0.0201)***
% IEP			-0.0618 (0.0112)***	-0.0103	-0.0611 (0.0117)***	-0.0702 (0.0115)***
Mean Teacher Experience			-0.0002 (0.0003)	-0.0024	-0.0001 (0.0003)	-0.0004 (0.0003)
Student-Teacher Ratio			-0.0006 (0.0003)**	-0.0093	-0.0004 (0.0003)	-0.0001 (0.0003)
Free and Reduced Lunch Ratio			-0.0133 (0.0050)***	-0.0051	-0.0047 (0.0051)	-0.0072 (0.0052)
Weapons Violations Per Student			-0.4763 (0.1525)***	-0.0010	-1.3929 (0.3329)***	-1.4205 (0.3457)***
Ratio of Teachers with Master's Degree			0.0338 (0.0062)***	0.0138	0.0329 (0.0065)***	0.0346 (0.0063)***
Mean Praxis Reading Percentile					0.0053 (0.0202)	
Mean Praxis Writing Percentile					0.0675 (0.0259)***	
Mean Praxis Math Percentile					-0.0262 (0.0158)*	
Mean NTE Common Knowledge Percent Correct						0.0544 (0.0321)*
Mean NTE General Knowledge Percent Correct						-0.0699 (0.0292)**
Mean NTE Professional Knowledge Percent Correct						0.0182 (0.0167)
Constant	0.9655 (0.0757)***	0.9655 (0.0757)***	2.2468 (0.1372)***		2.1158 (0.1455)***	2.1185 (0.1430)***
Observations	1159	1159	1148		1057	1002
Adjusted R-squared	0.8535	0.8534	0.8705		0.8798	0.8835

Absolute value of t statistics in parentheses * significant at 5%; ** significant at 1%

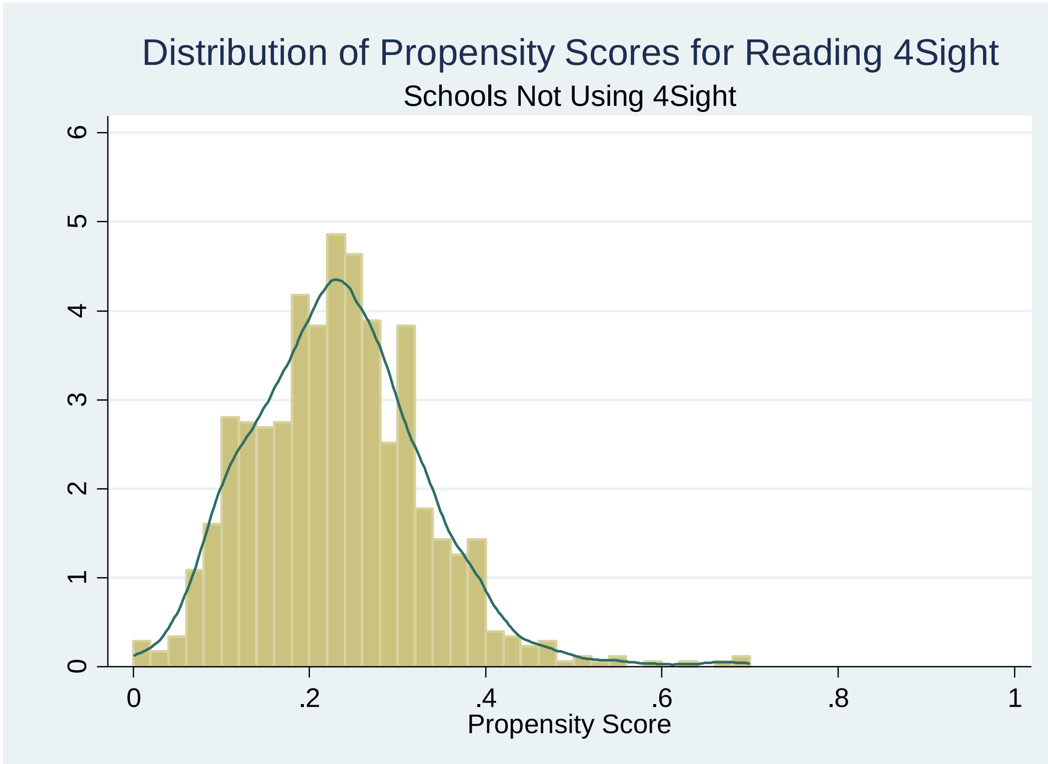
Figure 1.1: Propensity Scores for Math 4Sight Use†

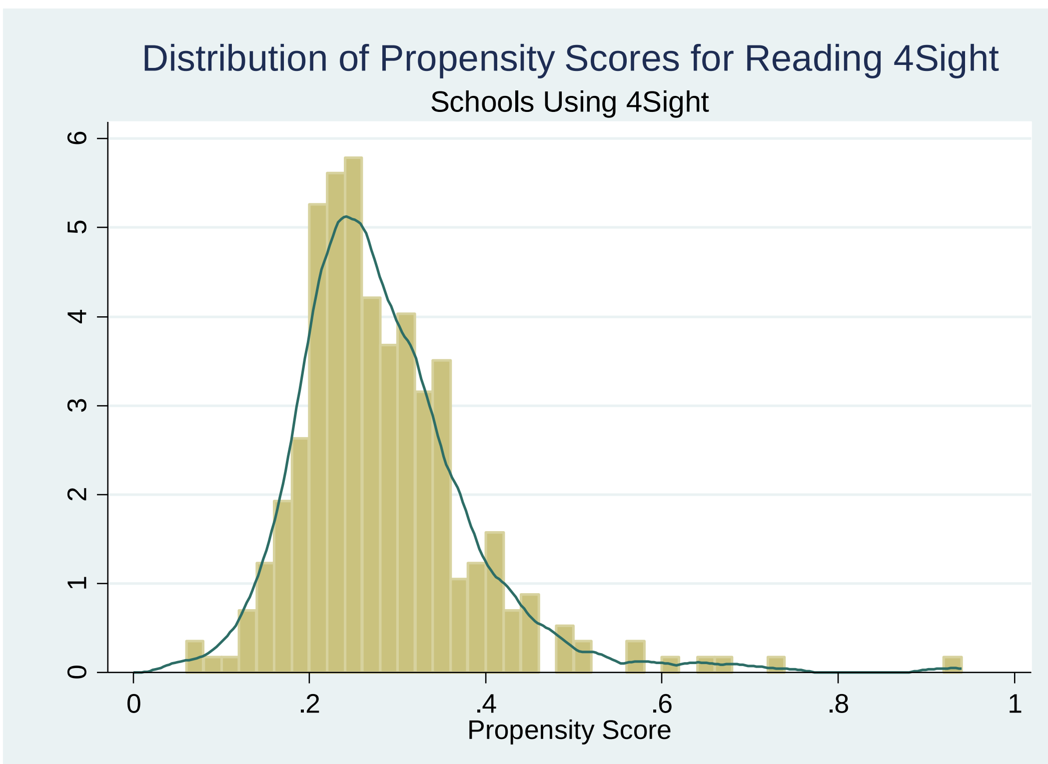




†The distributions of propensity scores for schools using 4Sight and schools not using 4Sight are statistically different at the 1% level. A two-sample Kolmogorov-Smirnov test was performed.

Figure 1.2: Propensity Scores for Reading 4Sight Use†





†The distributions of propensity scores for schools using 4Sight and schools not using 4Sight are statistically different at the 1% level. A two-sample Kolmogorov-Smirnov test was performed.

Table 1.22: School-Level Nearest-Neighbor Matching Model Results: ATT of Math 4Sight on Mean Math PSSA Scaled Scores

ATT	Standard Error†	Z-Statistic	N (Treatment)	N‡ (Control)
-12.432	**9.997**	**-1.24**	**279**	**209**

†Estimated using the Abadie and Imbens (2006) population variance estimator.
‡Actual nearest neighbor matches.

Table 1.23: School-Level Nearest-Neighbor Matching Model Results: ATT of Reading 4Sight on Mean Reading PSSA Scaled Scores

ATT	Standard Error†	Z-Statistic	N (Treatment)	N‡ (Control)
-19.013**	**9.472**	**-2.01**	**285**	**214**

†Estimated using the Abadie and Imbens (2006) population variance estimator.

‡Actual nearest neighbor matches.

**Statistically significant at the 5%-level.

1.9 Appendix

Table 1.A.1: Table of Reporting Categories, Anchors, and Eligible Content Covered on the 6th Grade Math PSSA*

Reporting Category	Assessment Anchor	Eligible Content	% of Eligible Content in Reporting Category Covered by 4sight
Numbers and Operations (A)	Demonstrate an understanding of numbers, ways of representing numbers, relationships among numbers and number systems. (A.1)	Represent common percents as fractions and/or decimals. (A.1.1.1)	91.6%
		Convert between fractions and decimals and/or differentiate between a terminating decimal and a repeating decimal. (A.1.1.2)	
		Represent a number in exponential form. (A.1.1.3)	
		Represent a mixed number as an improper fraction. (A.1.1.4)	
		Compare and/or order whole numbers, mixed numbers, fractions and/or decimals. (A.1.2.1)	
		Find the Greatest Common Factor (GCF) of two numbers and/or use the GCF to simplify fractions (A.1.3.1)	
		Find the Least Common Multiple (LCM) of two numbers and/or use the LCM to find the common denominator of two fractions. (A.1.3.2)	
		Use divisibility rules for 2, 3, 5, and/or 10 to draw conclusions and/or solve problems. (A.1.3.3)	
		Model percents using drawings, fraphs and/or sets. (A.1.4.1)	
	Understand the meanings of operations, use operations and understand how they relate to each other. (A.2)	Complete equation by using the following properties: associative, commutative, distributive and identity. (A.2.1.1)	
	Compute accurately and fluently and make reasonable estimates. (A.3)	Use estimation to solve problems involving whole numbers and decimals. (A.3.1.1)	
		Solve problems involving operations with whole numbers, decimals and fractions. (A.3.2.1)	
Measurement (B)	Demonstrate an understanding of measurable attributes of objects and figures, and the units, systems and processes of measurement. (B.1)	Determine and/or compare elapsed time to the minute. (B.1.1.1)	60%
	Apply appropriate techniques, tools and formulas to determine measurements. (B.2)	Use or read a ruler to measure to the nearest 1/16 inch or millimeter. (B.2.1.1)	
		Choose the more precise measurement of a given object. (B.2.1.2)	
		Find the perimeter of any polygon. (B.2.2.1)	
		Define, label and/or identify right, straight, acute and obtuse angles. (B.2.3.1)	
Geometry (C)	Analyze characteristics and properties of two-	*Identify, classify and/or compare polygons. (C.1.1.1)*	57.1%
		Identify and/or describe properties of all types of	

	and three-dimensional geometric shames and demonstrate understanding of geometric relationships. (C.1)	triangles. (C.1.1.2)	
		Identify and/or determine the measure of the diameter and/or radius of a circle. (C.1.1.3)	
		Identify and/or use the total number of degrees in a triangle, quadrilateral and/or circle. (C.1.1.4)	
		Identify, describe and/or label parallel, perpendicular or intersecting lines. (C.1.2.1)	
		Identify, draw and/or label points, planes, lines, line segments, rays angles and vertices. (C.1.2.2)	
	Locate points or describe relationships using the coordinate plane. (C.3)	Plot, locate or identify points in Quadrant I and/or on the x and y axes with intervals of 1, 2, 5 or 10 units. (C.3.1.1)	
Algebraic Concepts (D)	Demonstrate an understanding of patterns, relations and functions. (D.1)	Create, extend or find a missing element in a pattern displayed in a table, chart or graph. (D.1.1.1)	100%
		Determine a rule based on a pattern or illustrate a pattern based on a given rule. (D.1.2.1)	
	Represent and/or analyze mathematical situations using numbers, symbols, words, tables and/or graphs. (D.2)	Identify the inverse operation needed to solve a one-step equation. (D.2.1.1)	
		Solve a one-step equation. (D.2.1.2)	
		Match an equation or expression involving one variable, to a verbal math situation. (D.2.2.1)	
Data Analysis and Probability (E)	Formulate or answer questions that can be addressed with data and/or organize, display, interpret or analyze data. (E.1)	Analyze data and/or answer questions pertaining to data represented in frequency tables, circle graphs, double bar graphs, double line graphs or line plots. (E.1.1.1)	83.3%
		Choose the appropriate representation for a specific set of data. (E.1.1.2)	
		Display data in frequency tables, circle graphs, double-bar graphs, double line graphs or line plots using a title, appropriate scale, labels and a key when needed. (E.1.1.3)	
	Select and/or use appropriate statistical methods to analyze data. (E.2)	Determine the mean, median, mode and/or range of displayed data. (E.2.1.1)	
	Understand and/or apply basic concepts of probability or outcomes. (E.3)	Define and/or find the probability of a simple event. (E.3.1.1)	
		Determine/show all possible combinations involving no more than 20 total arrangements. (E.3.1.2)	

*Eligible content italicized are not covered by 4sight

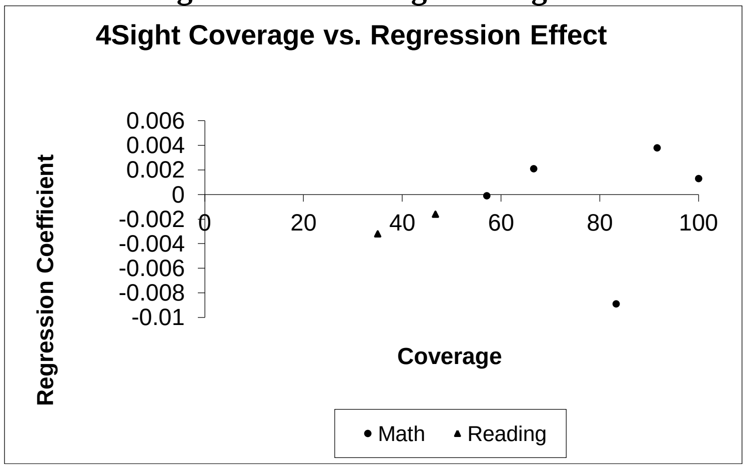
Table 1.A.2: Table of Reporting Categories, Anchors, and Eligible Content Covered on the 6th Grade Reading PSSA*

Reporting Category	Assessment Anchor	Eligible Content	% of Eligible Content in Reporting Category Covered by 4sight
Comprehension and Reading Skills (A)	Understand fiction appropriate to grade level. (A.1)	*Identify and/or apply meaning of multiple-meaning words used in text. (A.1.1.1)*	35%
		Identify and/or apply a synonym or antonym of a word used in text. (A.1.1.2)	
		Identify how the meaning of a word is changed when an affix is added; identify the meaning of a word from the text with an affix. (A.1.2.1)	
		Define and/or apply how the meaning of words or phrases changes when using context clues given in explanatory sentences. (A.1.2.2)	
		Make inferences and/or draw conclusions based on information from text. (A.1.3.1)	
		Cite evidence from text to support generalizations. (A.1.3.2)	
		Identify and/or explain stated or implied main ideas and relevant supporting details from text. (A.1.4.1)	
		Summarize the key details and events of a fictional text as a whole. (A.1.5.1)	
		Identify the author's intended purpose of text. (A.1.6.1)	
		Identify, explain, and/or describe examples of text that support the author's intended purpose. (A.1.6.2)	
	Understand nonfiction appropriate to grade level. (A.2)	*Identify and apply meaning of multiple-meaning words used in text. (A.2.1.1)*	
		Identify and apply meaning of content- specific words used in text. (A.2.1.2)	
		Identify and apply how the meaning of a word is changed when an affix is added; identify and apply the meaning of a word from the text with an affix. (A.2.2.1)	
		Define and/or apply how the meaning of words or phrases changes when using context clues given in explanatory sentences. (A.2.2.2)	
		Make inferences and/or draw conclusions based on information from text. (A.2.3.1)	
		Cite evidence from text to support generalizations. (A.2.3.2)	
		Identify and/or explain stated or implied main ideas and relevant supporting details from text. (A.2.4.1)	
		Summarize the major points, processes, and/or events of a nonfictional text as a whole. (A.2.5.1)	
		Identify the author's intended purpose of text. (A.2.6.1)	
		Identify, explain, and/or describe examples of text that support the author's intended purpose. (A.2.6.2)	

Interpretation and Analysis of Fictional and Nonfictional Text (B)	Understand components within and between texts. (B.1)	Identify, explain, interpret, compare, describe, and/or analyze components of fiction and literary nonfiction. (B.1.1.1)	46.7%
		Identify, explain, interpret, compare, describe, and/or analyze connections between texts. (B.1.2.1)	
	Understand literary devices in fictional and nonfictional text. (B.2)	Identify, explain, interpret, and/or describe examples of personification in text. (B.2.1.1)	
		Identify, explain, interpret, and/or describe examples of similes in text. (B.2.1.2)	
		Identify, explain, interpret, and/or describe examples of alliteration in text when its use is presumed intentional. (B.2.1.3)	
		Identify, explain, interpret, and/or describe examples of metaphors in text. (B.2.1.4)	
		Identify, explain, and/or describe the point of view of the narrator as first person or third person point of view. (B.2.2.1)	
		Explain, interpret, and/or describe the effectiveness of the point of view used by the author. (B.2.2.2)	
	Understand concepts and organization of nonfictional text. (B.3)	Identify, explain, and/or interpret statements of fact and opinion in nonfictional text. (B.3.1.1)	
		Identify exaggeration (bias) in nonfictional text. (B.3.1.2)	
		Identify, explain, and/or interpret how the author uses exaggeration (bias) in nonfictional text. (B.3.2.1)	
		Identify, explain, and/or interpret text organization, including sequence, question/answer, comparison/contrast, cause/effect, or problem/solution. (B.3.3.1)	
		Use headings to locate information in a passage, or identify content that would best fit in a specific section of text. (B.3.3.2)	
		Interpret graphics and charts and/or make connections between text and content of graphics and charts. (B.3.3.3)	
		Identify, explain, compare, interpret, describe, and/or analyze the sequence of steps in a list of directions. (B.3.3.4)	

*Eligible content italicized are not covered by 4sight

Figure 1.A.1: 4Sight Math Coverage vs. Regression Coefficient



Chapter 2: Labor Strife in Public Schools: Does it Affect Education Production?

Abstract

Teacher strikes and the right of public employees to collectively bargain are topics of frequent and heated debate in the public sphere, with little research available to inform the debate. In firms, the negative relationship between labor unrest and reduced productivity is well-documented; the purpose of this study is to explore whether there exists a similar, measurable relationship between labor strife and productivity in public schools. Using regression analysis to analyze data that includes teacher strikes and expired contracts over a seven-year period in Pennsylvania, I find that the pass rates on a district-level cohort's math tests decrease by about 1-2 percentage points in the year of a strike and by about 0.5 percentage points during a year that teachers work under an expired contract. Additionally, cohorts experiencing a strike during their 11th-grade year realize about a 2 percentage-point decrease in their graduation rate. In addition to improving upon the methodologies of previous teacher strike papers, this paper distinguishes between productivity loss due to strikes and that due to lengthy ongoing labor disputes that do not necessarily end in strike. Policy implications include making administrators aware of the possible effect of a strike on graduation rates and the need for better collection of data on collective bargaining by state agencies.

2.1 Introduction

In the 2009-10 school year, teachers or employees from eight Pennsylvania public school districts went on strike (Templeton 2011). In Pennsylvania, as in twelve other states (Weaver 2007), strikes are a legal and regulated part of collective bargaining, a process in which, every few years, teachers' unions and local school boards meet to agree upon pay, benefits, and working conditions. Theory and empirical evidence from manufacturing (Kleiner, Leonard, and Pilarski 1999; Mas 2007; Krueger and Mas 2003) and the service sector (Mas 2006) tell us that labor strife causes a decrease in employee productivity. Several papers (Baker 2011; Johnson 2011; Zwerling 2008; Thornicroft 1994; Zirkel 1992) have explored whether the same relationship that exists in firms, between production and labor strife, exists in public schools, between education production and teachers' labor disputes. These papers have found mixed evidence as to the empirical effect of teacher strikes on student test outcomes. The contribution of this paper to the literature is to further explore the relationship between education production and teachers' labor disputes, using data from Pennsylvania that includes strikes as well as long disputes over contracts that do not end in strike and two measureable student outcomes, performance on post-No Child Left Behind achievement tests, and graduation rates.

