UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays in Labor Economics and the Economics of Education

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Economics

by

Jaime Lynn Thomas

Committee in charge:

Professor Julie Cullen, Chair
Professor Julian Betts
Professor Nora Gordon
Professor Craig McIntosh
Professor Hugh "Bud" Mehan

2010

The Dissertation of Jaime Lynn Thomas is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

____

____

____

____

____
Chair

University of California, San Diego

2010

DEDICATION

For all the support and love he has shown me during my tenure as a graduate student, and for enabling me to achieve my dream of earning a Ph.D. by being a devoted husband and father, I dedicate this dissertation to Gregory Brian Thomas.

TABLE OF CONTENTS

Signature Page…………………………………………………………................. iii

Dedication……………………………………………………………………….. iv

Table of Contents………………………………………………………………... v

List of Tables……………………………………………………………………. viii

Acknowledgements……………………………………………………………... x

Vita... xi

Abstract of the Dissertation.. xii

Chapter 1.. 1

1.1 Introduction.. 2

1.2 Framework and Predictions.. 5

1.2.1 Four Labor-Market Knowledge Types.. 5

1.2.2 Positive Omitted Variable Bias and the Effect of Poor Labor-Market Knowledge.. 6

1.2.3 How Does Poor Labor-Market Information Affect Wages? The Role of Educational Attainment and Job Tenure.. 10

1.3 Empirical Methodology... 13

1.4 Data and Description of Variables.. 16

1.4.1 Primary Data Source... 16

1.4.2 Variables Used to Measure Labor-Market Knowledge....................... 16

1.4.3 Description of Student- and School-Level Variables.......................... 19

1.5 The Effect of Labor-Market Knowledge on Wages.............................. 23

1.5.1 Do Individuals End Up in Their Chosen Jobs?.................................. 23

1.5.2 Results Pertaining to Wages.. 24

1.5.3 The Role of Job Tenure and Educational Attainment.......................... 26

1.6 School Inputs Influencing Students' Information Capital............................ 28

1.7 Conclusions and Directions for Future Research.. 34

Appendix 1.1.. 50

Appendix 1.2.. 57

Appendix 1.3.. 71

References... 89

Chapter 2.. 93

2.1 Introduction... 94

2.2 Conceptual Framework.. 96

2.2.1 Class Type as an Input in an Education Production Function............. 96

2.2.2 The Combination-Class Treatment... 97

2.2.3 Confounding Factors Resulting from Nonrandom Assignment of Students and Teachers to Combination Classes... 99

2.3 Data... 100

2.4 Selection Issues... 101

2.4.1 School-Level Selection.. 101

2.4.2 Classroom-Level Selection.. 104

2.4.3 Student-Level Selection... 108

2.5 Results.. 110

2.6 Overall Impact... 113

2.7 Conclusion... 115

Appendix 2.1.. 128

References... 142

Chapter 3... 145

3.1 Introduction.. 146

3.2 Literature Review.. 147

3.3 Data, Household Sample, and School-Quality Variables............................. 150

3.4 Estimation Method and Results.. 153

3.5 Conclusion.. 157

References... 163

LIST OF TABLES

Table 1.1: Information Capital and Career Paths.. 36

Table 1.2: College and Noncollege Jobs.. 37

Table 1.3: Labor-Market Knowledge Types.. 38

Table 1.4: Means of Student and School Characteristics by Information-Capital Type.. 39

Table 1.5: Information Capital Predicts Holding a College Job............................... 42

Table 1.6: Effect of Labor-Market Knowledge on Wages, Comparing Overestimators to Noncollege-Track Students on Noncollege Jobs...... 43

Table 1.7: Effect of Labor-Market Knowledge on Wages, Comparing Underestimators to Noncollege-Track Students.................................... 44

Table 1.8: Labor-Market Information and Job Tenure, Comparing Overestimators to Noncollege-Track Students on Noncollege Jobs...... 45

Table 1.9: Labor-Market Knowledge and Educational Attainment, Comparing Overestimators to Noncollege-Track Students on Noncollege Jobs...... 46

Table 1.10: Labor-Market Information and Job Tenure, Comparing Underestimators to Noncollege-Track Students.................................... 47

Table 1.11: Labor-Market Knowledge and Educational Attainment, Comparing Underestimators to Noncollege-Track Students.................................... 48

Table 1.12: The Relationship Between School Inputs and Labor-Market Knowledge.. 49

Table 2.1: Types of First-Grade Classes Offered.. 117

Table 2.2: Comparison of Combination-Class Schools to Schools Offering Single-Grade Classes.. 118

Table 2.3: Means of Teacher Characteristics by Class Type................................... 119

Table 2.4: Modeling Teacher Selection.. 120

Table 2.5: Within-School Means of Classroom Characteristics by Class Type........ 121

Table 2.6: Means of Student Characteristics by Class Type.................................... 122

Table 2.7: Modeling Student Selection.. 124

Table 2.8: Four Outcome-Regression Models.. 125

Table 2.9: Coefficients of Interest from Four Outcome Regressions........................ 126

Table 2.10: Do 1-2 Students Benefit at the Expense of Other First Graders? Results from a Regression of Test Scores on a School-Type Dummy and School Characteristics... 127

Table 3.1: Ordinary Least Square Regression Results...... 159

Table 3.2: Conditional Logit Regression Results, Specification 1............................ 160

Table 3.3: Conditional Logit Regression Results, Specification 2............................ 161

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Julie Cullen for her indispensible research guidance. She is unfailingly generous with her time, and her comments and suggestions during our frequent meetings improved my research beyond what I thought possible. I would also like to acknowledge Professor Nora Gordon for her invaluable guidance and support, as well as Professor Julian Betts and the other members of my committee.