Assessing the relationship between labor strife and the productivity of schools requires measuring the effect of labor inputs on production in schools. Production in the schools is generally measured using achievement test scores, since these are well-documented predictors of future labor market outcomes.[26] A large literature examines achievement scores as the output of an education production function, in which an individual student's achievement is a function of endowed ability, family background, and school characteristics,[27] one of which would be the labor inputs of teachers. While assessing the impact of labor inputs on manufacturing production requires only that capital inputs be held fixed, identifying the impact of labor inputs on education production requires controlling for changes in student ability and family background. Students, their endowments, and their family backgrounds cannot be held fixed across

[26] See Cawley, Heckman, and Vytlacil (2001), for example.
[27] See Todd and Wolpin (2003) and the references therein.

different schools and over time; I account for this econometric difficulty by modeling the education production function and using either a cohort-based first-differences approach or a fixed-effects approach to control for unobserved variation in student endowments and backgrounds across different school districts.

Collective bargaining, and strikes in particular, can be extremely divisive among communities, often affecting the outcomes in school board elections and inconveniencing parents, students, and businesses when school is cancelled. The laws regarding bargaining and conflict resolution differ greatly from state to state; thirty-three states have collective bargaining laws for teachers and these laws vary in the scope of issues that may be included in bargaining: from wages and hours only to comprehensive curriculum plans and classroom management issues. Among states allowing collective bargaining there are an array of policies regarding impasse procedures; thirty-one states use third-party mediation, twenty-eight use fact-finding procedures, eighteen have voluntary arbitration, and four mandate arbitration (Krueger 2002). Teachers are allowed to strike in thirteen states and prohibited from striking in thirty-seven states, though not all of these have penalties for strikes (Weaver 2007), making it possible for some states to have banned teacher strikes but still experience them. In the state of Washington, strikes by public employees are illegal and yet they experienced eighty-four teacher strikes between 1972 and 2003 (Michaelis 2003). In Detroit, a September 2006 teacher strike deemed illegal by the state legislature went to court where the judge questioned the constitutionality of a law requiring judges to order striking public employees back to work and delayed making a decision (Macdonald and Jun 2006). Today, with education reform in the national spotlight, collective bargaining and the legal status of teacher strikes continue to make news as politicians and organizations spout opinions on both sides of the issue.[28] Measuring the effect of labor strife on education production provides policy-makers with the groundwork to make good decisions regarding collective bargaining in the public schools.

The empirical results of this study show that, in Pennsylvania, strikes have a negative impact on both test scores and graduation rates. At the grade-school and middle-school levels, a strike has a statistically significant, negative impact on the percentage of

[28] Most notably, the public and political debate in Wisconsin over Act 10 in 2011. See Kaufman (2012).

students that pass their annual math tests of about 1.1 percentage points, and at the high-school level the negative impact is about 2.2 percentage points. At neither level is there a significant effect on the percent of students passing their annual reading tests, though point-estimates in both cases are also negative. A strike also has a statistically significant and negative impact on graduation rate; the effect is about a 2.2 percentage-point decrease for students having experienced a strike in their 11th grade year. Finally, when teachers work under an expired contract for an extended period of time, this also has a statistically significant, negative effect of about 0.4 percentage points on the pass rates of elementary and middle school students on math tests, suggesting that it is not only the gap in instructional time causing a decrease in the percentage of students passing their exams, but also the effect of teachers' discontent with their working conditions.

2.2 Institutional Details

Pennsylvania is a particularly interesting state in terms of labor strife in schools because it consistently leads the country in number of teacher strikes. According to the Pennsylvania School Boards Association (PSBA), 21% of all public school strikes in the U.S. since 1971 have occurred in Pennsylvania. Since 1992, the state has experienced an average of twelve public school strikes annually. Pennsylvania has a rich history of unionization in its large steel, coal, textiles, and railroad industries,[29] which may help explain the abundance of teacher strikes occurring in the state; some have attributed causality to differences in the laws governing strikes in the state.[30] Regardless of the reason for the prevalence of teacher strikes, Pennsylvania presents an opportunity to study the effects of labor strife on the education production in schools.

Collective bargaining and strikes in Pennsylvania are governed by the Public Employe Relations Act of 1970 and Act 88, an amendment made in 1992. Together, these laws allow public employees and employers to engage in collective bargaining and define its scope and impasse protocols. Collective bargaining is a process in which two parties, the school district (employer) and the teachers' union (representing the

[29] See the *Pennsylvania Center for the Study of Labor Relations*, < http://www.hhs.iup.edu/laborcenter/>.
[30] Gamrat and Haulk (2006) attribute the difference between Pennsylvania and other states allowing teachers to strike to "no penalty for strikers…limited and untested voter referendum control over school spending or taxes and no Right to Work law."

employees), meet to agree upon a new contract. These contracts are generally specific to teachers, i.e., separate from the contracts of other employees within a school (such as custodians or secretaries). Each contract specifies effective and expiration dates; there is no predetermined length of time for which a contract will last (this varies by district) nor do consecutive contracts necessarily last the same amount of time.[31] The previous contract specifies a date by which negotiations on a new contract must begin; this date is usually about a year before the expiration date of the previous contract. A new contract may or may not be settled upon before the expiration date of the current contract, meaning that teachers may work indefinitely without an up-to-date contract. When this occurs, teachers work under the old contract until such time that a new contract is agreed upon; any changes in pay or benefits are usually applied retroactively upon agreement on a new contract.

In addition to providing a mandatory timeline for bargaining, Pennsylvania law provides mandatory, state-trained mediators if a contract is not agreed upon within a set period of time.[32] At this time, the state will also provide fact-finding. A strike may occur at any time during the negotiation process, as agreed upon by the teachers in the district.[33] Strikes are regulated minimally by law: 48-hour advance notice is required before any strike, and schools are required to make up for days missed due to a strike.[34] This occasionally results in spring strikes ending before resolution of bargaining issues. Importantly for my purposes, this requirement also ensures that any difference in educational outcomes between a striking district and a non-striking district cannot be attributed to fewer instructional days.

State and local policies such as Pennsylvania's Public Employe Relations Act determine the nature of the relationships between school boards and administrators (managers) and teachers (employees). States differ greatly in the policies they choose: some outlaw collective bargaining and teachers' unions altogether and provide state-wide

[31] In fact, contract length can be a bargaining point.
[32] Mediation is required if an agreement has not been reached within 45 days of the start of negotiations, or by 126 days before the budget submission date, around the end of February.
[33] More detailed information on the collective bargaining process can be found in Public School Negotiations: a Complete Guide to Collective Bargaining in Pennsylvania Public Education (Pennsylvania School Boards Association [1993]).
[34] Districts are required to meet a minimum period of instruction of 180 days by the later of June 15 or the last day of the scheduled school year in the district.

teacher pay scales; others allow collective bargaining but outlaw strikes. Empirical research on the effects of policy on student learning, the end goal of compulsory education and all laws surrounding it, is essential, useful, and informative for policymakers to have. In the interest of such research, I use data from Pennsylvania to empirically measure the effect of labor unrest arising from this particular state's collective bargaining policy on student outcomes.

2.3 Conceptual Framework

This research is related to empirical papers from the industrial relations literature studying the relationship between labor strife and production in manufacturing. Evidence from these studies demonstrates a negative relationship between labor unrest and productivity. Kleiner, Leonard, and Pilarski (2002) studied industrial relations in a commercial aircraft manufacturing plant, finding that "strikes, slowdowns, and tough union leaders influenced the productivity of this plant by both large percentages and absolute dollar amounts during the period they were occurring." Productivity within this plant took one to four months to return to pre-event levels. Mas (2008) found that workmanship at construction equipment factories that had experienced contract disputes in the past was significantly lower than that at factories without labor unrest; the products produced in facilities with contract disputes were "resold more often, received worse appraisal reports, and had lower list prices" than those produced in facilities without labor unrest The labor strife studied in this paper took place over a seven-year period in the 1990's and cost an estimated $400 million in lost revenues. Krueger and Mas (2004) studied a long strike and the hiring of replacement workers at a tire plant in the mid-1990's, finding that low product quality coincided with periods of labor strife, particularly when replacement workers worked side by side with returning strikers.

While the previous evidence is from the manufacturing sector and measures output in terms of the quality of a physical product, Mas (2006) provides a precedent for investigating the relationship between labor strife and productivity in the service sector, measuring police productivity in particular. Using final offer arbitration data from the New Jersey police, Mas finds a negative relationship between labor strife and police productivity; police officers who lose in arbitration had lower arrest rates and average sentence lengths following arbitration relative to those who win, after controlling for differences between the two groups. As in my paper, police "output" is not directly observable but measured using endogenous proxies that are noisy measures of actual police performance. If production in schools is analogous to production in firms, the findings from the manufacturing literature suggest that education production would similarly suffer during periods of labor strife. The Mas (2006) paper provides further evidence that the analogy extends to the public service sector, of which teachers providing educational services are a part.

In fact, Todd and Wolpin (2001) define the education production function as "an analogy between the knowledge acquisition process of a human being and the production process of a firm." Production in schools, however, differs from production in manufacturing in several ways. In manufacturing, the inputs are relatively simple: current labor and capital factors determine the quality and quantity of output. In education production, it is customary to define student *i*'s actual achievement in school at time *t* where (*A*) is a function of past and present family inputs (*X*), past and present school inputs (*S*), and the student's endowed ability (μ).

(1.1) $A_{ist} = f\left(X_{it}, S_{it}, \mu_i\right)$

Together, the cumulative nature of the inputs (which require a large amount of usually unavailable data) and the student's endowed ability (which is econometrically unobservable) cause the ceteris paribus empirical analysis of the education production function to be relatively difficult compared to analysis of the manufacturing production function. Furthermore, as opposed to manufacturing, in which the quality and quantity of output can be directly measured, a student's actual achievement in school is econometrically unobservable; achievement is observed only through a test score (*T*) that is a noisy measure of his or her actual achievement.

(1.2) $T_{ist} = g\left(A_{ist}, \varepsilon_{ist}\right)$

In a very general way, equations (1.1) and (1.2) describe the relationship between ability, inputs, achievement, and test scores for student *i*.

Several papers have explored the relationship between labor strife and school output in the past. Recently, Baker (2011) studied the impact of eleven teacher strikes on school-level cohorts of elementary students in Ontario, Canada. Baker uses first-differences regression and matching models to evaluate the effect of a strike on student test-scores, finding negative and significant effects for long strikes (greater than 10 days) occurring during a student cohort's 2^{nd}- or 6^{th}-grade year. While this paper is methodologically similar to Baker's, my data provide test outcomes for a greater span of grades on which to analyze the effect of strikes, observe many more strikes (56 as opposed to 11), and observe periods of labor strife that do not end in a strike (periods during which the contract is expired), and will thus contribute to the literature. Johnson (2011) studied strikes in Ontario during the same time-period as Baker, but also included an analysis of "work-to-rule" campaigns as part of his analysis; he finds that strikes have a negative impact on 3^{rd} and 6^{th} grade test scores, particularly in low-income schools, and that work-to-rule campaigns have a smaller, negative impact on test scores.

Additional studies from the economics of education suggest that a relationship might exist between teacher labor strife and education production. Using matched student-teacher panel data, Rockoff (2004) finds that teacher quality, as measured by their fixed effects on student outcomes, has a statistically significant effect on student achievement. A teacher's quality can be thought of as the element contained in S_{it} through which labor unrest could enter the education production function.[35] While the politics of collective bargaining creates obstacles to collecting data on teacher attitudes and activities during times of labor unrest, the few studies that have been done suggest a change occurs. Griffin, Tesluk, and Jacobs (1995) surveyed teachers in schools in Pennsylvania over a three-year period, during which time teachers were at different positions in the bargaining cycle. Using data on official bargaining year, contract settlement data, contract negotiation period, and surveys in which teachers rated their satisfaction with pay, benefits, administration, and teaching, the authors found that teachers are less satisfied with their pay and benefits in bargaining years than in the years directly preceding and proceeding. They also found that teachers within a school converge in their attitudes toward pay, benefits, and administration during contract negotiation. Carlton and Johnson (1980) surveyed school board members and teachers in a Virginia school district and found a substantial difference in their attitudes towards collective negotiations,

[35] Rowan, Chiang, and Miller (1997) found teachers affect student achievement positively in three ways: according to the teachers' ability, motivation, and work situation.

strikes, and sanctions, describing the relationship between teachers and school board members as a “have-have not paradigm.” Alternately, labor unrest might enter the education production function through a change in student attitudes[36] contained in X_{it}. There is evidence that student attitudes change during and after teacher strikes: studying student attitudes during and after an 8-week faculty strike at York University, Grayson (1997) found that a majority of students remained bitter towards the strike after it ended, and that student satisfaction with academic programs remained lower than before the strike.

2.4 Data

Student academic achievement is measured annually in Pennsylvania using the Pennsylvania System of School Assessment (PSSA). The PSSA is a standards-based assessment specifically aligned with curriculum and instruction in the public schools. For each grade level and subject area, “assessment anchors,” published and distributed by the Pennsylvania Department of Education (PDE), determine the skills and procedures that students will learn and be tested over. Since the inception of NCLB in 2001, the number of tests students take has increased; in 2001, students were tested in math and reading in grades 5, 8, and 11, and in writing in grades 6, 9, and 11. In 2010, students were tested in math and reading in grades 3-8 and 11, in writing in grades 5, 8, and 11, and in science in grades 4, 8, and 11. The math, reading, and science tests consist of multiple choice questions as well as some free-response questions in which students must explain their thinking in writing. The writing test consists of written essays. One of the explicit purposes of the PSSA is to “provide information to state policymakers…on how effective schools are in both promoting and demonstrating student proficiency of academic standards,” making the scaled scores for reading and math tests the most appropriate measure of educational output available for this study.

Raw scores from these tests are used to compute scaled scores, which adjust for test length and item difficulty, and can be used to roughly measure achievement on academic standards within and across years (Handbook for Report Interpretation 2002). A further requirement of NCLB is that states set cutoff scores for categorizing a student’s

[36] Summers and Wolfe (1977) found that student motivation, which they measured using lateness and unexcused absences, have a significant and positive relationship with achievement.

performance into one of four categories: below basic, basic, proficient, and above proficient. States then must set yearly goals stating the percentage of students required to pass the reading and math tests with a score of "proficient" or above in order for the school or district to be considered as having made "adequate yearly progress," or AYP. The federal law allows states to set their own trajectories in order to move from a relatively low pass-rate in 2002 to the goal of 100% proficiency in 2014. Pennsylvania's goals for math and reading proficiency pass rates can be seen in Figures 1 and 2.

The percentage of students with scores in each performance category on math and reading tests for the school years 2003-04 through 2009-2010, by district, were computed by the PDE. For the purpose of this paper, as well as AYP, the "pass rate" is the sum of the students scoring in the "proficient" and "above proficient" categories. Descriptive statistics for these district-level pass rates, by grade level, are presented in Tables 1 and 2. All tests show substantial increases in average pass rate over time, probably due to the increasing pressure associated with meeting AYP year after year. Figures 1 and 2 illustrate this time-trend. Additionally, the tests appear to become more difficult to pass as grade-level increases.[37]

An additional outcome of interest to the public is the graduation rate of schools and districts. Since a calculation of the graduation rate is required as part of meeting AYP, work is currently being done by the government to assure that all schools and districts will calculate their graduation rates the same way. Precise data on district drop-out rates for the years in my data set are not available in Pennsylvania; however, the CCD has collected data since the 2006-07 school year which allows it to construct an approximate graduation rate for each district.[38] Graduation rate will be used as an alternate dependent variable at the high-school level.

---

[37]The Fordham Institute issued a report in 2007 (Cronin, Dahlin, Adkins, and Kingsbury 2007) with evidence that improvement in state test scores over time are not usually a signal of improved learning. This makes it very important to control for the time-trend in this research. In addition, the report concludes that 8^{th} grade tests are significantly harder to pass in most states than those from previous grades, even taking into account the obvious increase in difficulty of the subject matter; my data confirms this observation. Pennsylvania was not one of the states studied in the Cronin et al report.

[38] The graduation rate is constructed by dividing the number of graduates in a district by the average number of students in the same cohort when they were in 8^{th}, 9^{th}, and 10^{th} grades:

$$\text{graduation rate} = graduates_t \Bigg/ \left(\frac{8^{th}\ graders_{t-4} + 9^{th}\ graders_{t-3} + 10^{th}\ graders_{t-2}}{3} \right).$$

Proxy Variables Measuring Labor Unrest

Labor unrest in the public schools is measured using data on teacher strikes and contract cycles by district and year. It seems reasonable to assume that a labor relations event, such as a strike or the expiration of a contract before the signing of a new contract, will trigger a change in teacher attitudes (an element of S_{it}) and/or student attitudes (an element of X_{it}) that accompany such an event. Ideally, the dates and lengths of strikes, when they occur in relation to the signing of a new contract, the outcomes of strikes, i.e., whether or not teachers "won" in the bargaining process, the start and end dates of all contracts, and the signing dates of all contracts (allowing me to observe periods when teachers worked without a contract) would all be used to create detailed proxy variables for labor unrest. Unfortunately, these data are not available, yet the Pennsylvania School Boards Association (PSBA) was able to provide me with the school years of all teacher strikes occurring between the 2003-04 and 2009-10 school years. Pennsylvania had 500 school districts during most of this period,[39] so over a seven-year period there were 56 strikes in 3498 school-district years, accounting for 1.6% of the observations.

Descriptive statistics comparing district-years with strikes over the eight-year period to the remaining, non-striking district-years are presented in Table 3. The statistically significant differences between these are a lower percentage of students eligible for free or reduced lunch and a higher pass rate on the 11th grade reading test for striking district-years compared to non-striking districts-years. Similar descriptive analysis is also available by district, rather than district-year, for the year 2009.[40] In that analysis, there were no statistically significant differences between the characteristics of districts that experienced at least one strike over the time period and districts that never experienced a strike during the time period. These numbers show that striking and non-striking districts are quite similar in their observed characteristics, and striking district-years and non-striking district-years are also very similar. A map of the geographic locations of striking districts, presented in Figure 3, shows that strikes do not follow a strong geographic pattern and are not significantly limited to either urban or rural areas. It does show that if

[39] In June of 2009, the Center Area and Monaca school districts combined to become Central Valley School District. Thus, from 2003-04 to 2008-09, Pennsylvania had 500 school districts, and in the 2009-10 school year there were 499 school districts. For the sake of this paper, I did not include Central Valley school district in any of the analyses, treating the 2009-10 school year as if there were 498 school districts.

[40] A table describing this data is available in the data appendix.

one district has more than one strike in the period, it is likely that other neighboring districts will also have more than one strike in the period.

The PSBA also provided me with teacher contract data, including expiration date, the term of the contract, and some information on salary scales, including the minimum and maximum pay for teachers with bachelor's degrees and master's degrees, and the number of steps on each pay scale, for the school years 2003-04 through 2009-10. The length of the contracts in the data set varies in time between one and nine years.

In any given year of PSBA contract data, there are some districts missing contract data. These missing data can be interpreted as years in which teachers in a district worked without a contract.[41] Table 4 shows the percentage of districts, in any given year, in which teachers worked an entire year without a contract. During any given school year during the period, an average of about 11% of districts have no current contract signed between the teachers' union and the school board.

Student, Family, and District Attributes

The data set contains control variables for each school district and year. Data on the percentage of students that qualify for free or reduced-price lunch (FRL), commonly used as a proxy for socio-economic status (SES),[42] and the percentage of students in each grade that are white, a proxy for family characteristics, come from the Common Core of Data (CCD). Data on spending per-student (in 2010 dollars), pupil-teacher ratio, and percentage of students on individualized education programs (IEPs) also come from the CCD, and serve as proxies for district attributes. Finally, the CCD provides data from the 2000 Census School District Demographic Project. This source provides data on per-capita income and the percentage of adults within a school district's boundaries that have a college education. Since these two variables are not time-varying, they are used only in the descriptive analysis comparing striking and non-striking school districts, and not in

[41] This data does not inform on partial years worked without a contract. Staff at the PSBA enters contract data into their computer system as they receive them, with about a two-month lag between the actual signing date and the date on which they receive a copy of the signed contract. This means that each contract appears in the data for the first time during the school year in which they received a signed copy and that districts must have gone almost an entire year without a contract before they will be reflected in the data as having been without a contract for a year.

[42] See Harwell and LeBeau (2010) for a discussion on the use of free-and-reduced-price lunch as a proxy for SES. I have chosen to use this measure despite its shortcomings because it is the only one readily available to me.

any of the regression analysis. Spending per-student data is also only used in the descriptive analysis.

Other Control Variables

A final set of dummy variables controls for test year. These help control for differences in the test and its scoring from year to year that are not corrected for in the scaling and selection of cutoff scores.

Instrumental Variables

Babcock, Want, and Loewenstein (1996) find that differences in the salary scales of "comparable" school districts that the unions and school boards, respectively, bring to the bargaining table are correlated with teacher strikes and uncorrelated with student outcomes in a district, and thus can be used as an instrument. The authors received survey data from about one hundred unions and school boards, obtaining their actual lists of "comparable" school districts used for collective negotiations, and were able to construct "comparables" for the remaining districts, using what they learned from the survey data, to some success. Given the descriptive statistics in Table 3 and the map in Figure 3, it seems to be a reasonable assumption that strikes are exogenous regressors in my empirical models. Nonetheless, an instrument correlated with the "strike" and "expired contract" variables but uncorrelated with the dependent variables would provide an alternate identification strategy without requiring the exogeneity assumption. Using data on each district's salary scales and a list of each district's neighboring districts,[43] I constructed lists of hypothetical "comparable" school districts for unions and school districts in order to construct potential instruments to strengthen my empirical analysis. The union "comparables" included the four highest-paying neighboring districts, and the school board "comparables" included the four lowest-paying neighboring districts. Four instruments resulted,[44] being the differences in the average comparables based on the minimum and maximum pay for teachers with Bachelor's degrees and the minimum and

[43] This list was constructed using a map of the Pennsylvania school districts obtained from the PDE's Office of Educational Technology, Division of Data Services.
[44] The construction of these instruments follows directly from evidence presented in Babcock, Want, and Loewenstein (1996). They found that each side used an average of 4.5 comparables from nearby, similar school districts, and that the unions tended to choose comparables with higher salaries, while school boards tended to choose those with lower salaries.

maximum pay for teachers with Master's degrees. Unfortunately, none of the four instruments were correlated in a statistically significant manner with either the strike or the expired contract variables,[45] thus IV estimation was not possible.

2.5 Empirical Model

The identification problem posed by lack of data on the large number of factors affecting each student's achievement in a school is generally addressed by using either a fixed-effects model[46] or a first-differences model,[47] both of which attempt to control for past inputs into education production. Here, inherent student abilities, which contain a student's unobserved history of inputs as well as immeasurable attributes, are allowed to vary by individual student. A typical first-differences model relating change in achievement to change in inputs is commonly linear in parameters and measures change in achievement (as measured by performance on a standardized test) of student *i* in school *s* and time *t* using the model:

$$\Delta T_{ist} = \Delta X_{ist}\beta_X + \Delta S_{st}\beta_S + \xi_{ist} \tag{1.3}$$

In this model, the change in academic achievement depends on the change in current family attributes (X), change in current school and teacher characteristics (S), and a random error (ξ). Here, inherent student abilities, which contain a student's unobserved history of inputs as well as immeasurable attributes, are allowed to cancel out of the difference equation.

While it would be desirable to observe individual students and characteristics, in this study I only have data aggregated to the school district-level. Thus to examine the effect of labor unrest on academic achievement, the model in equation (1.3) must be aggregated over individuals and schools within a district.

$$\overline{\Delta T_{dt}} = \overline{\Delta X_{dt}}\beta_X + \overline{\Delta S_{dt}}\beta_S + \overline{\xi_{dt}} \tag{1.4}$$

Remaining on the left-hand side of equation (1.4) is the change in average test performance for students in district *d* at time *t* and on the right-hand side, the change in average family characteristics, change in teacher and school characteristics, and random error for the district. The error term in equation (1.3) contains both random error and test

[45] First-stage regression results are available in the Appendix.
[46] For example, Todd and Wolpin (2006) or Goldhaber and Brewer (1997).
[47] For example, Rivkin, Hanushek, and Kain (2005) or Todd and Wolpin (2006).

measurement error. Any imprecision in the error term of equation (1.4), I assume, is uncorrelated with strikes or contract disputes, and thus does not pose a threat to identifying the effects of these variables.

I estimate the first-differences linear regression model in equation (1.4). Elements of $\overline{\Delta X_{dt}}$ include change in percent of students on FRL, change in percent of students that are white, and change in percent of students on an IEP.[48] Elements of $\overline{\Delta S_{dt}}$ include the change in pupil-teacher ratio, strike and lagged strike dummy variables, and a dummy variable indicating an expired contract. Also included are dummy variables to control for test year in order to pick up any systematic differences in performance common to all districts, specifically the upward time-trend visible in Figures 1 and 2. The first-differences model has an additional benefit beyond those offered in the fixed effects model for the case in which the data being used is aggregated to the district level; here, the change in student and school-level characteristics can also act as proxies for changes occurring in the student population over time. If I had student-level data, these changes would not be an issue; since the data is aggregated to the district-level, it is important to control for changes in the student population over time in order to identify the effect of a labor dispute on student performance.

The strike dummy variable (and lagged strike variables) attempt to pick up any changes in test scores due, for instance, to a change in teacher or student attitudes during the year of or years after a strike. The expired contract variable attempts to pick up changes in test scores due to changes in teacher attitudes related to ongoing contentious labor negotiations. For instance, it might be the case that working without a current contract causes teachers to feel that they need to prove their value to parents and administrators, and work harder to signal that they deserve a raise in pay. This might result in a boost to test scores during years without signed teacher contracts. Alternately, it might be the case that teachers feel disgruntled during tough labor negotiations, and their teaching might suffer because of it, causing test scores to fall in these years.

[48] The number of students on an IEP could also be thought of as an element of S_{dt}, since school policies (or changes therein) could dictate the prevalence of a student with a disability being identified as such.

2.6 Results

Table 5 presents estimates of the model presented in equation (1.4) using the three-year difference from 8th to 11th grade of cohorts eight, nine, ten, and eleven from Figure 4. In column (1), the difference in math pass rates is the dependent variable. The coefficient on a strike occurring sometime during the three-year period is -2.205 and significant at the 0.05 level, suggesting that a strike occurring in the past three years results in a 2.2 percentage-point decrease in the pass rate of a cohort affected by the strike. The coefficient on an expired contract occurring sometime during the three-year period is small and not statistically significant. In column (2), the difference in reading pass rates is the dependent variable. Here, the coefficient on a strike occurring sometime during the three-year period is small and insignificant; similarly, the coefficient on an expired contract is small and insignificant.

Also of interest at the high-school level is a district's graduation rate. Table 6 presents estimates of a district-level fixed-effect model using graduation rate as the dependent variable. In column (1), strike and two lagged strike dummy variables are the independent variables of interest, and percent of students on FRL, percent of students that are white, percent of students on an IEP, pupil-teacher ratio, and year indicators are used as controls. These estimates show that a strike in the current year (when the graduating cohort was in 12th grade) or two years ago (when the graduating cohort was in 10th grade) have a negative but insignificant effect on graduation rate, while a strike the previous year (when the graduating cohort was in 11th grade) has a significant effect, decreasing the graduation rate by about 2.2 percentage points. This implies that there is a lower graduation rate amongst students who experienced a teacher strike during their 11th-grade year, compared to those who did not experience a strike, or did experience a strike, but not during their 11th-grade year. Column (2), where the independent variable of interest is a pooled dummy variable of the three from column (1), confirms that the negative impact on graduation rate is not statistically significant for all grade-levels of high school students experiencing a strike. This regression shows a negative but insignificant effect overall of a strike in the past three years on graduation rate, implying that a strike in the 11th-grade year is particularly important in terms of dropping out.

Table 7 presents estimates of the model presented in equation (1.4) using one-year differences for student cohorts ending in 8th grade through 4th grade.[49] These include all consecutive-year pairs from cohorts 1-8 in Figure 4. Column (1) uses the difference in math pass rates as the dependent variable. Here, the coefficients on lagged strike and lagged expired contract are small and insignificant. The coefficient on strike is negative and significant, implying that a strike decreases the pass rate by 1.1 percentage points. The coefficient on expired contract is also negative and significant, implying that an expired teacher contract actually decreases the math pass rate by 0.4 percentage points. Column (2) uses the difference in reading pass rates as the dependent variable. Here again, the coefficients on lagged strike and lagged expired contract are small and insignificant; the coefficients on strike and expired contract are both negative but also insignificant.

2.7 Conclusions and Policy Implications

The purpose of this study is to explore whether there exists a measurable relationship between labor strife and productivity in public schools, similar to the well-documented, negative relationship between labor strife and production in firms (Kleiner, Leonard & Pilarski 1999; Mas 2007; Krueger & Mas 2003). My contribution is (1) to improve upon previous studies using econometric techniques currently used in the economics of education to control for the large number of cumulative inputs and unobserved endowed ability that factor into the ceteris paribus empirical analysis of education production (Todd and Wolpin 2001); (2) to add to the literature the analysis of non-strike observed labor strife, in this case, periods of time between signed contracts, and its effect on education production; (3) to investigate the impact of teacher labor strife on graduation rates; and (4) to use data benefiting from the increased frequency and participation in annual math and reading testing provided by post-NCLB education policy to perform this analysis.

The predominant model in this paper, used to analyze the impact of labor strife on the percentage of students passing a math or reading exam, is a district-level cohort first-differences model, similar to the school-level cohort first-differences model used by

[49] For example, a cohort's 3rd grade pass rate is subtracted from their 4th grade pass rate, and this is used as a value of the dependent variable.

Baker (2011). Using this model I find that a teacher strike significantly decreases the percentage of students passing their math tests in elementary- and middle- school grades by about 1.1 percentage points and in high-school by about 2.2 percentage points. I also find a smaller negative effect on the percentage of students passing their reading tests, but these are not statistically significant. These results are similar to the findings of Baker (2011) and Johnson (2011), both of whom found negative effects of strikes on Ontario 3rd and 6th grade test scores. One major difference between the data from Pennsylvania, used in this study, and the data from Ontario, in the two previously mentioned studies, is that in Pennsylvania, the school days missed due to a strike must be made up by law, whereas that is not the case in Ontario. Therefore a decrease in Pennsylvania test scores as a result of a strike cannot be attributed simply to fewer instructional days, but rather to something specific to a strike, such as an interruption to the flow of instruction, requiring teachers to backtrack and re-teach some material students forgot during their absence,[50] or a change in teacher or student attitudes before and/or after the strike. While it may seem counter-intuitive that strikes have a larger impact on math scores than on reading scores, Cooper, Nye, Charlton, Lindsay, and Greathouse (1996), in the context of knowledge lost over summer vacation, offer two possible explanations for the difference in magnitude and significance of the effects of a strike on math and reading tests: (1) during a period away from school, students are more easily able to keep up with their reading skills than their math skills; or (2) differences in the susceptibility to memory decay between the types of knowledge used for math and reading. I posit a third possible explanation: at the middle- and high-school levels, math skills are typically taught in one class by one teacher, whereas reading skills are taught in many classes and by many teachers.[51]

These results differ from those of Zwerling (2008), who also uses data from post-NCLB Pennsylvania and concludes that strikes have no effect on test score pass rates. Zwerling's methodological approach is to use either a district cross-sectional analysis with lagged test scores from a previous year but the same grade-level as a control

[50] Studies, such as those in the meta-analysis done by Cooper et al. (1996), find that students lose knowledge equivalent to one month of learning over summer vacation, and as a result teachers spend time re-teaching material at the beginning of each year. If the same principle applies to shorter vacations or unexpected time away from instruction, then even having the missed days made up would not entirely make up for the loss of knowledge occurring during a strike.

[51] For example, students use and learn reading skills in their English classes, but also in their social studies classes.

variable, which raises the question of endogeneity between the lagged independent variable and the error term, or a first-differences approach with both the dependent variable and its lag, again the same grade-level but a previous year, on the left-hand side of the regression. One issue this model does not address, is whether there are meaningful differences between the two different cohorts represented by a dependent variable and its lag in this context, i.e., a district's 8^{th} grade students in 2006 and the same district's 8^{th} grade students in 2007. The cohort model that I adopt does controls for changes in the cohort only as far as they can be measured using annual district-level demographic data, but has the advantage of actually comparing one group of students to the same group of students, outside of those changes, as opposed to one group of students to a different group of students, as in Zwerling's method.

Using this model I also find a significant negative impact of teachers working an entire year without a new contract on elementary- and middle-school student math and reading pass-rates, though these are significant again only for the math pass-rates. Since this is the marginal effect of teachers working without a contract, independent of teachers engaging in a strike, this negative impact on test scores cannot be due to the effect of missed days that are made up later in the year, and suggests that some other change related to the labor disagreement is impacting student test scores. Possible causes are changes in teacher attitudes, such as those observed by Griffin, Tesluk, and Jacobs (1995) or changes in teacher behaviors, such as the "work-to-rule" campaign observed and analyzed by Johnson (2011).

While several studies, cited previously, have analyzed the effect of strikes on student test scores, this is the first to study the impact of expired contracts on student test scores, and the findings suggest that there is an important negative relationship between the two. Table 4, annual data on expired teacher contracts in Pennsylvania, shows that an average of 11.4% of districts each year have gone at least one entire school year without the union and school board agreeing upon a new contract. Not only are these the districts at risk of experiencing a teacher strike and the negative effects thereof, but my findings suggest that just being without a contract for an extended period of time is associated with lower test pass rates. Certainly, this should be seen as an incentive to both teachers and school board members to agree upon a contract sooner rather than later, for the sake of student performance. However, without knowing exactly what it is about an expired contract that leads to lower test scores, it would be unreasonable to take this as evidence that teachers' rights to collectively bargain is the cause of the lower test scores; an expired contract is merely an observable signal that the attitude, behavior, or other change that causes the lower test scores is taking place. Research into the causes of contention in collective bargaining, such as that of Babcock, Want, and Loewenstein (1996) or Griffin, Tesluk, and Jacobs (1995), may be able to guide policy-makers in easing and shortening the collective bargaining process. A big open question for researchers is how collective bargaining itself affects student outcomes; until a natural experiment presents itself or experimental design becomes an option for rigorous econometric research into the topic, surveying and analyzing teacher attitudes and behaviors across multiple dimensions of their job (i.e., pay, benefits, quality of student) as well as across the rights to collectively bargain and strike, may provide some insight as to the magnitude of the underlying causes of decreased student performance during periods of labor strife.

The model I use to analyze the impact of labor strife on the graduation rate uses a fixed-effect model; a cohort-effects model is not possible since students only graduate from high school once. I find that a strike occurring during a cohort's 11^{th}-grade year of high school is correlated with about a 2.2 percentage-point decrease in that cohort's graduation rate. This is the first study to document a relationship between teacher labor disputes and the graduation rate, though since the enactment of NCLB in 2001 there have been a large number of studies on the various factors contributing to the graduation rate[52] as we attempt, as a nation, to simultaneously increase the rigor of our high school curricula and the percentage of students who graduate. This finding suggests that decreasing the number of school strikes will increase graduation rates. If teachers in a district do strike, administrators and teachers at the high-school level should engage in additional interventions aimed at preventing the marginal student, particularly those in the eleventh grade, from dropping out as a result.

The confidential nature of school data, including test score and graduation rate outcomes as well as demographic information, causes much educational research to be

[52] See, for instance, Heckman and LaFontaine (2010).

done at the school- or district-level. A study such as mine could be improved upon with the use of individual student-level data, which unfortunately is not available to me from Pennsylvania over the years of this study. Future studies performed using student-level data could inform us as to the type of student (based on demographic data or past test performance) that is most at risk of performing below expectation or dropping out of school as a result of teacher labor disputes and further guide policy-makers in preventing negative fallout. Additionally, more information on teacher contracts, such as a database including not only the effective and expiration dates of a contract, but also the date on which it was signed by both parties, as well as how the contract has changed from previous versions, would allow for a more nuanced analysis of the effects of collective bargaining on student outcomes. Ideally, such data would provide both the exact length of time that teachers worked without a contract as well as information on increases or decreases in pay, benefits, and time worked. This would require data collection on teacher contracts and negotiations at the state level, which would then be made available to researchers.

2.8 References

Babcock, Linda, Xianghong Want, and George Loewenstein. 1996. Choosing the wrong pond: Social comparisons in negotiations that reflect a self-serving bias. *The Quarterly Journal of Economics*, 111(1): 1-19.

Baker, Michael. 2011. Industrial actions in schools: Strikes and student achievement. NBER Working paper no. 16846. Cambridge, MA: National Bureau of Economic Research.

Carlton, Patrick and Richard Johnson. 1980. Collective bargaining and Virginia school board members: Perceptions and prognoses. *Peabody Journal of Education*, 57(2): 110-18.

Cawley, John, James Heckman, and Edward Vytlacil. 2001. Three observations on wages and measured cognitive ability. *Labour Economics*, 8(4): 419-42.

Cooper, Harris, Barbara Nye, Kelly Charlton, James Lindsay, and Scott Greathouse. 1996. The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66(3): 227-268.

Cronin, John, Michael Dahlin, Deborah Adkins, and G. Gage Kingsbury. 2007. *The Proficiency Illusion.* Thomas B. Fordham Institute. Accessed at http://www.edexcellence.net/publications/theproficiencyillusion.html.

Gamrat, Frank and Jake Haulk. 2006. Teachers strike; Pennsylvania strikes out. *Policy Brief*, 6(49). The Allegheny Institute for Public Policy. Accessed at http://alleghenyinstitute.org/administrator/components/com_policy/uploads/vol6no49.pdf .

Grayson, J. Paul. 1997. Follow-up survey of strike impact. Institute for Social Research, York University.

Griffin, Mark, Paul Tesluk, and Rick Jacobs. 1995. Bargaining cycles and work-related attitudes: Evidence for threat-rigidity effects. *The Academy of Management Journal*, 38(6): 1709-1725.

Harwell, Michael and Brandon LeBeau. 2010. Student eligibility for a free lunch as an SES measure in education research. *Educational Researcher*, 39: 120-131.

Hebdon, Robert and Robert Stern. 1998. Tradeoffs among expressions of industrial conflict: Public sector strike bans and grievance arbitrations. *Industrial and Labor Relations Review*, 51(2): 204-21.

Heckman, James and Paul LaFontaine. 2010. The American high school graduation rate: Trends and levels. *The Review of Economics and Statistics*, 92(2): 244-262.

Johnson, David R. 2011. “Do Strikes and Work-to-Rule Campaigns Change Elementary School Assessment Results?” *Canadian Public Policy*, Vol. 37, No. 4, pp. 479-494.

Kaufman, Dan. 2012. How did Wisconsin become the most politically divisive place in America? *The New York Times Magazine*, May 24, 2012. Accessed at http://www.nytimes.com/2012/05/27/magazine/how-did-wisconsin-become-the-most-politically-divisive-place-in-america.html.

Kleiner, Morris, Johnathan Leonard, and Adam Pilarski. 2002. How industrial relations affects plant performance: The case of commercial aircraft manufacturing.” *Industrial and Labor Relations Review*, 55(2):195-218.

Krueger, Alan and Alexandre Mas. 2004. Strikes, scabs and tread separations: Labor strife and the production of defective Bridgestone/Firestone tires. *Journal of Political Economy*, 112(2): 253-289.

Krueger, Carl. 2002 Unions/collective bargaining. StateNotes, June 2002. Education Commission of the States. Accessed at http://www.ecs.org/html/publications/documents/0203SN_Collection.pdf.

Macdonald, Christine and Catherine Jun. 2006. School strike law faces test. *The Detroit News*, September 7, 2006. Accessed at www.detroitnews.com.

Mas, Alexandre. 2006. Pay, reference points, and police performance. *Quarterly Journal of Economics*, 121(3): 783-821.

Mas, Alexandre. 2008. Labour unrest and the quality of production: Evidence from the construction equipment resale market. *Review of Economic Studies*, 71(1): 229-258.

Michaelis, Marsha. 2003. History of Washington teacher strikes. Evergreen Freedom Foundation. Accessed at www.effwa.org.

Pennsylvania Center for the Study of Labor Relations. 2007. Significant events in Pennsylvania labor history: 1780-2000. Indiana University of Pennsylvania. Accessed at http://www.hhs.iup.edu/laborcenter/.

Pennsylvania Department of Education. 2002. Pennsylvania system of school assessment handbook for report interpretation 2002 PSSA mathematics and reading assessment for grades 5, 8, and 11. Accessed at http://www.pde.state.pa.us.

Pennsylvania Department of Education. 2011. Pennsylvania consolidated state application accountability workbook. Accessed at http://www.education.state.pa.us/portal/server.pt/community/pennsylvania_department_of_education/7237.