Professor Kate Antonovics kindly provided useful advice and criticism on Chapters 1 and 2 of this dissertation. I would also like to thank Professors Gordon Dahl and Prashant Bharadwaj for their helpful comments.

I would also like to acknowledge fellow Economics Department graduate students Ben Gilbert, Lindsay Oldenski, Scott Borger, and Jason Shafrin. Their support, especially during our first year, helped me immeasurably.

Finally, I would like to thank Director of Student Affairs Rafael Acevedo, Facilities and Research Support Staff Member Michael Bacci, and Graduate Student Affairs Officers Rebecca Franco and Mary Jane Hubbard for their superb handling of administrative affairs and for making the Economics Department such a pleasant place to be.

Chapter 2 has been submitted for publication of the material as it may appear in the *Economics of Education Review*. The dissertation author was the sole author of this paper.

VITA

1998	Bachelor of Arts, Linguistics, Stanford University
1998-2000	Assistant Language Teacher, Japan Exchange and Teaching Program, Fukiage Town, Japan
2001-2002	Clinician, Lindamood Bell Learning Processes, Solana Beach, California
2003	Research Assistant, The Graduate School of International Relations and Pacific Studies, University of California, San Diego
2003-2004	Teaching Assistant, The Graduate School of International Relations and Pacific Studies, University of California, San Diego
2004	Master of Pacific and International Affairs, The Graduate School of International Relations and Pacific Studies, University of California, San Diego
2005-2010	Teaching Assistant, Department of Economics, University of California, San Diego
2008-2009	Instructor, Department of Economics and The Graduate School of International Relations and Pacific Studies, University of California, San Diego
2010	Doctor of Philosophy, Department of Economics, University of California, San Diego

FIELDS OF STUDY

Major Field: Economics

Studies in the Economics of Education
Professors Julian Betts, Julie Cullen, and Nora Gordon

Studies in Labor Economics
Professors Kate Antonovics, Julie Cullen, and Nora Gordon

ABSTRACT OF THE DISSERTATION

Essays in Labor Economics and the Economics of Education

by

Jaime Lynn Thomas

Doctor of Philosophy in Economics

University of California, San Diego, 2010

Professor Julie Cullen, Chair

This dissertation addresses three broad issues within the fields of labor economics and the economics of education: the accumulation of human and information capital, school quality, and policy-relevant analysis of classroom organization. At the secondary-school level, I document the importance of information capital, or accurate information about postsecondary and labor-market alternatives. At the elementary-school level, I

analyze the effect of combination classes and discuss different ways to measure school quality and the importance of these measures to parents of school-aged children.

In the first chapter, "Information Capital and Early-Career Wages," I define one measure of information capital acquired by students during high school and develop a framework through which I analyze the effect of this measure on educational attainment, job tenure, and wages. I also investigate the school-level characteristics that influence an individual's stock of information capital.

In the second chapter, "Combination Classes and Educational Achievement," I measure the effect of membership in a combination class in first grade on student achievement. I address the selection that occurs when implementing a combination class and find that first graders in 1-2 combinations can be expected to outperform single-grade students on math tests by one-seventh of a standard deviation. In addition, I find no evidence that first graders in schools offering combination classes perform worse than first graders in schools that do not offer such classes. Therefore, I conclude that combination classes may be a Pareto-improving option for school administrators.

In the last chapter, "Neighborhood Demographics, School Effectiveness, and Residential Location Choice," I investigate how neighborhood demographics and school effectiveness influence the residential location decisions of parents of different income levels. I find that low-income parents in the San Francisco Bay Area respond more strongly to school effectiveness than to neighborhood demographics, but that the reverse is true for high-income parents.

CHAPTER 1

INFORMATION CAPITAL AND EARLY-CAREER WAGES

Abstract: Traditional human capital theory posits that the larger the stock of a worker's human capital, the more productive the worker will be and the more the worker will earn. Information capital, the knowledge that individuals possess about the labor market and about their aptitudes and tastes for different levels of education and types of employment, is another component of an individual's skill set that affects productivity and wages. In this paper, I define one measure of information capital: labor-market knowledge captured by 12th graders' understanding of the educational requirements of the jobs they hope to hold at age 30. I demonstrate that inaccurate labor-market information affects wages through decreased job tenure, driven by individuals entering and leaving postsecondary school as they come to an accurate understanding of the educational requirements of their chosen jobs. I find that poor labor-market knowledge affects workers well into their twenties: despite having higher grades and test scores, workers who were mistaken about educational requirements in high school earn wages no higher than workers with accurate information. I also investigate the role of high school guidance counselors and vocational education faculty in students' information-capital acquisition, and show that schools can influence students' career aspirations and labor-market knowledge.

1.1 Introduction

Traditional human capital theory posits that the larger the stock of a worker's human capital, the more productive the worker will be and the more the worker will earn. Human capital refers to the skills and knowledge an individual acquires through education and experience, as well as the individual's innate abilities and values. The theory distinguishes between general and specific human capital. To this list should be added information capital: the knowledge that workers possess about the labor market and about their aptitudes and tastes for different levels of education and types of employment. Information capital is another component of an individual's skill set that affects productivity and wages.

In this paper, I define one measure of information capital: labor-market knowledge captured by 12^{th} graders' understanding of the educational requirements of the jobs they hope to hold at age 30. I demonstrate that inaccurate labor-market information affects wages through decreased job tenure, driven by individuals entering and leaving postsecondary school as they come to an accurate understanding of the educational requirements of the jobs they wish to hold. I find that poor labor-market knowledge affects workers well into their twenties: despite having higher grades and test scores, workers who were mistaken about educational requirements in high school earn wages no higher than workers in similar jobs who were not. I also investigate the role of high school guidance counselors and vocational education faculty in students' information-capital acquisition, and show that schools can influence students' aspirations and labor-market knowledge.