Pennsylvania School Boards Association. 1993. *Public School Negotiations: A Complete Guide to Collective Bargaining in Pennsylvania Public Education.* Mechanicsburg: Pennsylvania School Boards Association.

Rivkin, Steven, Eric Hanushek, and John Kain. 2005. Teachers, schools, and academic achievement. *Econometrica*, 73(2): 417-458.

Rockoff, Jonah. 2004. The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2): 247-252.

Rowan, Brian, Fang-Shen Chiang, and Robert Miller. 1997. Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education*, 70(4): 256-84.

Summers, Anita and Barbara Wolfe. 1977. Do schools make a difference? *American Economic Review*, 67(4): 639-52.

Templeton, Tom. 2011. School employee strikes. Pennsylvania School Boards Association. Accessed at https://www.psba.org/issues-advocacy/issues-research/contracts-strikes/school-employee-strikes.asp.

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

Todd, Petra and Kenneth Wolpin. 2007. The production of cognitive achievement in children: Home, school and racial test score gaps. *Journal of Human Capital*, 1(1): 91-136.

Weaver, Elizabeth. 2007. Pennsylvania teachers: Number one in strikes. Report #07-06. Allegheny Institute. Accessed at www.stopteacherstrikes.org/AlleghenyInstitute.pdf.

Zwerling, Harris L. 2008. Pennsylvania teachers' strikes and academic performance. *Journal of Collective Negotiations*, 32(3): 151-172.

2.9 Tables and Figures

Table 2.0.1: District-Level Math PSSA Pass Rate Descriptive Statistics

PSSA Reading	Observations	Mean	Standard Deviation	Minimum	Maximum
3rd Grade	2998	85.3	8.6	35.0	101.0
4th Grade	2497	83.5	9.7	26.2	100.0
5th Grade	3498	73.3	12.3	16.2	100.0
6th Grade	2499	76.3	12.0	13.9	100.0
7th Grade	2499	74.2	12.8	10.5	100.0
8th Grade	3499	69.4	13.9	13.0	98.2
11th Grade	3485	55.4	13.9	2.6	94.4

Figure 2.1: Time-Trend in Math PSSA Pass Rates

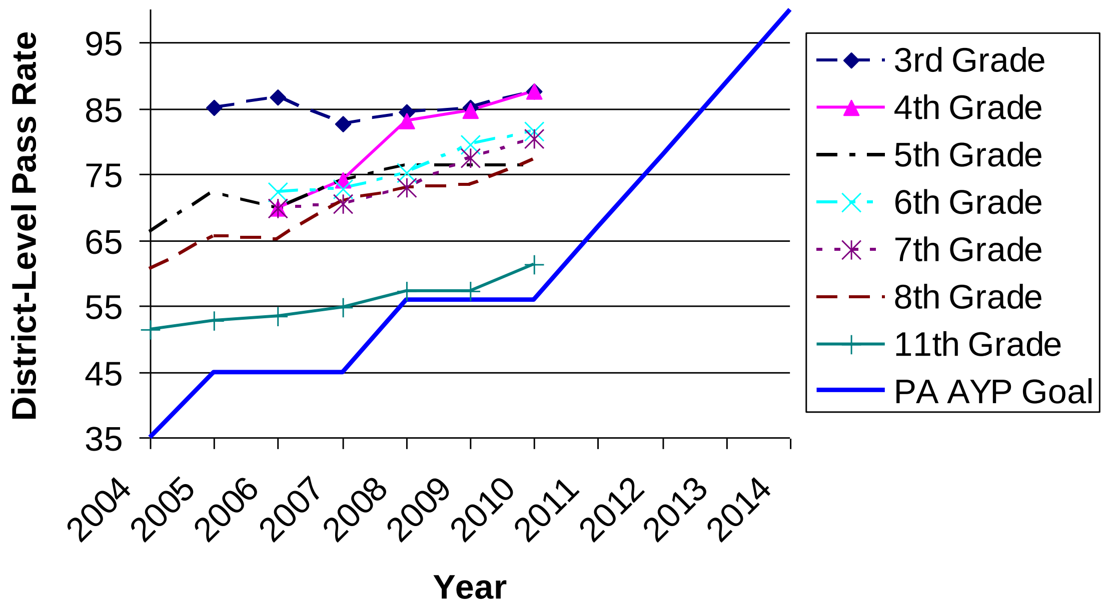


Table 2.0.2: District-Level Reading PSSA Pass Rate Descriptive Statistics

PSSA Reading	Observations	Mean	Standard Deviation	Minimum	Maximum
3rd Grade	2998	76.9	10.6	10.0	100.0
4th Grade	2497	73.9	11.2	16.7	100.0
5th Grade	3498	65.7	12.3	5.4	98.0
6th Grade	2499	69.9	12.0	6.8	97.4
7th Grade	2499	72.2	11.9	14.3	100.0
8th Grade	3499	76.7	12.2	17.9	99.1
11th Grade	3485	67.4	11.9	2.6	96.5

Figure 2.2: Time Trend in Reading PSSA Pass Rates

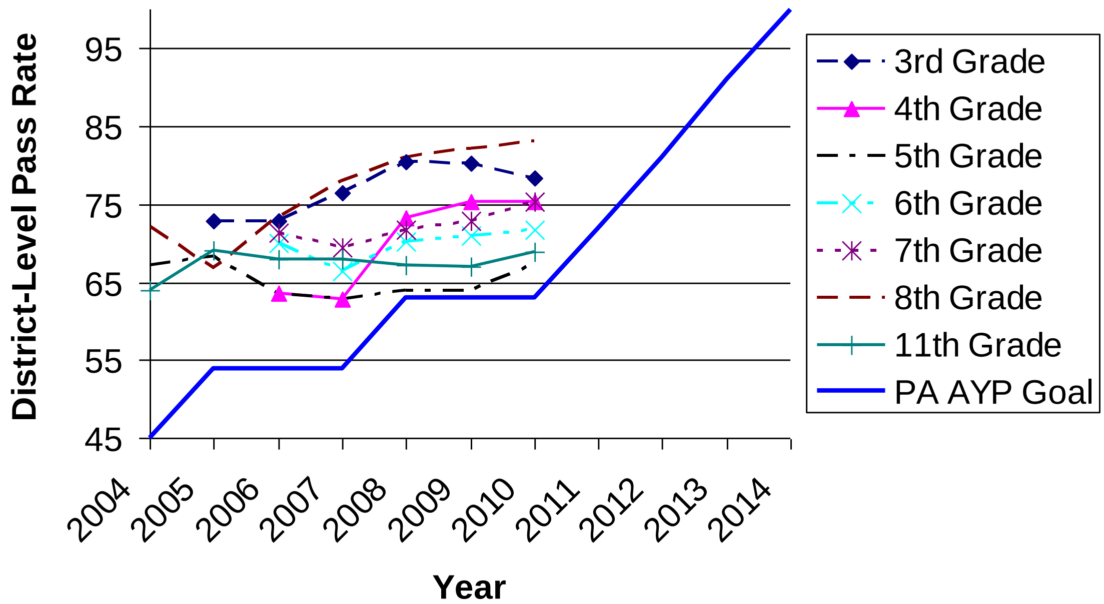


Table 2.0.3: Descriptive Statistics: Striking vs. Non-Striking Districts

	Mean Value for Non-Striking Districts		Mean Value for Striking Districts	
Number of Students	3462	(8360)	3451	(2929)
Percent of Students on Free or Reduced Lunch*	29.6%	(0.166)	25.1%	(0.165)
Percent of Students White	88.3%	(0.170)	89.8%	(0.150)
District Spending Per Student‡	$ 8,330	(1376)	$ 8,253	(1308)
Pupil-Teacher Ratio	14.7	(1.9)	15.2	(1.9)
Per-Capita Income†	$ 19,955	(5975)	$ 21,050	(5130)
Percent of Population College Graduates†	21.7%	(0.133)	24.7%	(0.118)
High School Graduation Rate	86.3%	(0.097)	87.7%	(0.056)
Pass Rate for 3rd Grade Math PSSA	85.3	(8.6)	85.9	(8.5)
Pass Rate for 4th Grade Math PSSA	83.6	(9.7)	82.2	(10.0)
Pass Rate for 5th Grade Math PSSA	73.3	(12.3)	73.6	(12.8)
Pass Rate for 6th Grade Math PSSA	76.3	(12.0)	74.7	(11.5)
Pass Rate for 7th Grade Math PSSA	74.2	(12.7)	72.5	(13.0)
Pass Rate for 8th Grade Math PSSA	69.4	(13.9)	71.0	(13.8)
Pass Rate for 11th Grade Math PSSA	55.4	(13.9)	57.7	(13.5)
Pass Rate for 3rd Grade Reading PSSA	76.9	(10.6)	77.4	(10.9)
Pass Rate for 4th Grade Reading PSSA	73.9	(11.2)	74.1	(12.8)
Pass Rate for 5th Grade Reading PSSA	65.7	(12.3)	66.4	(13.4)
Pass Rate for 6th Grade Reading PSSA	69.9	(12.0)	69.3	(12.8)
Pass Rate for 7th Grade Reading PSSA	72.2	(11.8)	72.1	(14.0)
Pass Rate for 8th Grade Reading PSSA	76.6	(12.2)	79.1	(9.6)
Pass Rate for 11th Grade Reading PSSA*	67.4	(11.9)	70.7	(9.9)

†Data taken from the Census 2000 School District Demographics Project.
‡In 2000 dollars.
*Means for the two groups are statistically different at the 0.05 level.

Figure 2.3: Map of Striking Districts

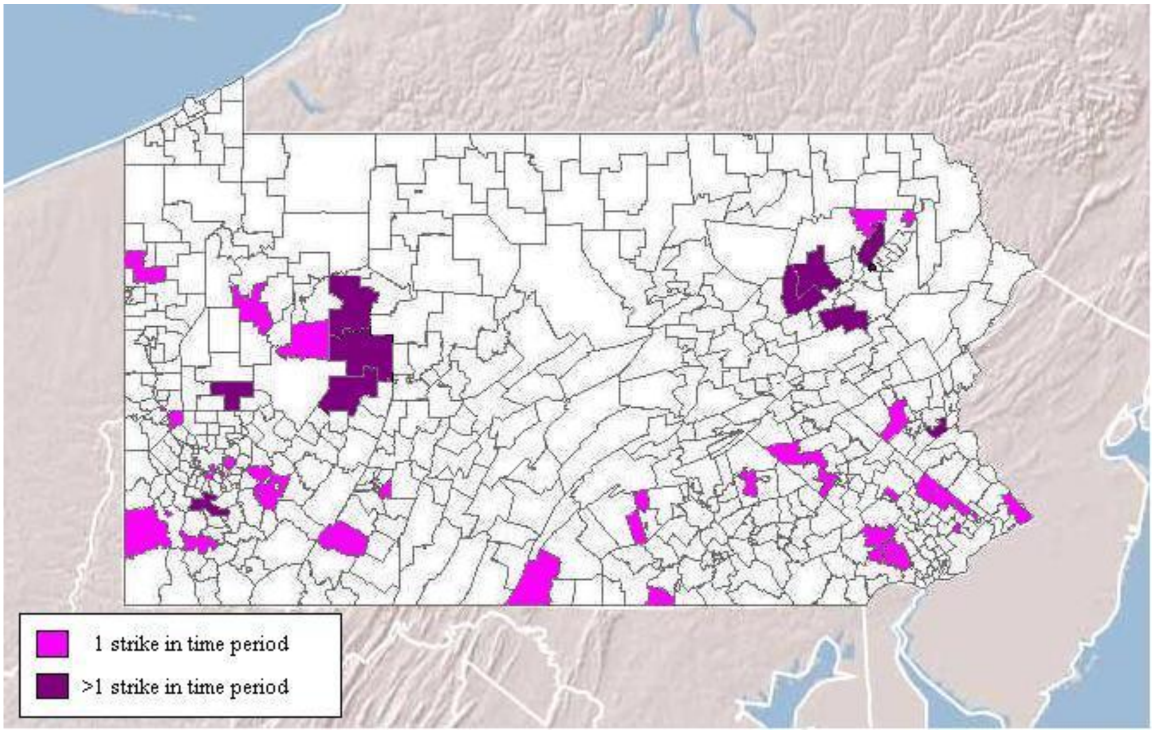


Table 2.0.4: Percent of Districts with Expired Contracts or Striking Each Year

Year	Expired	Striking
2004	8.8%	1.6%
2005	10.0%	1.2%
2006	11.2%	2.4%
2007	15.4%	2.6%
2008	7.8%	0.8%
2009	4.4%	1.4%
2010	13.9%	1.2%
TOTAL	11.2%	1.6%

Figure 2.4: Test Score Data Available, By Cohort

		School Year							Cohort Label
		2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	
Grade Level	3rd grade		x	x	x	x	x	x	↓
	4th grade			x	x	x	x	x	
	5th grade	x	x	x	x	x	x	x	1
	6th grade			x	x	x	x	x	2
	7th grade			x	x	x	x	x	3
	8th grade	x	x	x	x	x	x	x	4
	9th grade	No Tests Given							5
	10th grade								6
	11th grade	x	x	x	x	x	x	x	7
Cohort Label →						11	10	9	8

Table 2.0.5: Cohort Regression Results: 8th-11th Grade‡

	(1) **Difference in 11th and 8th Grade Math Pass Rates**	(2) **Difference in 11th and 8th Grade Reading Pass Rates**
Strike Occurred	-2.205**	-0.156
in 9th, 10th, or 11th Grade	(1.054)	(0.690)
Expired Contract Existed	0.145	-0.00566
in 9th, 10th, or 11th Grade	(0.494)	(0.429)
N	1,970	1,970
R-squared	0.0306	0.1992

‡First-differences regression (by district), robust standard errors in parentheses. The model also included the three-year change in: percent of students on free or reduced lunch, percent of students white, percent of students on an individualized education plan (IEP), and pupil-teacher ratio; dummy variables to control for test year, and a constant.
**Significant at the 0.05 level

Table 2.0.6: Regression Results: Graduation Rate†

	(1)	(2)
	Graduation Rate	
Strike This Year	-0.783	
	(0.986)	
Strike Last Year	-2.217**	
	(1.056)	
Strike Two Years Ago	-0.996	
	(0.994)	
Strike Occurred in Last Three Years		-0.966
		(0.777)
N	1,481	1,481
R-squared	0.0116	0.0093
Number of Districts	497	497

†Fixed-effects regression (by district) with clustered, robust standard errors. Control variables used: constant, percent of students on free or reduced lunch, percent of students white, percent of students on an individualized education plan (IEP), pupil-teacher ratio, year dummy variables.
**Significant at the 0.05 level

Table 2.7: Cohort Regression for One-Year First Differences, 4th-8th Grades[†]

	(1) **Change in Math Pass Rate**	(2) **Change in Reading Pass Rate**
Strike	-1.142**	-0.687
	(0.509)	(0.516)
Strike Last Year	0.0174	-0.623
	(0.620)	(0.455)
Expired Contract	-0.434*	-0.318
	(0.246)	(0.226)
Expired Contract Last Year	-0.0875	0.107
	(0.253)	(0.254)
N	10,869	10,869
R-squared	0.1673	0.4450

[†]First-differences regressions (by district) with clustered, robust standard errors in parentheses. The model also included the one-year change in: percent of students on free or reduced lunch, percent of students white, percent of students on an individualized education plan (IEP), and pupil-teacher ratio; dummy variables to control for test year and grade level, and a constant.
*Significant at the 0.10 level
**Significant at the 0.05 level

2.10 Appendix

Table 2.A.1: First-Stage Instrumental-Variables Regression Results

	(1)	(2)	(3)	(4)
		Strike		
Instrument: Difference in Average Comparable BA Minimum Salary	-1.9E-07 (1.11E-06)			
Instrument: Difference in Average Comparable BA Maximum Salary		3.25E-07 (5.17E-07)		
Instrument: Difference in Average Comparable MA Minimum Salary			1.42E-06 (8.82E-07)	
Instrument: Difference in Average Comparable MA Maximum Salary				9.98E-07 (6.98E-07)
Constant	0.0163*** (0.00273)	0.0149*** (0.00274)	0.0133*** (0.00271)	0.0136*** (0.00269)
Observations	3498	3498	3498	3498
R-squared	0	0	0.001	0.001

***Significant at the 0.01 level

Table 2.A.2: 2009 Descriptive Statistics: Never-Striking vs. Ever-Striking Districts

	Mean Value for Non-Striking Districts		Mean Value for Striking Districts	
Number of Students	3340	(7859)	3726	(3231)
Percent of Students on Free or Reduced Lunch	32.7%	(0.162)	28.8%	(0.180)
Percent of Students White	87.3%	(0.178)	87.2%	(0.168)
District Spending Per Student‡	$ 8,845	(1450)	$ 9,155	(1489)
Pupil-Teacher Ratio	14.2	(2.2)	14.2	(2.2)
Per-Capita Income†	$ 19,869	(6015)	$ 21,042	(5418)
Percent of Population College Graduates†	21.4%	(0.133)	25.0%	(0.124)
High School Graduation Rate	86.2%	(0.095)	88.2%	(0.082)
Pass Rate for 3rd Grade Math PSSA	85.2	(8.4)	86.2	(8.3)
Pass Rate for 4th Grade Math PSSA	84.7	(8.9)	85.0	(10.1)
Pass Rate for 5th Grade Math PSSA	76.3	(11.3)	77.0	(13.1)
Pass Rate for 6th Grade Math PSSA	79.5	(10.0)	79.9	(10.6)
Pass Rate for 7th Grade Math PSSA	77.4	(10.7)	77.6	(12.3)
Pass Rate for 8th Grade Math PSSA	73.3	(11.7)	74.1	(14.5)
Pass Rate for 11th Grade Math PSSA	57.1	(12.6)	58.5	(12.8)
Pass Rate for 3rd Grade Reading PSSA	80.2	(9.6)	81.8	(10.3)
Pass Rate for 4th Grade Reading PSSA	75.3	(10.7)	76.5	(13.0)
Pass Rate for 5th Grade Reading PSSA	67.1	(11.9)	69.4	(12.2)
Pass Rate for 6th Grade Reading PSSA	70.9	(11.4)	72.2	(12.0)
Pass Rate for 7th Grade Reading PSSA	72.8	(11.6)	74.0	(13.9)
Pass Rate for 8th Grade Reading PSSA	82.1	(9.2)	83.9	(9.8)
Pass Rate for 11th Grade Reading PSSA	66.7	(11.8)	69.1	(11.0)

†Data taken from the Census 2000 School District Demographics Project.
‡In 2000 dollars.

This table compares the descriptive statistics from the 2008-09 school year of districts that never strike during the period 2003-04 to 2009-10 to districts that strike at least once during the period. The year 2009 was chosen because it was a year in which all grades were tested and the variables requiring inputs from a subsequent year, district spending per student and high school graduation rate, were available. None of the variables are significantly different between striking and never-striking districts at the 10%-level.