This paper makes several contributions. First, I demonstrate the importance of information capital for both educational attainment and wages. Few studies have explored the effect of labor-market knowledge on educational attainment.[1] Ludwig (1999) focuses on inner-city youth and uses two measures of labor-market information: individuals' understanding of the job duties associated with nine different occupations and the difference between the average education level in a respondent's reported occupational goal and his or her reported educational aspirations. He finds that those with better information are more likely to graduate from high school. I extend this by documenting a link between labor-market information and postsecondary attainment.

A small body of literature examines the relationship between wages and information capital as measured by an individual's score on the Knowledge of World of Work (KWW) test and wages. KWW measures respondents' knowledge of the labor market by asking them about the duties, educational requirements, and relative earnings of ten occupations. Blackburn and Neumark (1992) find that labor-market knowledge does not explain inter-industry and inter-occupation wage differentials, but Polachek and Robst (1999) conclude that workers with better labor-market knowledge earn a larger proportion of their potential wages. I focus on labor-market knowledge relevant to the jobs students wish to hold and the effect of this labor-market knowledge on wages early in the individuals' careers.

Another contribution is to show that one channel through which inaccurate labor-market information affects wages is decreased job tenure. Farber (1999) investigates the

---

[1] A more substantial and currently quite active body of literature addresses the role of perceived returns to education in determining educational outcomes. See Manski (1989), Kaufmann and Attanasio (2009), Jensen (forthcoming), and Nguyen (2008).

roles of firm-specific capital and worker heterogeneity in mobility rates in explaining the following facts concerning job tenure in the United States: long-term employment relationships are common, most new jobs end early, and the probability of a job ending declines with tenure. I provide evidence that one facet of worker heterogeneity, poor labor-market information, is negatively related to job tenure early in workers' careers.

The other contributions of this paper are to present information capital as a novel output of an education production function and to provide preliminary evidence suggesting that schools can influence information-capital acquisition. An extensive body of literature considers measures of human capital such as student achievement as outputs produced by school inputs such as class size and teacher quality. Though a positive relationship between school resources and student achievement has been documented,[2] the debate over how specific school inputs affect achievement continues.[3] I consider the relationship between information capital and school inputs aimed directly at influencing students' career aspirations and labor-market knowledge—guidance counselors and vocational education faculty.

This paper proceeds as follows. Section 1.2 lays out the framework of this study and describes the predictions about the effects of information capital on wages via educational attainment and job tenure. Section 1.3 describes the empirical methodology. Section 1.4 describes the primary data source and the student- and school-level variables used throughout the analysis. Section 1.5 contains the results of regressions measuring the effect of labor-market knowledge on wages and discusses the role of job tenure and

---

[2] See, for example, Card and Payne (2002) or Rivkin, Hanushek, and Kain (2005).
[3] Alan Krueger and Erik Hanushek debate the effect of class size on student achievement in Mishel and Rothstein, eds. (2002); Rivkin, Hanushek, and Kain (2005) discuss the effects of teachers on reading and math achievement.

educational attainment. In Section 1.6, I analyze the relationship between information capital and school inputs such as guidance counselors and vocational faculty. Section 1.7 concludes.

1.2 Framework and Predictions

1.2.1 Four Labor-Market Knowledge Types

Educational attainment and wages are tied to students' career aspirations and their understanding of the educational requirements of their chosen careers. My measure of information capital comprises two components: a student's professional aspirations and her understanding of the educational requirements of her chosen job.

I classify jobs into two types: "college jobs" require individuals to hold a four-year college degree; "noncollege jobs" do not. In this simple framework, only those who graduate from a four-year institution can hold college jobs; anyone can hold a noncollege job.

In 12^{th} grade, students state their professional aspirations—college job or noncollege job—without necessarily understanding the job's educational requirements. At this time, they also declare how much education they believe is required for the job they wish to hold: a four-year college degree, or less than a college degree.

Combining these two dichotomous measures gives rise to four types of students. Students who are "not on the college track" aspire to a noncollege job and correctly believe that a college degree is not required for this job—their path to a noncollege job is straightforward. Students "on the college track," on the other hand, aspire to a college job and correctly believe that a college degree is required—their path to a college job is straightforward.

Students who overestimate the educational requirements of jobs aspire to a noncollege job but believe that a college degree is required. These "overestimators" have inaccurate labor-market information since their career goals and understanding of educational requirements do not line up, but these students do not close any doors for themselves by thinking college is required when it is not—college graduates can still hold noncollege jobs. Their job path, however, is not straightforward.

Students who underestimate educational requirements aspire to a college job but believe that a college degree is not required. In this simple framework, holding a college job without a college degree is impossible—the misalignment of these students' career goals and perceived educational requirements precludes them from attaining their chosen jobs. Unlike overestimators, "underestimators" face a barrier to achieving their professional goals. Like overestimators, their job path is not straightforward.

In the next subsection, I outline how to overcome one form of omitted variable bias in order to isolate the causal effect of poor labor-market knowledge on wages.

1.2.2 Positive Omitted Variable Bias and the Effect of Poor Labor-Market Knowledge

Suppose wages can be predicted according to the following reduced-form model:

$$y_i = X_i\beta + M_i\gamma + \alpha_i + \varepsilon_i, \qquad (1)$$

y_i is log hourly wages for individual i, X_i represents observable characteristics that affect wages such as ability, motivation, family background, and risk and rate-of-time preferences M_i is a dummy variable indicating that *i* is misinformed; that is, that *i* is either an overestimator or an underestimator. α_i represents unobservable characteristics, and ε_i is a mean-zero error.

Let $\bar{M}_1 = E\left(M_i \mid X_i, i \text{ is not on the college track}\right)$, $\bar{M}_2 = E\left(M_i \mid X_i, i \text{ is an overestimator}\right)$, $\bar{M}_3 = E\left(M_i \mid X_i, i \text{ is an underestimator}\right)$, and $\bar{M}_4 = E\left(M_i \mid X_i, i \text{ is on the college track}\right)$. $\bar{M}_1 = \bar{M}_4 = 0$ because these types have accurate information, and $\bar{M}_2 = \bar{M}_3 = 1$ since these types are misinformed. Now, let $\Delta\bar{M}_{2-1} = \bar{M}_2 - \bar{M}_1$, and $\Delta\bar{M}_{3-1} = \bar{M}_3 - \bar{M}_1$. Note that $\Delta\bar{M}_{2-1} = \Delta\bar{M}_{3-1} = 1$. Thus, γ can be interpreted as the effect of being misinformed.

In addition, let $\bar{\alpha}_1 = E\left(\alpha_i \mid X_i, i \text{ is not on the college track}\right)$, $\bar{\alpha}_2 = E\left(\alpha_i \mid X_i, i \text{ is an overestimator}\right)$, $\bar{\alpha}_3 = E\left(\alpha_i \mid X_i, i \text{ is an underestimator}\right)$, and $\bar{\alpha}_4 = E\left(\alpha_i \mid X_i, i \text{ is on the college track}\right)$. Finally, $\Delta\bar{\alpha}_{2-1} = \bar{\alpha}_2 - \bar{\alpha}_1$, and $\Delta\bar{\alpha}_{3-1} = \bar{\alpha}_3 - \bar{\alpha}_1$. $\Delta\bar{\alpha}_{2-1}$ and $\Delta\bar{\alpha}_{3-1}$ represent omitted-variable bias.

I will isolate the causal effect of labor-market knowledge by comparing outcomes across types. The expected difference in outcomes between group *j* and group *k* is given by

$$E\left(\Delta y_{j-k} \mid X\right) = \Delta\bar{M}_{j-k}\gamma + \Delta\bar{\alpha}_{j-k}. \qquad (2)$$

First, consider the comparison between overestimators (Type 2) and noncollege-track students (Type 1). These students share a career aspiration—neither type wants a college job—but the overestimators incorrectly believe that college is necessary. The expected difference in outcomes, given the X variables which are common to all types, is given by

$$E\left(\Delta y_{2-1} \mid X\right) = \Delta \bar{M}_{2-1}\gamma + \Delta\bar{\alpha}_{2-1} = \gamma + \Delta\bar{\alpha}_{2-1} \tag{3}$$

In Section 1.4, I show that students who believe that overestimators are higher-achieving and of higher socioeconomic status (SES) than noncollege-track students. Since it is likely that overestimators are positively selected on unobservables as well, $\Delta\bar{\alpha}_{2-1}$ should be positive.

My hypothesis is that γ is negative. Though overestimators are not closing any doors for themselves with their lack of understanding of the educational requirements of their chosen jobs, they do possess inaccurate labor-market information. Controlling for observable differences, if I find that overestimators have wages no greater than those of noncollege-track students, then inaccurate labor-market information outweighs any positive selection and I can conclude that poor labor-market information has a negative effect on wages.

Now, consider the comparison between underestimators and noncollege-track students. The expected difference in outcomes is

$$E\left(\Delta y_{3-1} \mid X\right) = \Delta \bar{M}_{3-1}\gamma + \Delta\bar{\alpha}_{3-1} = \gamma + \Delta\bar{\alpha}_{3-1} \tag{4}$$

Neither of these thinks that a college degree is necessary, but underestimators are incorrect in this belief since they want a college job. In Section IV, I show that underestimators have higher grades and test scores than noncollege-track students. Thus $\Delta\bar{\alpha}_{3-1}$ is likely to be positive—underestimators are likely to be positively selected on unobservables as well as observables.

I hypothesize that γ is negative since underestimators have inaccurate labor-market information. Controlling for observables, if I find that underestimators earn wages no greater than noncollege-track students, I can again conclude that the negative effect of inaccurate labor-market information outweighs any positive omitted variable bias.

The other possible comparisons do not allow me to draw meaningful conclusions about the effect of inaccurate labor-market knowledge on wages. Both noncollege- and college-track students have accurate labor-market information (i.e., $\Delta\overline{M}_{4-1} = 0 - 0 = 0$), and both overestimators and underestimators have inaccurate labor-market information (i.e., $\Delta\overline{M}_{3-2} = 1 - 1 = 0$). Comparing college-track students to underestimators or to overestimators, the omitted variables bias works in the same direction as any positive effect of accurate information. That is, since college-track students are positively selected relative to over- and underestimators, $\Delta\overline{\alpha}_{4-2}$ and $\Delta\overline{\alpha}_{4-3}$ are positive, and since college-track students have accurate labor-market information, the γs should be positive as well. Thus, finding that college-track students earn more than underestimators or overestimators tells us nothing about the effect of accurate labor-market knowledge since the positive omitted variables bias reinforces any positive effect of accurate information.

In the next subsection, I discuss possible mechanisms through which poor labor-market information affects wages.

1.2.3 How Does Poor Labor-Market Information Affect Wages? The Role of Educational Attainment and Job Tenure

In the framework I outline in this section, I focus on a subset of causal mechanisms through which inaccurate information affects wages. I show that inaccurate information can lead to decreased job tenure because of time spent in nonproductive education. That is, over- and underestimators make more missteps by entering and leaving postsecondary school as they come to an accurate understanding of the educational requirements of the jobs they wish to hold. This framework does not consider the very real possibility that more education could lead to more job tenure and higher wages due to returns to some college and/or differential exposure to unemployment. The predictions that this framework yields only apply if the negative effect of missteps dominates any positive effects of education. In Section 1.5, I show that the data do, in fact, bear out these predictions.