Table 2.A.3: One-Year Cohort Regression Results: Math Grades 4-8†

	(1) **8th Grade Pass Rate Minus 7th Grade Pass Rate**	(2) **7th Grade Pass Rate Minus 6th Grade Pass Rate**	(3) **6th Grade Pass Rate Minus 5th Grade Pass Rate**	(4) **5th Grade Pass Rate Minus 4th Grade Pass Rate**	(5) **4th Grade Pass Rate Minus 3rd Grade Pass Rate**
Strike	0.324	-1.768	-1.933**	-0.746	-1.309
	(1.141)	(1.946)	(0.917)	(1.025)	(0.844)
Strike Last Year	0.0618	-1.166	0.986	-0.207	0.179
	(1.077)	(1.450)	(1.149)	(1.202)	(0.769)
Expired Contract	-0.0952	-0.616	0.127	-0.764	-0.839**
	(0.554)	(0.631)	(0.576)	(0.577)	(0.403)
Expired Contract Last Year	-0.699	-0.0403	0.153	0.668	-0.488
	(0.525)	(0.599)	(0.585)	(0.650)	(0.454)
N	1,975	1,975	2,474	1,973	2,472
R-squared	0.0279	0.0378	0.0501	0.0197	0.1949

†Regressions run as first-differences by district, robust standard errors in parentheses. The model also included the one-year change in: percent of students on free or reduced lunch, percent of students white, percent of students on an individualized education plan (IEP), and pupil-teacher ratio; dummy variables to control for test year, and a constant.
*Significant at the 0.10 level
**Significant at the 0.05 level
***Significant at the 0.01 level

Table 2.A.4: One-Year Cohort Regression Results: Reading Grades 4-8†

	(1) **8th Grade Pass Rate Minus 7th Grade Pass Rate**	(2) **7th Grade Pass Rate Minus 6th Grade Pass Rate**	(3) **6th Grade Pass Rate Minus 5th Grade Pass Rate**	(4) **5th Grade Pass Rate Minus 4th Grade Pass Rate**	(5) **4th Grade Pass Rate Minus 3rd Grade Pass Rate**
Strike	-0.228	-2.427***	-1.300	-0.268	0.543
	(1.081)	(0.796)	(1.201)	(1.008)	(0.635)
Strike Last Year	-0.571	-0.543	-0.0994	-1.641*	-0.371
	(0.946)	(1.024)	(1.106)	(0.878)	(0.788)
Expired Contract	-0.262	0.261	0.0713	-1.321***	-0.384
	(0.508)	(0.526)	(0.502)	(0.492)	(0.441)
Expired Contract Last Year	0.659	-0.413	0.431	0.641	-0.670
	(0.521)	(0.498)	(0.557)	(0.609)	(0.537)
N	1,975	1,975	2,474	1,973	2,472
R-squared	0.0979	0.1029	0.0996	0.0476	0.1138

†Regressions run as first-differences by district, robust standard errors in parentheses. The model also included the one-year change in: percent of students on free or reduced lunch, percent of students white, percent of students on an individualized education plan (IEP), and pupil-teacher ratio; dummy variables to control for test year, and a constant.

*Significant at the 0.10 level

**Significant at the 0.05 level

***Significant at the 0.01 level

Chapter 3: Selection Bias and For-Profit Associate Degrees: Evidence from the NLSY Geocode

Abstract

The past decade has seen enormous growth in the for-profit higher education industry, and along with it, enormous debate over the relative costs and benefits of such an education. Utilizing the rich data from the NLSY97 Geocode merged with institutional data from IPEDS, I empirically analyze data on individuals with two-year degrees, estimate the average marginal earnings gain from a two-year degree, and compare the effects of degrees across institutional sector and across major area of study using OLS with family background and extensive demographic controls. I find evidence of selection at three levels: selection into college, selection into type of college, and selection into major area of study. Any estimates of labor market returns to these degrees will be biased until future research unravels and models these selection mechanisms and processes. This chapter provides a first look into the differential inputs and outputs of for-profit and public two-year degree programs. I find that a two-year degree results in an 8.1 percent average marginal earnings gain over a high-school diploma, and that the sector of the degree-granting institution alone does affect this gain. I also find that earnings gains vary greatly by major; individuals with "academic" degrees experience no significant earnings gains while individuals with "vocational/technical" degrees on average experience a 32.7 percent earnings gain. I find statistical differences in the marginal earnings gains across institutional sector within major fields of study, suggesting that attending a for-profit does matter when major field of study is taken into account. Policy-makers should take note that this preliminary analysis of the returns (which can be thought of as the private benefits) to public and for-profit degrees does not provide unambiguous evidence in favor of one sector over another, but rather a first look into the "black box" in which students, institutions, and major areas of study come together and jointly determine labor market outcomes.

This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. The views expressed here do not necessarily reflect the views of the BLS.

3.1 Motivation

For-profit education is a hot topic in the world of higher education. News coverage frequently posits that for-profit schools take advantage of their students via misleading facts and predatory marketing[53] and the sector has been the subject of a recent negative U.S. Senate committee report.[54] Despite this negative publicity, for-profit colleges, often through online degree programs, offer the promise of higher-educational services to a population that has been limited in the past by their geography and the physical location of competitors, such as community colleges, by the hours they work, or their family obligations at home. Amidst this political controversy, the number of students pursuing education from for-profit institutions has risen rapidly in recent years,[55] and scholarly research in the field of for-profit education lags behind public opinion.

Traditionally, the for-profit sector was made up of proprietary schools,[56] but over the past 40 years they have seen unprecedented growth, both in terms of size and market share. Deming, Goldin, and Katz (2012) describe this recent growth, wherein fall enrollments at for-profit schools have grown over 100-fold and their percent of enrollment among degree-granting schools has grown from less than 1 percent to 9 percent from 1970 to 2009. At the same time, for-profit schools have expanded their market, which used to offer only more traditional certificate programs, to actual two- and four-year degree programs. This has come with a simultaneous growth in federal dollars received by for-profit colleges and universities. The University of Phoenix, one of the largest "chain" for-profits, received \$933 million in Pell Grants in 2009, "making it the biggest recipient of federal dollars for financial aid."[57] The rising amount of federal dollars going to the for-profit sector has caused both congressional and presidential investigation into the performance of for-profits in preparing their students for "gainful employment," and has resulted in a change of policy regarding the allocation of Title IV

[53] See Lewin (2010) or Greenblatt (2012), for example.
[54] See Lewin (2012).
[55] According to the Carnegie Foundation for the Advancement of Teaching, 77% of new higher education institutions arising between 2005 and 2010 are private, for-profit institutions.
[56] The for-profit sector is not new: about a quarter of the regionally accredited for-profits are more than 100 years old. (Kinser 2005)
[57] Kreighbaum (2010).

funds to schools that fail to prepare their students to successfully pay back their federal student loans.

An answer to the question of how for-profit institutions affect individuals' earnings potentials, *ceterus paribus*, would be of great import to policy-makers determining how to allocate federal Title IV funds across the many types of institutions competing to educate individuals who qualify for this money. It would also be useful to prospective students who, according to neoclassical economic theory, engage in a set of rational decisions regarding their schooling choices in order to maximize their stream of net future earnings. Unfortunately, selection bias in educational attainment is a well-documented phenomenon (Card 1999) and as I will show in this chapter, the selection bias into for-profit colleges occurs on three distinct levels: selection of high-school graduates or GED-holders into college; selection of college attendees into different types of colleges;[58] and selection into a major area of study. Unraveling the underlying selection processes, which may include a combination of preference-based self-selection or ability-based institutional selection, is not a goal of this chapter but remains an open question for further research. Instead, this chapter empirically describes a nationally representative longitudinal sample of two-year degree completers and dropouts to their high school (or equivalent) educated counterparts, both before and after they pursue their educations.

My descriptive analysis finds that two-year degree completers come from households and families with significantly higher socio-economic status (SES) and have significantly higher (proxy) measures of innate ability than other high-school educated individuals, both before and after eliminating individuals who go on to pursue further education (beyond a two-year degree) from the sample. Among two-year degree completers, those with degrees from public colleges do not differ significantly from those with degrees from private, not-for-profit colleges, but come from households and families with significantly higher SES and have higher measures of ability than those with degrees from for-profit colleges. They are also about 20 percent more likely to pursue a bachelor degree than for-profit two-year degree-holders. Furthermore, I find some evidence of

[58] In this chapter I will distinguish between three different types of colleges: public, for-profit, and private, not-for-profit colleges.

selection within major area of study across types of colleges. Among individuals pursuing degrees in a health-related field, those entering for-profit colleges have significantly lower proxy measures of endowed ability than those entering public colleges. Not surprisingly, earnings outcomes appear to be significantly lower for health-related for-profit graduates than health-related public graduates, both descriptively and after regression analysis controlling for observed demographics, family background, and ability proxies. In addition, for-profit graduates of health degree programs are much less likely than their public analogues to be working in a health-related field after completing their degree. With strong evidence of selection bias in favor of the public colleges, it is not reasonable to conclude that for-profit health-related degrees are causing these graduates to have lower earnings, though it is a possibility; unambiguously, the individuals with for-profit health degrees earn less than their public health-degree holding counterparts.

Despite the evidence of tri-fold selection into colleges and majors wherein individuals obtaining for-profit degrees are relatively disadvantaged compared to those with public degrees, descriptive statistics and controlling for observed covariates show that for-profit degree-holders do not always have worse outcomes than their public counterparts. I find business and vocation/technical majors from both for-profit and public colleges earn significantly more than their high-school educated counterparts. Neither descriptive statistics nor regression analysis finds evidence that either public or for-profit graduates in these fields earn significantly more than the other. I find that for-profit graduates in business or vocational/technical fields are more likely than their public counterparts to be employed full-time after obtaining a degree, and business majors from for-profits are less likely to report not working once they hold their degrees.

I find evidence of differences between the labor market outcomes of public and for-profit degree-holders using a nationally representative sample of young adults, but the evidence does not indicate that a public two-year degree is unilaterally correlated with higher earnings than a for-profit two-year degree, or *vice versa*. Given strong evidence of selection into types of colleges and majors, none of my findings should be interpreted as causal effects. However, the selection I find implies a negative bias on the earnings estimates of for-profit degree-holders and among some major fields of study, these individuals are still performing as well or better than their publicly-educated counterparts. A major finding of this chapter is that the field of study is crucial to determining differential labor market outcomes across types of colleges. I recommend that future research focus on the mechanisms of selection into types of colleges and major fields of study.

3.2 Conceptual Framework

Since the rapid growth of and controversy over the for-profit sector is a recent phenomenon, few papers describe the demographics and earnings of individuals with degrees from these institutions, but those that do are recent or concurrent to this chapter. Deming, Goldin, and Katz (2012) describe the for-profit sector in great detail using data from the Beginning Postsecondary Students (BPS) longitudinal survey, which includes information on the schools, students, programs, and relationship to the federal government. They find that for-profits educate a greater percentage of minorities, disadvantaged, and older students than other types of institutions, but that these students frequently end up unemployed, with lower earnings and higher debt-burdens than do similar students at other schools. A concurrent working paper, Lang and Weinstein (2013), also uses the BPS, using ordinary least squares (OLS) and matching models to evaluate the returns to certificates, associate degrees, and bachelor degrees at for-profit and not-for-profit institutions, controlling for major field of study. They find that the differences in returns across majors are greater than the differences in returns across institutional sectors. The BPS has the benefit of being expressly collected to investigate first-time students of post-secondary education programs, with the most recent cohort containing information on over 16,700 students at three points in time. One drawback of using this data to study for-profit students is that it is not nationally representative; in particular, it does not include information on students who are returning to post-secondary education after a gap in education. Deming et al report that 65.1 percent of for-profit students are over age 24 compared to only 40.4 percent of students at two-year public colleges;[59] this indicates that for-profit students are less likely than public students to be enrolled in post-secondary education for the first time.

[59] Deming, Goldin, and Katz, p.147.

The data for this paper comes from the National Longitudinal Surveys of Youth 1997 (NLSY97) Geocode, which consists of a nationally representative sample of approximately 9,000 youths. While the smaller sample size is a disadvantage compared to the BPS, in the NLSY I am able to observe individuals who enter college, leave, and return again, some of whom complete degrees at for-profit institutions. Another concurrent working paper, Cellini and Chaudhary (2012), also uses the NLSY97 Geocode, evaluating the returns to associate's degrees and unfinished work towards associate's degrees. Using an individual fixed-effects approach, the authors find some evidence that students with degrees from for-profit institutions experience higher returns than public sector graduates and that degree completion is an important determinant of for-profit quality and student success.

While Cellini and Chaudhary (2012) attempt to identify the causal effects of education on earnings using individual fixed-effects, following the lead of Jacobson, LaLond, and Sullivan (2005) and Jepsen, Troske, and Coomes (2011), the NLSY97 data is less suited to this methodology than the data used in the other two studies. The individuals in the Jacobson et al paper are displaced workers who attend community college later in life, finishing their schooling at a mean age of 35.9 (Jacobson et al, p.276), and the individuals in the Jepsen et al paper enter community college between the ages of 20 and 60 (Jepsen et al, p.8). Both studies identify the effects of education on earnings, at least in part, using the fixed effects of individuals with a substantial work history prior to enrollment. Respondents in the NLSY97 data, on the other hand, are between the ages of 25 and 29 in 2009, the final year of Cellini and Chaudhary's (2012) analysis. As such, even if an individual worked for several years before attending college, his or her wage history is minimal, associated with a young age, and has not necessarily been established well enough to provide a baseline off of which to identify an earnings fixed-effect.

The NLSY is notorious for its wealth of data, so while it does not yet[60] provide data on individuals who attend for-profit, two-year degree programs after establishing a lengthy earnings history, it lends itself well to OLS analysis with family background and

[60] Simply because the sample is not yet old enough.

other demographic control variables, which I will use in this paper.[61] This model extends Mincer's (1974) human capital earnings function to represent an individual *i*''s log earnings (y) in time t as a linear additive function of education (S), experience (A), and other demographic information (X) as in equation (1).

$$(1) \qquad \log y_{it} = a + bS_{it} + c \cdot f(A_{it}) + d \cdot g(X_{it}) + e_{it}$$

When the elements of S_{it} are indicator variables as they will be in this chapter (discussed in detail in section 3.4), its coefficient is essentially measuring a treatment effect. In this case, the "treatment" is a two-year college degree. The estimate of this treatment effect will be an unbiased estimate of the population treatment effect under certain conditions including the absence of selection into treatment. As discussed in the previous section, there is tri-fold selection into colleges and major fields of studies, so it will be unreasonable to interpret my estimates as causal effects.

Cellini and Chaudhary (2012) focus on individuals who work towards or complete two-year, or associate, degrees, noting that 18 percent of all associate degrees are granted by for-profit colleges, putting them in direct competition with community colleges[62] who, along with four-year public institutions, confer 76 percent of all associate degrees. In contrast, for-profits only confer 5 percent of all bachelor degrees (Deming, Golding, and Katz 2012). In this chapter I focus on individuals who complete two-year degrees. This decision is necessitated by the data; there are simply not enough individuals in the NLSY97 who report earning certificates or bachelor degrees from for-profit institutions to perform a meaningful analysis of their characteristics. Additionally, NLSY97 respondents are still relatively young (30 or younger as of 2010), and two-year degree earners will have a more established work history from which to draw earnings data than would four-year degree holders.

This chapter, like Cellini and Chaudhary (2012) and Lang and Weinstein (2013), is primarily interested in the benefits to a for-profit education. Policy-makers must also

[61] Such a model, including additional assumptions and the probability limit of the estimator, is discussed in detail in Card (1999).
[62] Cellini (2009) looks at competition between public and private two-year colleges by assessing the impact of increased funding for public colleges on the market for two-year college degrees. She finds that increasing funding for public schools diverts private students into the public sector and causes some proprietary schools to exit the market. Bailey, Badway, and Gumport (2003) find that the for-profits are more of a complement than a substitute to community colleges.

take into account the costs of for-profit education, which are discussed in several recent papers. Cellini (2010) examines the relationship between federal, military, and state financial aid grants and the entry of for-profit colleges, finding that an increase in the dollar amount of aid encourages for-profit entry into a market, though not at the expense of the public sector. Cellini (2012) estimates the taxpayer costs of for-profit education and public education, determining that the taxpayer costs are greater when providing a student with a public education than with a for-profit education, although the costs to the student are higher when educated by a for-profit institution. Cellini and Goldin (2012) find evidence in a multi-state study looking at both Title IV eligible and ineligible schools that the Title IV eligible schools raise tuition in order to maximize profit on the available federal student aid.

3.3 Data

The NLSY97 Geocode consists of a nationally representative sample of approximately 9,000 youths who were 12 to 16 years old as of December 31, 1996 and have been interviewed on an annual basis since 1997 on topics including labor market behavior and educational experiences as well as extensive demographic and family background information. In this chapter, I analyze data from rounds 1-14, corresponding to the years 1997-2010. To this I merge data on the sectors of institutions of higher education at which these individuals enrolled from the Integrated Postsecondary Education Data System (IPEDS). The NLSY97 and IPEDS data are publically available; the Geocode variables are restricted-access data.

Since I am interested in the marginal returns to a two-year degree, I limit the sample to individuals who have earned at least a high school diploma or GED by 2010, as these provide an appropriate comparison group for two-year degree earners. Additionally, I eliminate individuals whose reported average hourly wage is greater than $1000, as these individuals are either outliers or the data has been miscoded.[63] Descriptive statistics for the remaining 7,665 individuals in the sample are displayed in Table 3.1. The first column represents the entire sample, and descriptive statistics detail individual characteristics (sex, race, ethnicity, and whether or not the individual was born

[63] This results in the elimination of 45 individual/year observations.

in a foreign country, whether or not English was the individual's second language (ESL), a percentile score for the Armed Services Vocational Aptitude Battery (ASVAB)[64]) , family background characteristics (parents' household net worth in 1997[65], biological mother's and father's highest grade completed, and the age of the individual's mother at the birth of her first child and at the birth of the individual), and schooling-related characteristics (number of years spent as a part-time student and as a full-time student and whether an individual earned a GED, certificate, two-year degree, or bachelor's degree). Of this sample, 8.2 percent, or 629 individuals, earn two-year degrees by 2010. Descriptive statistics for this sample are displayed in the column (2) of Table 3.1. T-tests demonstrate that the sample of two-year degree earners differs significantly from the sample of non-two-year degree earners in sex (significantly less males earn two-year degrees), race (significantly less black and more "other race" individuals earn two-year degrees), ASVAB percentile (higher for two-year degree earners), and mothers' characteristics (mothers of two-year degree earners have more education and were older at first birth and at the birth of the individual). In addition, significantly fewer two-year degree earners had earned a GED than non-two-year degree earners, and (unsurprisingly) two-year degree earners have spent more time as students than non-two-year degree earners.

For most of the analysis, I have further limited the sample to individuals who never earn a bachelor or higher degree. Descriptive statistics for these 6,106 remaining individuals are displayed in the third column of Table 3.1, and statistics for 505 two-year degree earners without bachelor or higher degrees are in the fourth column. T-tests demonstrate additional statistical differences in the means of these two groups than the two groups which included bachelor and higher degree earners. With the exception of mother's highest grade completed and other race, all of the differences from the larger sample remain statistically different in the second sample; in addition, two-year-degree earners are more white, come from households with higher net worth, and have fathers with more education than non-two-year degree earners when individuals with bachelor or higher degrees are eliminated from the sample. Overall, two-year degree earners appear

[64] All respondents were given the opportunity to take the ASVAB.
[65] These have been adjusted to 2010 dollars using the CPI-U for ease of interpretation.

to differ from non-college degree earners in expected ways: they come from families with higher socio-economic status (SES), they are less likely to have earned a GED as opposed to a high-school diploma, and they are more likely to be women.[66]

Data from the NLSY97 Geocode are merged by UnitID[67] with institutional data from the Integrated Postsecondary Education Data System (IPEDS) by academic year of degree completion. This data provides the sector of each institution and allows me to categorize each student's degree as having been earned from a public, for-profit, or private, not-for-profit institution. Most individuals earn only one two-year degree; for individuals who earn more than one degree, I define them based on the final degree that they earn.

Descriptive statistics by type of institution attended are shown in Table 3.2. Columns (1) through (3) include individuals who go on to earn a bachelor or higher degree; in fact, the most notable statistical difference between public and for-profit students is their propensity to earn a bachelor degree. 23.7 percent of public degree holders go on to earn a bachelor degree, as opposed to only 3.9 percent of for-profit degree holders.[68] The remaining statistical differences between public and for-profit students suggest that public students are of slightly higher SES than for-profit students; they are more white, perform higher on the ASVAB, their mothers have completed more education and were older at the individual's birth, and they have spent more time as a student, both part-time and full-time, over the entire time-period.[69] Column (3) presents descriptive statistics on private, not-for-profit students. While there are few of these students (twenty-five), they do not differ statistically from the public students in any meaningful way.[70] Columns (4) through (6) of Table 3.2 eliminate students who go on to earn a bachelor or higher degree. Here the statistical differences between public and for-

[66] A recent report by the NCES, *Higher Education: Gaps in Access and Persistence Study*, examines the growing male-female higher-education gap.
[67] The UnitID is a unique identification number assigned to a postsecondary educational institution by IPEDS; the NLSY 1997 Geocode includes UnitID for all colleges an individual attended.
[68] Cellini (2009) discusses the role of community colleges in encouraging future enrollment in four-year institutions through transferability of coursework; Deming, Goldin, and Katz (2012) find that comparable students enrolled at for-profit schools are less likely to obtain a bachelor degree than public students.
[69] Deming, Goldin, and Katz (2012) find, similarly, that for-profits educate a larger fraction of minority and disadvantaged students.
[70] While they do differ statistically in percentage of Asian students, students with a GED, and students with a certification, this is possibly because of the small sample size and the lack of observations in each of these categories.

profit students remain the same as in columns (1) and (2), with the exception of mother's age at birth, which is no longer statistically different. Surprisingly, the differences in time spent as a student remain significant, even when the bachelor students, who were overwhelmingly public students, are eliminated. This could be caused by public students taking longer to earn their degrees than for-profit students; alternately, given the higher propensity of public students to pursue higher degrees, some of these students may have pursued but never earned bachelor degrees.

Column (1) of Table 3.3 displays descriptive statistics on labor market outcomes for the entire sample. Here and for the regression analysis of labor market outcomes, I have limited the sample to observations in which the individual is at least age twenty-five; most individuals have completed their education and begun working full time by this age. These variables include average hours worked per week and average hourly wage,[71] which I have computed using the average over the first five jobs worked in the year.[72] Multiplying these together generates average weekly earnings, the main outcome variable of interest. Full-time employment is defined as working an average of thirty-five or more hours per week during the year; any employment is defined as working an average more than zero.[73]

The final two variables, number of biological children living in the individual's household and outside the individual's household, are used as independent variables for the regression analysis. Column (2) displays the same statistics for two-year degree holders. The two-year degree holders differ significantly from non-two-year degree holders in average hours worked per week (they work slightly less), any employment (slightly more likely to be employed), and number of biological children both in and outside the household (less likely to have either). Note that their hourly wage and weekly earnings are not significantly different from those of non-two-year degree holders. Columns (3) and (4) eliminate holders of bachelor and higher degrees; the statistical differences between degree holders and non-degree holders remain the same.

[71] Average hourly wage has been adjusted to 2010 dollars using the CPI-U.
[72] This mirrors the data methodology of Cellini and Chaudhary (2012).
[73] Again, these calculations mirror those of Cellini and Chaudhary (2012), and are computed using the average over the first five jobs worked in the year.

Table 3.4 presents labor market outcome descriptive statistics for two-year degree holders by institution type. In the sample including bachelor and higher degree holders (columns (1) through (3)), the only statistically significant difference between public and for-profit students is that for-profit students are about 10 percent more likely to be employed full-time post-education than public students. Once bachelor and higher degree holders are removed from the sample (columns (4) through (6)), for-profit degree holders work about two additional hours per week compared to public degree holders, and are about 14 percent more likely to be employed full-time and 6 percent more likely to have any employment than public degree holders. Despite more hours and a higher likelihood of employment, there is no statistically significant difference between the earnings or wages of these groups. Also as in Table 3.2, private, not-for-profit students do not to differ significantly from public students.

Data on college major by type of institution is presented in Table 3.5. One key difference between public and for-profit students appears to be the number of students with majors in the academic/other/unknown category.[74] Here I have broken the health-related majors into two categories: nursing and non-nursing. Public two-year health-related degrees are much more likely than for-profit degrees to be in nursing; there are very few observations of for-profit nursing degrees.[75] Table 3.6 presents selected descriptive statistics by major field of study and institutional status. The first three columns include individuals who go on to earn bachelor and higher degrees. There are statistical differences in the types of students pursuing health-related degrees in public and for-profit institutions; for-profit students have an average ASVAB percentile about 15 percentage-points below their public counterparts, and their post-degree weekly earnings are lower by about $300. These differences hold up in the final three columns, after individuals who go on to earn a bachelor's or higher degree are removed from the sample. No statistically significant differences exist between public and for-profit degree earners for these variables in any of the three remaining major areas of study.

[74] Breneman, Pusser, and Turner (2006) find that for-profit institutions award a disproportionate share of degrees in pre-professional, technical, and vocational fields as a result of their expansion into areas with a low supply of college options and a high demand for skilled workers.
[75] For the rest of the analysis, I combine the nursing and non-nursing majors into one category. I have also performed the analysis with nursing majors separated from non-nursing majors, which does not change the results. These are available from the author upon request.

Table 3.9 reports the industry of employment for two-year degree holders, by major and institution type, at age twenty-six. The first row displays the percentage of individuals for whom there was no response to this survey question or the response was marked as uncodable; these observations are dropped and the percentages in the remaining rows for each column sum to 100 percent. Among the individuals with health-related majors, 25 percent of for-profit degree-holders were not employed at age 26, compared to only 5 percent of public degree-holders. Additionally, 68.3 percent of public degree-holders report working in the educational, health, and social services industry, compared to just 30 percent of their for-profit counterparts. Among those with business-related majors, 16.7 percent of public degree-holders were not working at age 26, compared to zero percent of for-profit degree-holders, suggesting that a for-profit business degree makes one more employable than a public business degree. The other difference of note is in the services/public administration industry, where 37.5 percent of for-profit business degree-holders were employed, compared to 19.4 percent of public business degree-holders. Finally, among vocational/technical degree-holders, 14.8 percent of for-profit degree-holders reported working in information and communications, compared to only 5.7 percent of public degree-holders; a finer analysis of the data shows that the highest earning vocational/technical degree-holders have studies and work in the field of information technology, and that these individuals hold degrees from for-profit colleges.

3.4 Empirical Methodology

For the regression analysis of labor market outcomes, I approach the model from equation (1) considering the subset of the population described in columns (3) and (4) of Table 3.1: individuals with at least a high school diploma (or GED) who have not earned a bachelor or higher degree. Additionally, I limit the sample to observations in which these individuals are age twenty-five or older, in order to allow the entire sample to complete school and to eliminate the early years in which individuals are less established in their careers and earnings trajectory. In particular, I am interested in the average marginal effect of a two-year degree (compared to a high school diploma) on earnings, whether the average marginal effect from a two-year degree earned from a for-profit institution differs from a public or private, not-for-profit degree, and how major area of study affects the average marginal benefit of a two-year degree, given the degree-granting institution's profit status. To investigate these questions, I will estimate the equation:

(2) $\log w_{it} = \alpha + \beta I_i + \gamma X_{it} + a_{it} + \varepsilon_{it}$.

where w_{it} is an individual's average weekly earnings in year t, X_{it} is a vector of in individual's demographic characteristics (sex, race, ethnicity, ASVAB percentile, ESL, foreign-born, mother's age at first birth, mother's age at birth, the highest grade completed by biological mother and father, the geographic region of residence, the number of biological children in and out of the household and an interaction between sex and number of biological children inside the household[76]), and a_{it} is a set of age-specific indicator variables. I_i is a vector of indicator variables for GED and non-degree certifications plus either: (a) any two-year degree; (b) two-year degree by institution type (public, for-profit, or private, not-for-profit); (c) two-year degree by major type (academic/other, business-related, health-related, or vocational/technical); or (d) two-year degree by institution type[77] and major type (i.e., public health-related degree, for-profit vocational/technical degree). A limitation of the log-earnings model is that non-earners (the willingly or unwillingly unemployed) cannot be included in the estimation; as a result, all coefficients on educational degrees arising from this estimating equation must be interpreted as average marginal returns conditional upon employment.

3.5 Results

The OLS results are presented in Table 3.7. Column (1) presents coefficients on GED and a two-year degree in the absence of demographic and family background covariates.[78] In this specification, holding a GED is associated with 12.9 percent lower weekly earnings than holding only a high school diploma and holding a two-year degree

[76] The sample contains no women with biological children living outside the household.
[77] Here I will eliminate private, not-for-profit degree earners from the sample, as their numbers are small, and their numbers within each major category too small upon which to estimate meaningful average marginal effects.
[78] All specifications include a set of age-specific dummy variables. An indicator variable for a non-degree certificate is also included in all specifications; these coefficients are not reported because they were small and not statistically significant, possibly due to the small percentage of individuals who reported holding such a certificate. Lang and Weinstein (2013) also analyze returns to a certificate and find small, positive point estimates that are not statistically significant.

is associated with a 6.9 percent increase in weekly earnings compared to holding a high school diploma, though the second coefficient is not statistically significant. Column (2) presents the same coefficients from the full model (including all demographic and family background covariates). The demographic and family background characteristics have accounted for much of the difference between GED and high school diploma holders; in the remaining specifications (which all include the full model), a GED is associated with 5.5 to 5.9 percent lower weekly earnings but this effect is no longer statistically significant.[79] In the full model, a two-year degree is associated with an 8.1 percent bump in earnings and is statistically significant; this is comparable to the findings of others in the literature (Kane and Rouse 1995, Jacobson LaLonde, and Sullivan 2004, Cellini and Chaudhary 2012). Columns (3) through (5) also present coefficients from the full model. In column (3), separate variables indicate a two-year degree from a public institution (associated with a 7.9 percent increase in earnings), a for-profit institution (associated with an 8.4 percent increase in earnings), or a private, not-for-profit institution (associated with an 8.7 percent increase in earnings). For this model, I tested the hypothesis that earnings from public, for-profit, and private, not-for-profit institutions were equal and was unable to reject the null at the ten-percent level. This finding implies that the type of degree-granting institution alone is not an important factor in determining returns to a two-year degree and differs from the findings of Cellini and Chaudhary (2012) and Lang and Weinstein (2013), both of whom find some evidence that for-profit graduates experience higher returns than their public-sector counterparts.[80] In column (4), separate variables indicate a two-year degree in one of four fields: academic/other/unknown (associated with a 6.7 percent decrease in earnings), business (associated with a 20 percent increase in earnings), health-related (associated with a 13.5 percent increase in earnings), or vocational/technical (associated with a 32.7 percent increase in earnings). On their own, only the coefficients on business and vocational/technical degrees are statistically significant, the hypothesis that earnings are

[79] Previous literature, such as Cameron and Heckman (1993), finds the earnings of GED-holders to be equal to those of high-school dropouts.

[80] Lang and Weinstein (2013) do not exclude individuals who go on to earn bachelor or higher degrees from their analysis of returns to two-year degrees, and seem to argue that not-for-profit two-year degree holders are earning less on average than their for-profit counterparts because they are in school continuing to pursue a bachelor degree.

equal for the four major fields of study is rejected at the one percent level. Previous literature has also found large, statistically significant differences in returns to different fields of study (Black, Sanders, and Taylor 2003, Jacobson, LaLonde, and Sullivan 2004, Jepsen, Troske, and Coomes 2012, Lang and Weinstein 2013).

In column (5) I have dropped private, not-for-profit degree-earners from the analysis and calculated separate coefficients for each major/institution-type pair. In this specification, academic/other/unknown degrees have negative average marginal earnings (4.3 percent less than diploma-holders for public degree-holders and 26.4 percent less for for-profit degree-holders). The hypothesis that these two coefficients are equal cannot be rejected at the ten-percent level. Business-related degrees have positive, statistically significant average marginal earnings (16.5 percent more than diploma-holders for public degree-holders and 24.3 percent higher for for-profit degree-holders), but the difference between returns to public and for-profit is not statistically significant. Health-related degrees have statistically different average marginal earnings by institutional status; public health degree-holders earn an average of 28.2 percent more than high school diploma holders, while for-profit health degree-holders earn an average of 4.5 percent less than high-school diploma holders. These coefficients are statistically different at the one-percent level.[81] Vocational/technical degree holders have positive, statistically significant average marginal earnings (17.7 percent for public degree-holders and 41.9 percent for for-profit degree-holders), but again the coefficients are not statistically different from each other. Overall, I tested the hypothesis:

$$H_o: \begin{array}{c} \beta_{public\ academic} = \beta_{for-profit\ academic} \\ \beta_{public\ business} = \beta_{for-profit\ business} \\ \beta_{public\ health} = \beta_{for-profit\ health} \\ \beta_{public\frac{vo}{tech}} = \beta_{for-profit\frac{vo}{tech}} \end{array}$$

$$H_a: \begin{array}{c} \beta_{public\ academic} \neq \beta_{for-profit\ academic} \\ \beta_{public\ business} \neq \beta_{for-profit\ business} \\ \beta_{public\ health} \neq \beta_{for-profit\ health} \\ \beta_{public\frac{vo}{tech}} \neq \beta_{for-profit\frac{vo}{tech}} \end{array}$$

and rejected the null hypothesis at the five percent level, suggesting that the major/institutional status pairing has overall statistically significant explanatory power.

---

[81] I have also run this model separating the nursing majors from the non-nursing health-related majors and the differences across institutional type remain for both categories of health major. This analysis is available from the author upon request.

Lang and Weinstein (2013) also control for major field of study in their analysis of returns to certificates, associate's degrees, and bachelor's degrees at for-profit and not-for-profit institutions, finding that this does not eliminate a non-trivial but statistically insignificant difference between the returns to a for-profit and a not-for-profit certification. My preferred specification differs from theirs in that the major field is interacted with the type of degree-granting institution, and my coefficients are statistically significant while theirs are not. These coefficients tell us that that the institutional status of the degree-granting college does matter, in terms of earning power, once we take into account the major field of study, particularly for degrees in health-related fields.

Table 3.8 presents the full model as specified in column (5) of Table 3.7 for a selection of other labor market outcomes: log hourly wage, log hours worked per week, full-time employment, and part-time employment. In column (1), a negative and statistically significant coefficient on GED suggests that the lower average marginal earnings of GED-earners can be attributed to a lower hourly wage. Health degrees from public institutions and vocational/technical degrees from both public and for-profit institutions are associated with positive, statistically significant average marginal hourly wages. In column (2), the only degree that appears to have a statistically significant effect on hours worked per week are for-profit business-related degrees (14.3 percent more hours worked per week). Columns (3) and (4) are linear probability models of full-time and part-time employment and include higher numbers of individuals and observations due to the exclusion of zero-earners from the log weekly earnings, log hourly wage, and log hours per week specifications. Column (3) suggests that individuals with for-profit business degrees and for-profit vocational/technical degrees are 22 and 23.1 percent more likely, respectively, to find full-time employment than high school diploma-holders. Column (4) suggests that public degree holders in any major field are marginally more likely to find any employment (ranging from 4.4 percent more likely with an academic degree to 10.7 percent more likely with a business degree); additionally, for-profit degree-holders with majors in health or vocational/technical fields are 13 and 15 percent more likely, respectively, to find any employment.

3.6 Replication of Cellini and Chaudhary (2012)

Because Cellini and Chaudhary (2012) also use the NLSY97 Geocode, I have provided Table 3.10 through Table 3.15, which roughly replicate their methodology but use my own sample selection.[82] For this part of the analysis, I have identified individuals in my sample who report working towards an associate degree but have not completed the degree. Since these models use individual fixed-effects to identify earnings differentials between public and for-profit college work and degrees,[83] this sample includes only individuals having earned or worked towards a tow-year degree, excluding those who go on to earn higher degrees, and also excluding those with no earnings history prior to enrolling in a two-year degree program. As a result, the sample is smaller than my original sample. Table 3.10 displays descriptive statistics; comparing these to Table 3.2, it is clear that adding non-completers to the sample of two-year completers results in lowering the mean SES of the sample, as measured by racial composition, ASVAB percentiles, and parents' income in 1997. The samples are statistically different in most aspects, either choosing for-profit schools or being chosen by these schools. Interestingly, the percentage of completers is significantly higher among for-profit attendees (35.8 percent compared to 11.9 percent of public attendees).

Table 3.11 presents descriptive statistics on labor market outcomes before beginning work on a two-year degree (pre-education) and after either completing a two-year degree, or the final year during the sample period in which the individual was working on a two-year degree (post-education).[84] Using their sample selection methods, Cellini and Chaudhary (2012) find few statistically significant differences between the public and for-profit individuals in their Table 1; conversely, I find all outcome coefficients except average hours worked per week to be statistically different between the two groups in the post-education period. In particular, I find average weekly earnings and average wages to be lower among for-profit attendees than among public attendees

---

[82] These tables correspond to Table 1 through Table 3A in Cellini and Chaudhary (2012). I have also attempted to purely replicate their paper. These attempts and comments thereupon can be found in my working paper Turner (2013).
[83] See Cellini and Chaudhary (2012) for a full description of the model and data methodology. See Turner (2013) for further details on the difference between my methodology and that of the original authors.
[84] This differs from the methodology used by Cellini and Chaudhary (2012) to determine pre- and post-education periods. See Turner (2013) for further details.

post-education. This pre/post analysis of two-year attendees tells a somewhat different story than Table 3.4, in which I described only two-year completers and found very few statistical differences, suggesting that the non-completers of degrees at public and for-profit colleges differ much more than the degree-completers.

Table 3.12 presents estimates of returns to college attendance using individual fixed-effects models analogous to those in Table 2A of Cellini and Chaudhary (2012). The findings are very similar to those of the original authors, with returns to some college ranging from 6.5 to 18.1 percent over pre-college earnings. As in the original authors' table, the coefficients in column (4), which excludes earnings from ages sixteen and seventeen, are not statistically different than zero, suggesting that much of the identification in the remaining models/columns is off of individuals' earnings at a very young age. The coefficients in columns (6) and (7), in which for-profit and public attendees, respectively, are separated and run in their own regressions, are much higher than estimates from the column (2) in which they are pooled, suggesting that the age- and year-fixed effects differ significantly for the two groups. This may be a result of an unidentified, underlying selection structure. Table 3.13 is analogous to Table 2B in Cellini and Chaudhary (2012) and includes coefficients for degree completion as well as college attendance. Again, my findings are similar to those of the original authors; a notable difference is that they find separate, statistically different effects of college attendance and degree completion across most specifications, while I find positive, statistically significant coefficients only on college attendance. Column (8) is my own addition, excluding non-completers from the analysis. Here the sample size is only 262 individuals, and neither the coefficient on degree completion nor on for-profit degree completion are statistically different than zero; I suspect this is because the model lacks explanatory power due to the low number of individuals with short earnings history rather than because completion of a two-year degree has no effect on earnings.

Table 3.14 is the analogue of Cellini and Chaudhary's Table 3A; they find positive, statistically significant effects of college attendance on weekly earnings, hours worked per week, and the probabilities of full-time or any employment. I find positive, statistically significant effects of college attendance on weekly earnings, average wage, and hours worked per week, but none on the probability of employment. Table 3.15

includes coefficients on degree completion and is analogous to Table 3B in Cellini and Chaudhary. They find statistically significant effects of attendance and completion that differ by type of institution across all outcomes; using my data selection methods, I find quite different results. The only coefficients that differ statistically from zero are post-attendance (7 percent higher) and for-profit post-attendance (8.1 percent lower) on log weekly earnings (suggesting that after working towards a public degree, individuals increase their earnings potential while after working towards a for-profit degree the two effects cancel each other out), and post-attendance on log hourly wages (3.9 percent higher).

Table 3.16 compares the base model from Table 3.12, column (2), in which post-attendance raises weekly earnings by 7.5 percent and for-profit attendance does not statistically differ from public attendance, to a model that separates the effects of attendance and for-profit attendance by major area of study. Here I find no earnings effect of working towards an academic/other/unknown degree from either sector. In business-related or vocational/technical study, post-attendance earnings gain 14.1 and 14.9 percent respectively with no statistically significant difference associated with for-profit attendance. Health-related study results in a 7.5 percent increase in earnings, and an additional 14.4 percent decrease (resulting in about a 7 percent decrease overall) in earnings to for-profit attendees. These results mirror my findings from Table 3.7, suggesting that while for-profit attendance alone may not impact earnings, in conjunction with the major area of study (particularly in health-related fields), for-profit attendance may impact earnings differentially than public attendance.

Similarly, Table 3.17 compares the base model from Table 3.13, column (2), which includes coefficients on college attendance and degree completion, to a model which separates these coefficients by major area of study. I find that working toward or completing an academic/other/unknown degree has no statistically significant effect on log weekly earnings. Working toward a business-related degree increases weekly earnings on average by 13.8 percent, but working toward that degree at a for-profit institution lowers that figure by 15.3 percent of earnings, suggesting that there is no earnings effect for these individuals. Working towards a health degree seems to have no statistically significant effect on earnings unless the work is done at a for-profit institution, in which case earnings decrease by 24.2 percent on average. Work towards a vocational/technical degree seems to boost earnings by 13.1 percent and completing the degree boosts earnings by another 14.5 percent, with no statistical difference in earnings effects for for-profit attendees. These results are, again, similar in some ways to my findings from Table 3.7, though as in Table 3.13, I find little evidence of a degree completion effect while college attendance is included in the model. Again, I suspect this is due more to the low sample size and short earnings histories of individuals in the sample rather than because completion of a two-year degree has no effect on earnings. My inclusion of major area of study in the fixed-effects models demonstrates two points: (1) that the analysis of earnings across graduates and attendees from different institutional sectors requires finer tuning to determine which types of students do well in which types of institutions; and (2) that more data, either including additional individuals or the same individuals observed over longer periods of time, is needed in order to make an individual fixed-effects approach effective in discerning the causal effects of a for-profit education.