I will illustrate the possible job paths for students of each labor-market knowledge type. Consider a simple, three-period example. Life does not end at $t=3$, but wages will be measured at that time. Acquiring a college degree takes two periods—if a worker decides to (re-) join the labor force after attending college for only one period, that worker must work in the noncollege job. While in college, students earn 0. For new hires, the college job pays w^C and the noncollege job pays w^N, with $w^C > w^N$. This assumption is in line with the literature on the return to a college degree. For example, plotting age-earnings profiles using CPS data, Card (1999) shows that at zero years of experience, college-educated workers earn higher wages than high school graduates.

Altonji and Williams (2005) and others show that job tenure has a positive effect on wages. In my framework, wages rise with job tenure at a rate ρ per period. At $t=0$, all students graduate from high school.

In this simple framework, once an individual's job choice and educational requirements are aligned, that individual does not switch jobs. In addition, immediately out of high school, individuals pursue a path dictated by their perceived educational requirements. Thus, students not on the college track head straight for the noncollege job and do not switch, and students on the college track head straight for the college job via college. Overestimators and underestimators, however, face circuitous paths to their chosen jobs. While in practice, all groups will be learning about their preferences for jobs and education, misinformed individuals have more to learn about the constraints they may face in attaining their chosen jobs. Thus the relative rates of missteps should match the predictions in this section.

Table 1.1 illustrates all possible career paths and gives the wages earned by each type of worker in each time period. Students on the noncollege track work for three periods. At the end of the third period, these students earn $w^N(1+\rho)^3$. Students on the college track attend college for two periods and work in the college job for one period. At the end of three periods, these students earn $w^C(1+\rho)$.

Overestimators believe a college degree is required so they attend college immediately after graduating from high school. From here, their paths diverge. Some overestimators realize that a college degree is not required for the noncollege job, and join the labor force in the second period. They work for two periods and at the end of this time, earn $w^N(1+\rho)^2$. Some overestimators re-evaluate their career goals and decide they want a college job, finish college, and earn $w^C(1+\rho)$ in the last period.

Underestimators do not believe a college degree is required, so they work the period immediately after high school in the noncollege job, the only one open to them.[4] From there, they can follow one of three paths. Some re-evaluate their career goals and stick with the noncollege job. At the end of period three, these workers earn $w^N(1+\rho)^3$. Others realize that a college degree is required for the college job, and decide to go to school. After attending school for one year, some of these decide that schooling is too costly and return to the noncollege job. These workers earn $w^N(1+\rho)^2$ at the end of the last period. Finally, some underestimators stay in school for two periods. These earn 0 at the end of the third period.

This framework predicts that overestimators who end up on the noncollege job earn lower wages than noncollege-track individuals on the same job because they have accumulated less job tenure. They have lower job tenure because they spent some time in nonproductive education before re-evaluating their career plans, so I can also predict that overestimators—even those on the noncollege job—will have higher educational attainment than noncollege-track students.

Turning to the comparison between underestimators and noncollege-track students, noncollege-track students have accumulated at least as much job tenure in the noncollege job as underestimators, because some underestimators give college a try

[4] To keep this framework as simple as possible while preserving its usefulness in understanding the roles of educational attainment and job tenure, I ignore the possibility of voluntary unemployment due to w^N being less than the individual's reservation wage.

before returning to the noncollege job, and some stay in school. Thus, the framework in this section predicts that, on average, underestimators will earn wages less than or equal to the wages of students not on the college track but will have (weakly) more educational attainment and (weakly) less job tenure.

Section 1.3 describes the empirical methodology I use to test these predictions.

1.3 Empirical Methodology

In this section, I describe the empirical models I employ in order to measure the effect of labor-market knowledge on wages and test the predictions from Section 1.2. First, I compare overestimators to noncollege-track students. These students share a job aspiration in 12th grade—both want a noncollege job. Section 1.2 predicts that overestimators who end up in the noncollege job will earn lower wages than noncollege-track students in the same job, and have higher educational attainment and lower job tenure. Thus I restrict the sample to individuals who desired a noncollege job in 12th grade and ended up at a noncollege job in their mid-twenties. (I will describe this method of job classification in more detail in Section 1.5). I estimate the following:

$$y_{is} = \alpha + \beta_2 T_{2is} + \lambda X_{is} + \delta_s + \varepsilon_{is}\ , \qquad (5)$$

where y_{is} is the outcome of interest for student *i* at school *s*—log hourly wage, educational attainment, or job tenure. T_{2is} is a dummy denoting a student who does not want a college job but thinks the job requires a college degree—an overestimator. (Since I restrict the sample to those students not desiring a college job in 12th grade, T_{1is}—denoting noncollege-track students—is the omitted category).

Next, I compare underestimators to noncollege-track students. These share a belief that a college degree is not required for their chosen jobs. Section 1.2 predicts that underestimators will earn (weakly) lower wages than noncollege-track students, have (weakly) more educational attainment, and have (weakly) lower job tenure. This prediction does not depend on the type of job the individuals hold in their mid-twenties. I restrict the sample to the two types of interest and estimate

$$y_{is} = \alpha + \beta_3 T_{3is} + \lambda X_{is} + \delta_s + \varepsilon_{is}, \qquad (6)$$

where y_{is} defined as in (5), and T_{3is} denotes underestimators, with noncollege-track students as the omitted category. In both (5) and (6), X_{is} is a vector of student characteristics.