3.7 Discussion

This chapter takes advantage of the wealth of family background information available in the NLSY97 and the ability to link institutional data from IPEDS to the NLSY through additional variables available through the restricted-access Geocode data to empirically analyze the differences between individuals pursuing different types of two-year degrees at different types of colleges, and to estimate the average marginal effects of completing different types of two-year degrees on the labor market outcomes of young adults. In addition, I replicate a concurrent working paper, Cellini and Chaudhary (2012), using the same underlying data but different sample selection techniques, and I extend their analysis to investigate different major areas of study within each type of college.

There is obviously selection into two-year colleges; Table 3.1 shows that two-year completers differ significantly from high school completers even before removing the high achievers (those who go on to complete a bachelor or higher degree) from the sample. Among two-year completers, there is selection into the type of college, as shown in Table 3.2. For-profit students have lower SES than public students and are much less likely to go on to earn a bachelor degree. However, Table 3.10 shows that among degree-attendees, for-profit students are more likely to complete their two-year degree than their public counterparts. Additionally, there is selection into type of major. Table 3.6 shows that individuals pursuing for-profit health-related degrees differ significantly from their public counterparts while individuals pursuing other majors do not differ significantly from their public counterparts. Table 3.5 shows that students in public two-year programs are 20 percent more likely to choose an academic/other/unknown major than students in for-profit institutions, while students in for-profits are twice as likely to be studying in a vocational/technical field, so selection into type of college and selection into type of major are not independent. Previous studies have identified the nature of some of the selection mechanisms: financial (Cellini 2009, 2010); convenience (Bailey, Badway, and Gumport 2003); recruitment (Breneman, Pusser, and Turner 2006); desire to transfer to a four-year college program (Deming, Goldin, and Katz 2012); product differentiation (Kinser 2005). Other mechanisms, such as ability grouping and other aspects of personal preference no doubt also play a part in the selection process. Unraveling and modeling the selection process is an essential next step for future research into the analysis of for-profit higher education.

My findings indicate that major area of study has a differential effect on the returns to a degree, and that earnings effects can differ within major area study depending on the institutional status of the degree-granting college. In particular, a two-year degree in an academic/other/unknown area of study does not have a statistically significant impact on earnings regardless of sector of institution. Business-related and vocational/technical degrees provide positive and statistically significant returns of 20 and 32.7 percent respectively; evidence from OLS regression does not indicate statistically different earnings returns for public and for-profit degree-holders within these two areas of study, but analysis of the industries these degree-holders end up working in suggests that for-profit business-related majors are less likely to be unemployed than their public counterparts, and for-profit vocational/technical majors are more likely to procure high-paying jobs in information technology than their public counterparts. Additionally, regression analysis finds that for-profit business and vocational/technical majors are more likely to find full-time employment than their public counterparts. This suggests that in these fields for-profit degree-earners may have some labor market advantages over their public counterparts, especially considering that there is most likely a negative bias in the OLS estimates due to possible selection on unobservables.

OLS and fixed-effects regressions provide strong evidence that individuals with health-related degrees from for-profit institutions earn more than 20 percent less than their public degree-holding counterparts. In addition, relatively few of these individuals end up working in a health-related field (30 percent versus almost 70 percent of public degree-holders) and many more do not work at all following degree completion (25 percent versus only 5 percent of public degree-holders). Because of the selection going on, the causality is unclear, but differential outcomes of for-profit and public students exist and depend upon the major area of study.

The descriptive analysis provided in this paper is a first step toward unraveling the selection bias that exists in the matching of individuals with two-year institutions and major fields of study. Previous and concurrent studies that attempt to correct for selection bias using matching models can only account for selection on observables; those using individual fixed-effects require longer earnings histories than are currently available for individuals studying in the for-profit sector. What is the underlying structure of the selection into public and for-profit schools and furthermore into major areas of study? This is the question that needs to be answered before we can identify the causal effects of a for-profit education. Policy-makers should take note that this preliminary analysis of the returns (which can be thought of as the private benefits) to public and for-profit degrees does not provide unambiguous evidence in favor of one sector over another, but rather a first look into the "black box" in which students, institutions, and major areas of study come together and jointly determine labor market outcomes.

3.8 References

Bailey, T., Badway, N., and Gumport, P. J. *For-Profit Higher Education and Community Colleges.* Stanford, Calif.: National Center for Postsecondary Improvement, 2003.

Black, Dan A., Seth Sanders, and Lowell Taylor. 2003. "The Economic Reward for Studying Economics." *Economic Inquiry* 41(3): 365-377.

Breneman, David W., Brian Pusser, and Sarah E. Turner, eds. 2006. *Earnings from Learning: The Rise of For-Profit Universities.* Albany, NY: State University of New York Press.

Cameron, S. V. and J. J. Heckman. 1993. "The Nonequivalence of High School Equivalents." *Journal of Labor Economics* 11(1, Part 1): 1-47.

Card, David. 1999. "The Causal Effect of Education on Earnings." Chapter 30 of the *Handbook of Labor Economics*, Vol. 3A, eds. Orley Ashenfelter and David Card. North-Holland: Amsterdam.

Cellini, Stephanie Riegg. 2009. "Crowded Colleges and College Crowd-Out: The Impact of Public Subsidies on the Two-Year College Market," *American Economic Journal: Economic Policy* 1(2): 1-30.

Cellini, Stephanie Riegg. 2010. "Financial Aid and For-Profit Colleges: Does Aid Encourage Entry?" *Journal of Policy Analysis and Management* 29(3): 526-52.

Cellini, Stephanie Riegg. 2012. "For-Profit Higher Education: An Assessment of Costs and Benefits." *National Tax Journal*, 65(1), 153-180.

Cellini, Stephanie Riegg and Latika Chaudhary. The labor market returns to a private two-year college education. Working paper, April 2011.

Cellini, Stephanie Riegg and Latika Chaudhary. 2012. "The labor market returns to a for-profit college education." NBER Working Paper 18343.

Cellini, Stephanie Riegg and Claudia Goldin, 2012. "Does Federal Student Aid Raise Tuition?: New Evidence on For-Profit Colleges." NBER Working Paper 17827.

Deming, David J., Claudia Goldin, and Lawrence F. Katz. 2012. The for-profit postsecondary school sector: Nimble critters or agile predators? *Journal of Economic Perspectives.* Vol. 26, No. 1, pp.139-164.

Greenblatt, Mark. "Whistle-Blower: For-Profit College Operator Allegedly Inflates Job Placement Rates." *ABC News*. 26 Nov. 2012. Web. 28 Nov. 2012. <abcnews.go.com>.

Hammermesh, Daniel S. 2007. "Viewpoint: Replication in Economics." *Canadian Journal of Economics*. Vol. 40, No. 3A, pp. 715-733.

Integrated Postsecondary Education Data System. http://nces.ed.gov/ipeds/ Accessed August 3, 2011.

Kinser, K. (2005), A profile of regionally accredited for-profit institutions of higher education. New Directions for Higher Education, 2005: 69–83.

Kreighbaum, Andrew. “Congress Shines Light on For-Profit Education Sector as Industry Makes Lobbying Surge.” *OpenSecrets.org Center for Responsive Politics.* 23 June 2010. Web. 7 Feb. 2013.

Lang, Kevin, and Russell Weinstein. 2013. “The Wage Effects of Not-For-Profit and For-Profit Certifications: Better Data, Somewhat Different Results.” NBER Working Paper 19135.

Lewin, Tamar. "Senate Committee Report on For-Profit Colleges Condemns Costs and Practices." *New York Times*. 29 July 2012. Web. 28 Nov. 2012. <nytimes.com>.

Lewin, Tamar. “For-Profit Colleges Mislead Students, Report Finds.” *New York Times.* 3 Aug. 2012. Web. 28 Nov. 2012.

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings.* New York: Columbia University Press for the National Bureau of Economic Research.

Parsad, B., and Lewis, L. 2008. Distance Education at Degree-Granting Postsecondary Institutions: 2006–07 (NCES 2009–044). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, D.C.

Ross, Terris, Grace Kena, Amy Rathbun, Angelina KewalRamani, Jijun Zhang, Paul Kristapovich, and Eileen Manning. 2012. Higher Education: Gaps in Access and Persistence Study (NCES 2012-046). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, D.C.

Ruch, R. Higher Ed, Inc.: The Rise of the For-Profit. Baltimore, Md.: Johns Hopkins University Press, 2001.

Turner, Abby Clay. 2013. “Some Comments on the Replication of ‘The Labor Market Returns to a For-Profit College Education.’” Working Paper.

Symonds, William C., Robert Schwartz, and Ronald F. Ferguson. 2011. Pathways to prosperity: Meeting the challenge of preparing young Americans for the 21st century. Cambridge, MA: Pathways to Prosperity Project, Harvard University Graduate School of Education.

Waits, Tiffany and Laurie Lewis. 2003. Distance Education at Degree-Granting Postsecondary Institutions: 2000-2001. NCES 2003017.

3.9 Tables and Figures

Table 3.1: Descriptive Statistics

	All HS Completers		***All Two-Year Completers***		***HS Completers, without Bachelor's Completers***		***Two-Year Completers, without Bachelor's Completers***	
	(1)		**(2)**		**(3)**		**(4)**	
Male	50.2%	(50.0)	42.7%	(49.5)***	52.3%	(49.7)	44.0%	(49.7)***
White	51.0%	(50.0)	52.8%	(50.0)	47.5%	(50.0)	52.5%	(50.0)**
Black	25.3%	(43.5)	20.8%	(40.6)***	27.7%	(44.7)	22.4%	(41.7)***
Asian	1.8%	(13.4)	1.6%	(12.5)	1.3%	(11.3)	1.6%	(12.5)
Hispanic	19.7%	(39.8)	21.3%	(41.0)	21.5%	(41.1)	20.6%	(40.5)
Other race	2.1%	(14.4)	3.5%	(18.4)**	1.9%	(13.6)	3.0%	(17.0)
Foreign born	12.6%	(33.1)	14.3%	(35.1)	12.5%	(33.1)	12.9%	(33.5)
ESL	6.0%	(23.7)	7.8%	(26.8)	6.6%	(24.7)	6.7%	(25.1)
ASVAB Percentile	0.4876	(.2846)	0.5161	(.2499)***	0.4325	(0.2720)	0.5057	(0.2551)***
HH net worth 1997	$98,868	(142,733)	$94,889	(126,906)	$79,579	(122,639)	$92,542	(127,393)**
Father's highest grade completed	12.9	(4.0)	12.8	(2.9)	12.4	(4.2)	12.8	(2.9)***
Mother's highest grade completed	12.7	(3.3)	12.5	(2.8)**	12.4	(3.3)	12.4	(2.6)
Mom's age at birth of first child	23.0	(4.9)	23.5	(5.1)***	22.4	(4.8)	23.4	(5.1)***
Mom's age at birth	25.7	(5.5)	26.7	(5.4)***	25.2	(5.5)	26.1	(5.3)***
Years spent as part-time student	0.4	(1.0)	0.9	(1.4)***	0.5	(1.0)	0.9	(1.5)***
Years spent as full-time student	2.1	(2.2)	3.8	(2.1)***	1.4	(1.8)	3.4	(1.9)***
Earn GED	12.8%	(33.4)	4.0%	(19.6)***	16.0%	(36.7)	4.6%	(20.9)***
Earn a certification	1.2%	(10.9)	1.3%	(11.2)	1.3%	(11.3)	1.0%	(9.9)
Earn a two-year degree	8.2%	(27.5)	-		8.3%	(27.5)	-	
Earn a bachelor degree	20.2%	(40.2)	19.7%	(39.8)	-		-	
N	***7,665***		***629***		***6,106***		***505***	

Standard deviations in parentheses.
* Statistically different from non-two-year completers at the 10%-level.
** Statistically different from non-two-year completers at the 5%-level.
*** Statistically different from non-two-year completers at the 1%-level.

Table 3.2: Descriptive Statistics for Two-Year Completers, by Institutional Status

	All Two-Year Completers						*Two-Year Completers, without Bachelor's Completers*					
	Public		***For-Profit***		***Private, Not-For-Profit***		***Public***		***For-Profit***		***Private, Not-For-Profit***	
Variable	**(1)**		**(2)**		**(3)**		**(4)**		**(5)**		**(6)**	
Male	41.9%	(49.4)	47.6%	(50.2)	44.0%	(50.7)	43.1%	(49.6)	47.5%	(50.2)	47.4%	(51.3)
White	54.2%	(49.9)	42.7%	(49.7)**	64.0%	(49.0)	55.0%	(49.8)	42.4%	(49.7)**	57.9%	(50.7)
Black	20.6%	(40.5)	25.2%	(43.7)	12.0%	(33.2)	22.5%	(41.8)	25.3%	(43.7)	10.5%	(31.5)
Asian	1.3%	(11.2)	2.9%	(16.9)	0.0%	(0.0)**	1.1%	(10.5)	3.0%	(17.2)	0.0%	(0.0)**
Hispanic	20.3%	(40.3)	26.2%	(44.1)	20.0%	(40.8)	18.6%	(39.0)	26.3%	(44.2)	26.3%	(45.2)
Other race	3.6%	(18.7)	2.9%	(16.9)	4.0%	(20.0)	2.8%	(16.5)	3.0%	(17.2)	5.3%	(22.9)
Foreign born	14.8%	(35.6)	14.0%	(34.9)	12.5%	(33.8)	13.4%	(34.1)	12.4%	(33.1)	11.1%	(32.3)
ESL	7.5%	(26.3)	8.5%	(28.1)	9.5%	(30.1)	6.0%	(23.8)	7.7%	(26.8)	12.5%	(34.2)
ASVAB Percentile	0.5302	(0.2437)	0.4391	(0.2610)**	0.5390	(0.2553)	0.5196	(0.2489)	0.4400	(0.2605)**	0.5410	(0.2822)
HH net worth 1997	$99,442	(127,610)	$80,019	(118,782)	$98,571	(141,281)	$97,312	(126,869)	$79,714	(121,537)	$101,801	(148,901)
Father's highest grade completed	12.9	(2.8)	12.5	(2.9)	12.7	(3.1)	12.9	(2.8)	12.4	(2.9)	12.6	(3.3)
Mother's highest grade completed	12.8	(4.9)	12.1	(2.5)*	12.6	(2.2)	12.8	(5.2)	12.1	(2.5)*	12.5	(2.2)
Mom's age at birth of first child	23.7	(5.1)	23.1	(4.8)	22.4	(5.4)	23.6	(5.1)	23.0	(4.9)	21.8	(5.9)
Mom's age at birth	26.6	(5.5)	25.7	(4.8)*	25.1	(4.4)	26.4	(5.4)	25.6	(4.9)	25.3	(4.6)
Years spent as part-time student	1.1	(1.6)	0.6	(1.0)***	0.8	(1.2)	1.1	(1.6)	0.6	(1.0)***	0.8	(1.3)
Years spent as full-time student	4.4	(2.0)	3.5	(1.7)***	4.3	(2.0)	4.0	(2.0)	3.3	(1.6)***	4.1	(2.2)
Earn GED	3.6%	(18.7)	5.8%	(23.5)	0.0%	(0.0)***	4.4%	(20.6)	5.0%	(22.0)	0.0%	(0.0)***
Earn a certification	1.3%	(11.2)	1.9%	(13.9)	0.0%	(0.0)**	0.8%	(9.1)	2.0%	(14.1)	0.0%	(0.0)*
Earn a bachelor degree	23.7%	(42.6)	3.9%	(19.4)***	24.0%	(43.6)	-		-		-	
N	***472***		***103***		***25***		***360***		***99***		***19***	

Standard deviations in parentheses.
* Statistically different from Public at the 10%-level.
** Statistically different from Public at the 5%-level.
*** Statistically different from Public at the 1%-level.

Table 3.3: Outcome Descriptive Statistics, Age 25+

Variable	*All HS Completers* (1)		*All Two-Year Completers* (2)		*HS Completers, without Bachelor's Completers* (3)		*Two-Year Completers, without Bachelor's Completers* (4)	
Average Weekly Earnings	$611.54	(785.93)	$583.46	(431.20)	$576.73	(746.33)	$564.76	(444.93)
Average Hourly Wage	$17.44	(21.09)	$17.50	(16.42)	$16.75	(21.65)	$17.28	(17.45)
Average Hours per Week	35.5	(10.0)	33.8	(9.8)***	35.1	(9.8)	33.3	(9.9)***
Full-Time Employment	49.2%	(39.6)	48.1%	(40.1)	47.3%	(39.6)	45.6%	(40.0)
Any Employment	74.2%	(36.5)	80.3%	(31.7)***	71.6%	(37.7)	78.0%	(33.5)***
Age	26.5	(0.8)	26.5	(0.7)	26.5	(0.7)	26.5	(0.7)
# Biological Children in HH	0.6	(0.9)	0.5	(0.8)***	0.7	(1.0)	0.6	(0.9)***
# of Biological Children Outside of HH	0.1	(0.5)	0.0	(0.2)***	0.2	(0.5)	0.1	(0.2)***
N	***7,665***		***629***		***6,106***		***505***	

Standard deviations in parentheses.
* Statistically different from non-two-year completers at the 10%-level.
** Statistically different from non-two-year completers at the 5%-level.
*** Statistically different from non-two-year completers at the 1%-level.

Table 3.4: Outcome Descriptive Statistics for Two-Year Completers, by Institution Type, Age 25+

	All Two-Year Completers						*Two-Year Completers, without Bachelor's Completers*					
	Public		***For-Profit***		***Private, Not-For-Profit***		***Public***		***For-Profit***		***Private, Not-For-Profit***	
Variable	**(1)**		**(2)**		**(3)**		**(4)**		**(5)**		**(6)**	
Average Weekly Earnings	$579.99	(448.12)	$549.36	(328.64)	$659.55	(421.25)	$553.51	(467.91)	$544.30	(329.09)	$657.51	(457.16)
Average Hourly Wage	$17.55	(17.01)	$16.41	(15.88)	$17.93	(8.54)	$17.25	(18.39)	$16.35	(16.18)	$17.86	(8.70)
Average Hours per Week	33.5	(10.0)	34.8	(9.5)	33.8	(9.8)	32.7	(10.1)	34.7	(9.5)*	32.9	(11.1)
Full-Time Employment	46.7%	(40.4)	56.9%	(38.9)**	43.3%	(36.4)	42.7%	(40.1)	56.9%	(38.8)***	37.6%	(36.5)
Any Employment	80.1%	(31.9)	83.4%	(29.7)	77.7%	(32.4)	76.8%	(34.2)	83.2%	(29.9)*	78.3%	(33.0)
Age	26.5	(0.7)	26.6	(0.7)	26.6	(0.7)	26.5	(0.7)	26.6	(0.7)	26.5	(0.7)
# Biological Children in HH	0.5	(0.8)	0.6	(0.9)	0.8	(1.0)	0.5	(0.8)	0.6	(0.9)	0.9	(1.0)
# of Biological Children Outside of HH	0.0	(0.2)	0.1	(0.4)	0.0	(0.0)***	0.0	(0.2)	0.1	(0.4)	0.0	(0.0)***
N	***469***		***103***		***25***		***357***		***99***		***19***	

Standard deviations in parentheses.
* Statistically different from Public at the 10%-level.
** Statistically different from Public at the 5%-level.
*** Statistically different from Public at the 1%-level.

Table 3.5: Area of Major for Two-Year Completers, by Institution Type

	All Two-Year Completers			*Two-Year Completers, without Bachelor's Completers*		
Major	***Public***	***For-Profit***	***Private, Not-For-Profit***	***Public***	***For-Profit***	***Private, Not-For-Profit***
Academic/Other/Unknown	54.7%	30.8%	37.5%	52.8%	28.0%	38.9%
Business	13.0%	18.3%	8.3%	12.9%	19.0%	5.6%
Health (Non-Nursing)[a]	7.9%	19.2%	4.2%	8.7%	20.0%	5.6%
Nursing[a]	10.7%	2.9%	25.0%	12.1%	3.0%	16.7%
Vocational/Technical	13.7%	28.8%	25.0%	13.5%	30.0%	33.3%
N	***468***	***104***	***24***	***356***	***100***	***18***

[a] Together the health (non-nursing) and the nursing majors create the health-related major category used in the rest of the analysis.