I am interested in the causal interpretation of β_2 and β_3, which tell me the effect of being an overestimator and an underestimator, respectively, on the outcome of interest, relative to noncollege-track students. Here, it is important to note that focusing on comparisons between over- or underestimators and noncollege-track individuals allows me to isolate the causal effect of poor labor-market knowledge on wages because the negative misinformation effect moves in the opposite direction of the positive omitted variable bias. To the extent, however, that educational attainment is positively associated with any omitted variables, and in turn is negatively associated with job tenure, I am not able to ascribe a causal interpretation to the β s in regressions with educational attainment and job tenure as outcomes.

In addition to the positive omitted variable bias described in Section 1.2, another barrier to causal interpretation of the β s is the possibility that poor labor-market information is a proxy for variables such as “flakiness” or poor estimation ability that may negatively affect wages. In order to address this source of bias, I use several different specifications of the models in (5) and (6), adding more and more variables to X_{is} each time. In particular, I add variables measuring noncognitive traits and risk and rate-of-time preferences in order to ensure that my information-capital measure is not just a proxy for undesirable individual characteristics.

Unobserved school characteristics can also be a source of omitted variable bias. I address this source of bias by including school fixed effects, δ_s. Thus I am measuring within-school differences in wages as a function of information-capital type and student characteristics.

Because type is determined in high school, labor-market type coefficients are difficult to interpret if I include student-level measures also determined in high school. This is because high school performance and participation measures such as test scores, grades, and participation in extracurricular activities may be codetermined with information-capital type. For example, if I receive poor grades, I may decide that a college job is not for me. Conversely, if I decide a college job is not for me, I may put forth less effort in school and earn lower grades. Or, if I think college is required for my job, I may put forth more effort and earn higher grades, or decide to participate in extracurricular activities. Thus I exclude variables codetermined with information-capital type because they prevent meaningful interpretation of the β s in (5) and (6).

12th grade standardized test scores, however, are arguably not codetermined with students’ information-capital type. Because they are not observed by future employers or college admissions committees, scores on these tests depend less on students' motivation and career and educational aspirations than do grades and participation in extracurricular activities. In order to account for experiences in high school affecting postsecondary and labor-market outcomes through ability but not through career aspirations or labor-market knowledge, I include 12th grade standardized test scores in one of the specifications in Section 1.5.

Section 1.4 contains a description of my primary data source and a detailed description of the variables used throughout the analysis.

1.4 Data and Description of Variables

1.4.1 Primary Data Source

My primary data source is the National Education Longitudinal Study of 1988 (NELS), a nationally representative sample of 27,805 eighth-grade students interviewed in 1988. Follow-ups took place in 1990 (when most were in 10th grade), 1992 (12th grade), 1994, and 2000 (when the respondents' average age was 26). The study contains data from detailed student, parent, and school administrator questionnaires, as well as high school transcript data and information on postsecondary and labor-market outcomes.

1.4.2 Variables Used to Measure Labor-Market Knowledge

In order to measure labor-market knowledge, I classify students into types based on their answers to two 1992 survey questions. The first asks about job goals: "Which of the categories below comes closest to describing the job or occupation that you expect or plan to have … when you are 30 years old?" I classify jobs into college and noncollege jobs by mapping detailed occupations from the March 1992 CPS to the jobs listed in the 1992 NELS survey. (Please see Appendix 1.1 for this mapping). If at least 60 percent of the individuals in a job have a bachelor's degree or more according to the CPS, I classify that job as a college job. Table 1.2 contains these job classifications.

As the other component of my labor-market knowledge measure, I consider students' responses to the question, "How much education do you think you need to get the job you expect or plan to have when you are 30 years old?" which immediately follows the question on career aspirations in the 1992 NELS survey. If students answer "4 or 5 year college degree" or more, I classify them as perceiving that their chosen job requires a college degree.

Combining the answers to these two questions, I construct a measure of labor-market knowledge that can take four values. Table 1.3 shows that 23.8 percent of the respondents are not on the college track, 22.0 percent are overestimators, 5.5 percent are underestimators, and 48.7 percent are on the college track.[5]

A potential criticism of this binary classification of jobs is that I may have misclassified a number of students. For example, I classify Bill, a student who says he wants to be a "Professional (e.g., accountant, registered nurse, engineer)" but who does not plan on graduating from college, as an underestimator. Bill may, in fact, plan to be a registered nurse and attain this goal by attending two years of nursing school after high school. Thus he has correctly estimated the educational requirements of his chosen job, and should be classified as a student not on the college track rather than an underestimator.

[5] 15,511 students have nonmissing observations for this measure.

In order to address this criticism, I have repeated the analysis in Section 1.5 with continuous measures of labor-market knowledge which I briefly describe here. Instead of assigning each student to a type, I assign each student a probability of being in each type as follows. First, define

$$c_i = \begin{cases} 0 \text{ if student } i \text{ thinks a college degree is not required} \\ 1 \text{ if student } i \text{ thinks a college degree is required} \end{cases}. \quad (7)$$

Then, let p_i be the percent of individuals in the U.S. in student *i*'s chosen job with a B.A. or more (taken from the March 1992 CPS). Now, the probability of being in each type is given by

$$\begin{aligned} &\Pr(\text{Noncollege track})_i = (1-c_i)(1-p_i) \\ &\Pr(\text{Overestimator})_i = c_i(1-p_i) \\ &\Pr(\text{Underestimator})_i = (1-c_i)(p_i) \\ &\Pr(\text{College track})_i = c_i p_i \end{aligned} \quad . \quad (8)$$

For example, instead of being unequivocally placed into category 3 as an underestimator, Bill (from the example above) would receive the following values:

$$\begin{aligned} &\Pr(\text{Noncollege track})_{Bill} = 1(1-0.66) = 0.34 \\ &\Pr(\text{Overestimator})_{Bill} = 0(1-0.66) = 0 \\ &\Pr(\text{Underestimator})_{Bill} = 1(0.66) = 0.66 \\ &\Pr(\text{College track})_{Bill} = 0(0.66) = 0 \end{aligned} \quad . \quad (9)$$

Since 66 percent of individuals in the "Professional (e.g., accountant, registered nurse, engineer)" category have a college degree according to the CPS, and since Bill does not think college is required for his job, he has a 34 percent chance of being correct and a 66 percent chance of being incorrect. In other words, he has a 34 percent chance of being a student not on the college track with accurate labor-market knowledge, and a 66 percent chance of being an underestimator.

Since my results in Section 1.5 are not sensitive to using these continuous measures (see Appendix 1.2 for the continuous results), I choose to describe and report results for the discrete measure because it is consistent with my framework in Section 1.2 and because it lends itself to more natural discussion and interpretation.

1.4.3. Description of Student- and School-Level Variables

Recall from (5) and (6) that I regress the outcome of interest on labor-market knowledge dummies and student characteristics. In order to show that my information capital measure is not a proxy for unobserved abilities or skills, I use several different specifications of the models in (5) and (6), adding successively more covariates from one to the next. I do this by partitioning X , the vector of student characteristics, into four different groups of variables: $X1$, $X2$, $X3$, and $X4$.

$X1$ contains the following eighth-grade academic ability, achievement, and coursetaking measures: a reading and math standardized test score composite, GPA, reading, math, and science proficiency measures, a dummy variable indicating whether a student was held back in a grade prior to eighth, and a dummy variable indicating that the student took algebra in eighth grade. According to the 2008 Brown Center Report on American Education, during the 1990s and the 2000s, the percentage of American eighth graders taking algebra has nearly doubled. The impetus for this increase came during the Clinton Administration which made universal eighth grade algebra a national goal in order to enable students to succeed in higher-level math courses in high school. Thus, taking algebra in eighth grade is an important predictor of academic orientation and readiness for more advanced high school math courses.[6]

$X1$ also contains two noncognitive or personality-trait measures: locus of control and self-concept. These measures have begun to receive attention in the economics literature as important influences on schooling decisions and wages (see, for example, Heckman, Stixrud, and Urzua, 2006). Students' answers in the eighth-grade survey to six questions eliciting the degree to which they feel they can control what happens to them are used to construct the locus of control score.[7] The higher the score, the more the student feels he can control events. Answers to seven questions on students' feelings of worthiness or self-esteem are used to construct the self-concept score,[8] with a higher score indicating more self-esteem.

$X2$ contains basic student and family characteristics: age, gender, race and ethnicity, and family SES. Family SES is a composite of father's and/or mother's education level, father's and/or mother's occupation, and family income.

$X3$ contains variables measuring students' household environments and risk and rate-of-time preferences: dummies indicating that a student's home language is non-

[6] Though there is some evidence that students who take algebra in eighth grade outperform other students, recent research calls into question the value of eighth grade algebra for under-prepared students (2008 Brown Center Report on American Education). It is also important to note that offering eighth-grade algebra reflects a school's resources, not just an individual student's academic orientation or high-school readiness.
[7] Students respond to the following statements by choosing from a four-point Likert Scale (strongly agree, agree, disagree, strongly disagree): "I don't have enough control over the direction my life is taking," "In my life, good luck is more important than hard work for success," "Every time I try to get ahead, something or somebody stops me," "My plans hardly ever work out, so planning only makes me unhappy," "When I make plans, I am almost certain I can make them work," and "Chance and luck are very important for what happens in my life."
[8] The self-concept statements are as follows: "I feel good about myself," "I feel I am a person of worth, the equal of other people," "I am able to do things as well as most other people," "On the whole, I am satisfied with myself," "I certainly feel useless at times," "At times, I think I am no good at all," and "I feel I do not have much to be proud of."

English only/non-English dominant, that the student lived in a single-parent household in eighth grade, that the student often discussed his or her studies with parents in eighth grade, and that the student smoked in eighth grade. I include the smoking dummy to capture risk and time preferences, since smoking has been linked to both high discount rates and low levels of risk aversion; see, for example, Fersterer and Winter-Ebmer (2003) and Ida and Goto (2009). $X4$ contains just one variable: a 12th grade reading and math standardized test score composite.

In Section 1.5, I use five different specifications of (5) and (6). The first contains the relevant information-capital dummy and school fixed effects as the only right-hand-side variables. Specification (2) includes these as well as $X1$, specification (3) adds $X2$, specification (4) adds $X3$, and specification (5) adds $X4$.

The school-level variables I analyze in Section 1.6 are the number of guidance counselors in student *i*'s school divided by tenth grade enrollment in the school, the number of vocational education faculty divided by tenth grade enrollment,[9] whether the school is a vocational school, the number of AP courses offered, the percent of the previous year's graduates attending 2- and 4-year schools, tenth grade enrollment, student-teacher ratio, percent non-White, percent receiving free or reduced-price lunch, dummies for urban, suburban, or rural location, and regional dummies (Northeast, North Central, South, and West).[10]

---

[9] The tenth-grade (1990) NELS survey reports the number of guidance counselors and vocational education faculty as categorical variables with 1 = none, 2 = 1-5, 3 = 6-10, 4 = 11-15, and 5 = over 15. I assign the value 0 to category 1, 3 to category 2, 8 to category 3, 13 to category 4, and 15 to category 5.

[10] Because the school administrator surveys from students' tenth grade year had much higher response rates than the 12th grade surveys, I use tenth grade school-level measures.