Table 3.6: Selected Descriptive Statistics for Two-Year Completers, by Major and Institution Type

	All Two-Year Completers						*Two-Year Completers, without Bachelor's Completers*					
Major	***Public***		***For-Profit***		***Private, Not-For-Profit***		***Public***		***For-Profit***		***Private, Not-For-Profit***	
			ASVAB Percentile						***ASVAB Percentile***			
Humanities/Other	0.5284	(0.2478)	0.4678	(0.2711)	0.5086	(0.2558)	0.5152	(0.2508)	0.4748	(0.2704)	0.5204	(0.3015)
Business	0.5476	(0.2398)	0.4539	(0.3044)	a		0.5154	(0.2510)	0.4539	(0.3044)	a	
Health	0.5456	(0.2256)	0.3843	(0.1948)***	0.6057	(0.3251)	0.5404	(0.2333)	0.3843	(0.1948)**	a	
Vocational/Technical	0.4961	(0.2593)	0.4302	(0.2680)	0.4695	(0.2344)	0.5058	(0.2738)	0.4302	(0.2680)	0.4695	(0.2344)
			HH Net Worth 1997 (in 2010 $)						***HH Net Worth 1997 (in 2010 $)***			
Humanities/Other	$98,566	(129,777)	$87,146	(128,344)	$83,278	(91,565)	$99,597	(129,771)	$87,358	(138,117)	$83,278	(91,565)
Business	$105,896	(139,194)	$73,717	(69,011)	a		$99,421	(146,789)	$73,717	(69,011)	a	
Health	$85,029	(102,128)	$73,580	(124,525)	$212,960	(237,348)	$86,600	(106,237)	$73,580	(124,525)	a	
Vocational/Technical	$114,019	(137,117)	$80,229	(135,367)	$43,738	(40,319)**	$101,297	(126,222)	$80,229	(135,367)	$43,738	(40,319)**
			Weekly Earnings, Age 25+						***Weekly Earnings, Age 25+***			
Humanities/Other	$525.98	(424.18)	$494.25	(303.31)	$591.48	(385.15)	$501.59	(456.56)	$469.55	(296.17)	$533.35	(390.72)
Business	$616.54	(313.00)	$527.09	(270.93)	a		$549.46	(283.29)	$527.09	(270.93)	a	
Health	$726.28	(637.14)	$411.73	(185.04)***	$583.67	(372.33)	$687.52	(633.00)	$411.73	(185.05)***	a	
Vocational/Technical	$585.62	(281.14)	$724.18	(401.43)	$913.11	(552.05)	$578.53	(296.04)	$720.42	(394.73)	$913.11	(552.05)
			Earn a Bachelor's Degree									
Humanities/Other	25.9%	(43.9)	11.4%	(32.3)**	20.0%	(42.2)						
Business	25.0%	(43.7)	0.0%	(0.0)***	a							
Health	15.1%	(36.0)	0.0%	(0.0)***	42.9%	(53.5)						
Vocational/Technical	25.4%	(43.9)	0.0%	(0.0)***	0.0%	(0.0)***						

a = Less than 5 observations in this category.
Standard deviations in parentheses.
* = Statistically different than the mean for public students at the 0.1 level
** = Statistically different than the mean for public students at the 0.05 level
*** = Statistically different than the mean for public students at the 0.01 level

Table 3.7: OLS Regression Results: Effects on Log Weekly Earnings

	Coefficients on Log Weekly Earnings				
Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
GED	-0.129***	-0.059	-0.059	-0.055	-0.057
	(0.040)	(0.039)	(0.039)	(0.039)	(0.039)
Two-Year Degree	0.069	0.081*			
	(0.046)	(0.045)			
Public Two-Year Degree			0.079*		
			(0.045)		
For-Profit Two-Year Degree			0.084		
			(0.093)		
Private, Not-For-Profit Two Year Degree			0.087		
			(0.490)		
Major: Academic/Other/Unknown				-0.067	
				(0.065)	
Major: Business Field				0.200***	
				(0.070)	
Major: Health Field				0.135	
				(0.099)	
Major: Vocational/Technical				0.327***	
				(0.089)	
Public: Academic/Other/Unknown					-0.043
					(0.066)
For-Profit: Academic/Other/Unknown					-0.264
					(0.207)
Public: Business Field					0.165**
					(0.083)
For-Profit: Business Field					0.243**
					(0.116)
Public: Health Field					0.282***
					(0.068)
For-Profit: Health Field					-0.045
					(0.095)
Public: Vocational/Technical					0.177*
					(0.107)
For-Profit: Vocational/Technical					0.419***
					(0.154)
Includes Demographic Controls	No	Yes	Yes	Yes	Yes
Includes Private, Not-For-Profit Students	Yes	Yes	Yes	Yes	No
R-squared	0.0282	0.1245	0.1245	0.1287	0.1235
N	8,297	8,297	8,297	8,297	8,264
N	2,454	2,454	2,454	2,454	2,443

Robust standard errors, clustered by individual, in parentheses. All Specifications control for part- or full-time school attendance and include age dummy variables. Demographic controls include sex, race, ethnicity, ASVAB percentile, ESL, foreign-born, mother's age at first birth, mother's age at birth, the highest grade completed by biological mother and father, the geographic region, number of biological children in and out of the household and interaction with sex. I have run these regressions with and without year dummy variables; the year does not seem to matter as long as age is included and the dollars are CPI-U adjusted.

* = Statistically significant at the 0.1 level

** = Statistically significant at the 0.05 level

*** = Statistically significant at the 0.01 level

Table 3.8: OLS Regression Results: Effects on Other Outcomes

	Coefficients on Other Outcomes			
	(1)	**(2)**	**(3)**	**(4)**
Variables	**Log of Hourly Wage**	**Log of Hours Worked per Week**	**Full Time Employment**	**Any Employment**
GED	-0.064*	0.013	-0.009	-0.007
	(0.034)	(0.020)	(0.021)	(0.017)
Public: Academic/Other/Unknown	-0.033	-0.005	-0.041	0.044*
	(0.049)	(0.038)	(0.042)	(0.024)
For-Profit: Academic/Other/Unknown	-0.280	-0.036	-0.002	-0.032
	(0.202)	(0.064)	(0.086)	(0.088)
Public: Business Field	0.088	0.080	0.074	0.107***
	(0.077)	(0.061)	(0.104)	(0.031)
For-Profit: Business Field	0.100	0.143***	0.220***	0.049
	(0.095)	(0.048)	(0.079)	(0.049)
Public: Health Field	0.355***	-0.079	-0.012	0.059*
	(0.056)	(0.062)	(0.060)	(0.031)
For-Profit: Health Field	-0.051	0.010	-0.063	0.130***
	(0.085)	(0.056)	(0.112)	(0.025)
Public: Vocational/Technical	0.129*	0.050	0.120	0.097***
	(0.075)	(0.061)	(0.073)	(0.029)
For-Profit: Vocational/Technical	0.279**	0.142	0.231**	0.150***
	(0.114)	(0.087)	(0.089)	(0.020)
Includes Demographic Controls	Yes	Yes	Yes	Yes
Includes Private, Not-For-Profit Students	No	No	No	No
R-squared	0.1168	0.0650	0.0640	0.0341
No. of Observations	8,275	8,388	9,580	9,580
No. of Individuals	2,444	2,456	2,631	2,631

Robust standard errors, clustered by individual, in parentheses. All specifications control for part- or full-time school attendance and include age dummy variables. Demographic controls include sex, race, ethnicity, ASVAB percentile, ESL, foreign-born, mother's age at first birth, mother's age at birth, the highest grade completed by biological mother and father, the geographic region, number of biological children in and out of the household and interaction with sex. I have run these with and without year dummy variables; the year does not seem to matter as long as age is included and the dollars are CPI-U adjusted.

* = Statistically significant at the 0.1 level

** = Statistically significant at the 0.05 level

*** = Statistically significant at the 0.01 level

Table 3.9: Two-Year Completers: Industry Working In At Age 26, by Major and Institution Type

	Business-Related Majors		*Health-Related Majors*		*Technical/Vocational Majors*	
	Public	***For-Profit***	***Public***	***For-Profit***	***Public***	***For-Profit***
No Response/Uncodable[a]	18.2%	5.9%	16.7%	9.1%	23.9%	10.0%
Not Working[b]	16.7%	-	5.0%	25.0%	2.9%	3.7%
Agriculture, Forestry, Fishing and Hunting; Mining	2.8%	-	-	-	5.7%	-
Construction; Manufacturing; Transportation and Warehousing	8.3%	12.5%	1.7%	15.0%	25.7%	22.2%
Trade[c]	13.9%	6.3%	8.3%	15.0%	5.7%	18.5%
Information and Communications	5.6%	-	-	-	5.7%	14.8%
Finance, Insurance, Real Estate, and Rental and Leasing	22.2%	18.8%	-	-	5.7%	3.7%
Professional, Scientific, Management, Administrative, and Waste Management Services	5.6%	12.5%	3.3%	5.0%	14.3%	18.5%
Educational, Health and Social Services	5.6%	12.5%	68.3%	30.0%	2.9%	-
Services; Public Administration	19.4%	37.5%	13.3%	10.0%	31.4%	18.5%

[a] Categories exclusive of No Response/Uncodable are calculated as a percentage of those reporting a category or not working.
[b] Valid non-response is interpreted as Not Working.
[c] The trade category includes jobs in retail.

Table 3.10: Descriptive Statistics for Fixed-Effects Analysis

	Public		For-Profit	
Variable	**Mean**	**St. Dev.**	**Mean**	**St. Dev.**
Male	47.0%	(49.9)	43.3%	(49.7)
White	45.9%	(49.9)	38.1%	(48.7)**
Black	26.7%	(44.2)	34.4%	(47.6)**
Asian	1.5%	(12.3)	0.9%	(9.6)
Hispanic	23.5%	(42.4)	23.3%	(42.3)
Other	2.4%	(15.3)	3.3%	(17.8)
Foreign-Born	14.6%	(35.4)	15.1%	(35.9)
Primary Language Not English	5.8%	(23.4)	8.8%	(28.4)
ASVAB Percentile	0.4453	(0.2474)	0.4021	(0.2478)**
Parents' Income in 1997[a]	$60,179	(46,573)	$55,005	(44,611)
Ever Earned a 2-year Degree	11.9%	(32.4)	35.8%	(48.1)***
Major: Academic/Other/Unknown	51.0%	(51.0)	21.9%	(41.1)***
Major: Business-Related	14.0%	(34.7)	20.9%	(40.8)*
Major: Health-Related	16.1%	(36.8)	23.7%	(42.6)***
Major: Vocational/Technical	18.9%	(39.1)	33.5%	(47.3)***
N	1552		215	

[a] In 2010 dollars.
Standard deviations in parentheses.
* = Statistically different than public at the 0.1 level
** = Statistically different than public at the 0.05 level
*** = Statistically different than public at the 0.01 level

Table 3.11: Pre- and Post-Education Summary Statistics

	Pre-Education				Post-Education			
	Public		For-Profit		Public		For-Profit	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Weekly Earnings[a]	$238.34	(371.47)	$241.54	(165.65)	$538.37	(1039.86)	$471.10	(325.72)*
Average Wage[a]	$9.28	(14.93)	$8.71	(5.50)	$15.40	(27.13)	$13.47	(12.08)**
Average Hours per Week	25.6	(11.6)	27.5	(11.7)***	35.1	(10.3)	35.4	(9.9)
Full Time	17.8%	(38.2)	21.9%	(41.3)***	46.7%	(50.0)	50.6%	(50.0)**
Any Employment	67.7%	(46.8)	70.7%	(45.6)*	70.9%	(45.4)	76.9%	(42.2)***
Age	18.3	(2.2)	18.7	(2.4)***	24.9	(2.6)	25.1	(2.6)***

[a] In 2010 dollars.
Standard deviations in parentheses.
* = Statistically different than public at the 0.1 level
** = Statistically different than public at the 0.05 level
*** = Statistically different than public at the 0.01 level

Table 3.12: Fixed-Effects Analysis: Returns to College Attendance, Log Weekly Earnings[a]

	OLS	**Individual FE**					
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
FP*Post	-0.021	-0.018	-0.007	0.009	-0.069		
	(0.035)	(0.033)	(0.033)	(0.034)	(0.044)		
Post	0.154***	0.075***	0.065**	0.031	0.130***	0.181***	0.131***
	(0.025)	(0.025)	(0.026)	(0.027)	(0.031)	(0.040)	(0.016)
Notes	Loaded OLS	Baseline FE	FE with flexible controls	Age 18 and over only	Dropping years in school	For-Profits Only	Publics Only
No. Obs.	16,796	16,796	16,796	14,831	12,271	2,115	14,681
No. Individuals	1,763	1,763	1,763	1,757	1,741	215	1,548

[a]This table is a replication of Cellini and Chaudhary (2012), Table 2A, p. 35.
Robust standard errors clustered at the individual level in parentheses.
* = Statistically significant at the 0.1 level
** = Statistically significant at the 0.05 level
*** = Statistically significant at the 0.01 level
All regressions include age and year fixed effects, and other than column (5) they also include and indicator variable for the years individuals report being in school. Column (1) is the fully loaded OLS including the same controls used by Cellini and Chaudhary (2012): race, gender, region, ability, foreign language and income. Missing data has been set to data with additional dummy variables indicating which variables were missing. Column (2) is the baseline individual fixed-effects model with age and year fixed-effects. Column (3) adds interactions of age with race, gender and region to the baseline model. Column (4) restricts the baseline model to age 18 and over. Column (5) drops the years in school. Column (6) includes only for-profit students and column (7) includes only public sectors students.

Table 3.13: Fixed-Effects Analysis: Returns to College Attendance and Completion, Log Weekly Earnings[a]

	OLS				**Individual FE**			
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FP*Post*Degree	0.127	0.120	0.094	0.134*	0.074	0.189***		0.049
	(0.084)	(0.077)	(0.076)	(0.078)	(0.106)	(0.060)		(0.064)
FP*Post	-0.067	-0.081*	-0.063	-0.059	-0.108**	0.106**		
	(0.043)	(0.042)	(0.041)	(0.043)	(0.053)	(0.050)		
Post*Degree	-0.004	0.063	0.071	0.054	0.055		0.071	-0.011
	(0.051)	(0.050)	(0.050)	(0.049)	(0.065)		(0.050)	(0.080)
Post	0.154***	0.070***	0.060**	0.028	0.127***		0.123***	
	(0.026)	(0.026)	(0.026)	(0.027)	(0.032)		(0.017)	
Notes	Loaded OLS	Baseline FE	FE with flexible controls	Age 18 and over only	Dropping years in school	For-Profits Only	Publics Only	Degree Completers Only
No. Obs.	16,796	16,796	16,796	14,831	12,271	2,115	14,681	2,565
No. Individuals	1,763	1,763	1,763	1,757	1,741	215	1,548	262

[a]This table is a replication of Cellini and Chaudhary (2012), Table 2B, p. 36.
Robust standard errors clustered at the individual level in parentheses.
* = Statistically significant at the 0.1 level
** = Statistically significant at the 0.05 level
*** = Statistically significant at the 0.01 level

All regressions include age and year fixed effects, and other than column (5) they also include and indicator variable for the years individuals report being in school. Column (1) is the fully loaded OLS including the same controls used by Cellini and Chaudhary (2012): race, gender, region, ability, foreign language and income. Missing data has been set to data with additional dummy variables indicating which variables were missing. Column (2) is the baseline individual fixed-effects model with age and year fixed-effects. Column (3) adds interactions of age with race, gender and region to the baseline model. Column (4) restricts the baseline model to age 18 and over. Column (5) drops the years in school. Column (6) includes only for-profit students and column (7) includes only public sectors students. Column (8) is my own addition to the table, including only degree-completers.

Table 3.14: Fixed-Effects Analysis: Effects of College Attendance on Other Labor Market Outcomes[a]

Variable	Log Weekly Earnings (1)	Log Hourly Wages (2)	Log Hours per Week (3)	Full-Time Employment (4)	Any Employment (5)
FP*Post	-0.018	-0.023	0.001	0.032	0.036
	(0.033)	(0.027)	(0.025)	(0.024)	(0.024)
Post	0.075***	0.042**	0.030*	0.012	0.010
	(0.025)	(0.019)	(0.018)	(0.017)	(0.017)
No. Obs.	16,796	16,816	16,997	22,579	22,579
No. Individuals	1,763	1,763	1,763	1,767	1,767

[a]This table is a replication of Cellini and Chaudhary (2012), Table 3A, p. 37.
Robust standard errors clustered at the individual level in parentheses.
* = Statistically significant at the 0.1 level
** = Statistically significant at the 0.05 level
*** = Statistically significant at the 0.01 level
All regressions include age and year fixed effects and an indicator variable for the years individuals report being in school.

Table 3.15: Fixed-Effects Analysis: Effects of College Attendance and Completion on Other Labor Market Outcomes[a]

Variable	Log Weekly Earnings (1)	Log Hourly Wages (2)	Log Hours per Week (3)	Full-Time Employment (4)	Any Employment (5)
FP*Post*Degree	0.120	0.059	0.031	0.068	0.021
	(0.077)	(0.065)	(0.058)	(0.055)	(0.054)
FP*Post	-0.081*	-0.054	-0.023	0.006	0.034
	(0.042)	(0.033)	(0.032)	(0.028)	(0.030)
Post*Degree	0.063	0.031	0.045	-0.000	-0.022
	(0.050)	(0.039)	(0.034)	(0.030)	(0.028)
Post	0.070***	0.039**	0.027	0.012	0.013
	(0.026)	(0.019)	(0.018)	(0.017)	(0.017)
No. Obs.	16,796	16,816	16,997	22,579	22,579
No. Individuals	1,763	1,763	1,763	1,767	1,767

[a]This table is a replication of Cellini and Chaudhary (2012), Table 3B, p. 38.
Robust standard errors clustered at the individual level in parentheses.
* = Statistically significant at the 0.1 level
** = Statistically significant at the 0.05 level
*** = Statistically significant at the 0.01 level
All regressions include age and year fixed effects and an indicator variable for the years individuals report being in school.

Table 3.16: Fixed-Effects Analysis: Returns to College Attendance by Major, Log Weekly Earnings

Variable		(1)	(2)
All Majors	**FP*Post**	-0.018	
		(0.033)	
	Post	0.075***	
		(0.025)	
Academic/ Other/ Unknown	**FP*Post**		0.049
			(0.064)
	Post		0.032
			(0.029)
Business-Related	**FP*Post**		-0.078
			(0.074)
	Post		0.141***
			(0.045)
Health-Related	**FP*Post**		-0.144*
			(0.079)
	Post		0.075*
			(0.043)
Vocational/ Technical	**FP*Post**		-0.018
			(0.057)
	Post		0.149***
			(0.038)
No. Obs.		16,796	16,796
No. Individuals		1,763	1,763

Robust standard errors clustered at the individual level in parentheses.
* = Statistically significant at the 0.1 level
** = Statistically significant at the 0.05 level
*** = Statistically significant at the 0.01 level
All regressions include age and year fixed effects and an indicator variable for the years individuals report being in school.

Table 3.17: Fixed-Effects Analysis: Returns to College Attendance and Completion by Major, Log Weekly Earnings

Variable		(1)	(2)
All Majors	**FP*Post*Degree**	0.120	
		(0.077)	
	FP*Post	-0.081*	
		(0.042)	
	Post*Degree	0.063	
		(0.050)	
	Post	0.070***	
		(0.026)	
Academic/ Other/ Unknown	**FP*Post*Degree**		-0.153
			(0.141)
	FP*Post		0.109
			(0.079)
	Post*Degree		-0.038
			(0.084)
	Post		0.039
			(0.030)
Business-Related	**FP*Post*Degree**		0.151
			(0.167)
	FP*Post		-0.153*
			(0.080)
	Post*Degree		0.041
			(0.107)
	Post		0.138***
			(0.048)
Health-Related	**FP*Post*Degree**		0.142
			(0.171)
	FP*Post		-0.242**
			(0.101)
	Post*Degree		0.190
			(0.118)
	Post		0.047
			(0.045)
Vocational/ Technical	**FP*Post*Degree**		0.106
			(0.119)
	FP*Post		-0.096
			(0.072)
	Post*Degree		0.145*
			(0.080)
	Post		0.131***
			(0.039)
No. Obs.		16,796	16,796
No. Individuals		1,763	1,763

Robust standard errors clustered at the individual level in parentheses.
* = Statistically significant at the 0.1 level
** = Statistically significant at the 0.05 level
*** = Statistically significant at the 0.01 level
All regressions include age and year fixed effects and an indicator variable for the years individuals report being in school.