Table 1.4 contains the means of each of these variables by labor-market-knowledge type. In terms of eighth-grade GPA, reading, math, and science proficiency, and eighth- and 12th-grade standardized test scores, students on the college track are the highest achieving, followed by overestimators, then underestimators, and finally, students not on the college track. The same pattern holds for SES, though the difference between underestimators and noncollege-track students is not statistically significant. Considering the predictions on wages, it is interesting to note that both underestimators and overestimators are higher-achieving than noncollege-track students, and in addition, overestimators are of significantly higher SES, even though these two types are predicted to have wages no greater than the noncollege-track students (at least on the noncollege job).

Table 1.4 also shows the importance of using school fixed effects to estimate the effect of information-capital type on wages. As measured by variables such as the percent of the previous year's graduates attending four-year colleges and student-teacher ratio, college-track students and overestimators appear to attend schools that have more resources and are more academically oriented than schools attended by noncollege-track students and underestimators. If schools differ on unobservable characteristics as well, merely including these school-level variables as covariates in the wage regressions will not eliminate omitted variable bias. For this reason, I include school fixed effects in the regressions discussed in the next section.

1.5 The Effect of Labor-Market Knowledge on Wages

In this section, I demonstrate that poor labor-market knowledge has a negative effect on wages, and this effect appears to operate through educational attainment and job tenure as outlined in Section 1.2. First, since some of the predictions from Section 1.2 depend on the job an individual holds in her mid-twenties, I show that information-capital type is an important predictor of job type. I then present the results on wages, job tenure, and educational attainment. In the tables in this section, I report only the coefficients on the information-capital dummies. Appendix 1.3 contains full regression results.

1.5.1 Do Individuals End Up in Their Chosen Jobs?

In this subsection, I analyze whether information-capital type predicts holding a college job in one's mid-twenties. I first classify the jobs that employed respondents report in the 2000 survey as college or noncollege jobs. Recall that, in order to classify 12th grade job aspirations, I map detailed occupations from the March 1992 CPS to the jobs listed in the 1992 survey, classifying as a "college job" one in which at least 60 percent of individuals have a bachelor's degree or more. I use the same method to classify the respondents' jobs as of 2000, but the job categories given in the 2000 survey are much harder to link to CPS job categories. Thus I also calculate the percent of individuals within each job with a B.A. or above according to the 2000 NELS survey. If I use a 55 percent cutoff with the latter method, discrepancies between the two methods are minimized. Appendix 1.1 contains these mappings.

Table 1.5 contains these results. I use the same methodology discussed in Section III, equations (5) and (6), except I include the full set of information-capital dummies: T_1

denotes noncollege-track students, T_2 denotes overestimators, T_3 denotes underestimators. T_4, college-track students, is the omitted category. The dependent variable is a dummy that equals one if the individual holds a college job in 2000. Linear probability results are reported; probit results show the same signs and significance levels.

Table 1.5 reveals that students on the college track in 12th grade are the most likely to hold a college job in their mid-twenties, followed by overestimators. According to unreported F-tests, overestimators are significantly more likely to hold a college job than noncollege-track students in all of the specifications. Though point estimates suggest that overestimators are more likely than underestimators to hold a college job, coefficients are not significantly different in specifications (3) through (5). Underestimators are no more or less likely to hold a college job than noncollege-track students—the coefficients on Type 3 and Type 1 are not significantly different in any of the specifications.

1.5.2 Results Pertaining to Wages

Now that I have demonstrated that information-capital type is an important predictor of the job an individual will hold in his mid-twenties, I now analyze its effect on wages. Table 1.6 compares overestimators to noncollege-track students, as in equation (5). The dependent variable is log hourly wage in 2000, when the average age of respondents is 26. OLS results are reported. I restrict the sample to students who aspired to a noncollege job in 12th grade and who were employed in a noncollege job in 2000. When no controls are included, overestimators appear to have higher wages than noncollege-track students—this is consistent with my statements in Section 1.2 that overestimators are positively selected relative to noncollege-track students on both observables and unobservables.

Once controls are added, point estimates are consistent with the predictions of Section 1.2: overestimators earn lower wages than noncollege-track individuals on the noncollege job. This difference is not significant at conventional levels in specifications (2) through (4), but it has economic significance. Relative to students who had the same career aspirations and an accurate understanding of educational requirements in high school, individuals who overestimated educational requirements earn approximately 70 cents less per hour, translating to an annual income difference of approximately $1400.[11]

Table 1.7 contains the results from the comparison between underestimators and noncollege-track students.[12] Point estimates accord with the predictions in Section 1.2: underestimators earn wages no higher than noncollege-track students, though the coefficient on the Type 3 dummy is not significantly different from zero in specifications (3) and (4). The lack of statistical significance belies the economic significance of this difference. Underestimators can be expected to earn approximately $1.50 per hour less than students not on the college track, an annual difference of approximately $3000.

I do not include a correction term to control for nonrandom selection into the labor force in the regressions reported in Tables 1.6 and 1.7. This is because information-

[11] If I do not condition on job type in 2000, the coefficient on the overestimator dummy is positive but very close to zero, and not significant. This coefficient suggests an hourly difference in wages of approximately three cents, and an annual difference of about $58. Even though a sizeable percentage of overestimators go to college and end up on a college job, overall they earn wages virtually indistinguishable from those earned by noncollege-track students, few of whom end up going to college and holding college jobs.

[12] Recall that, for this comparison, the framework in Section 1.2 does not require me to condition on type of job in 2000: as of period 3, all underestimators are either on the college job or in college. If I do condition on holding a noncollege job in 2000, results are qualitatively similar.

capital type is not a significant predictor of being employed in 2000. I run the same set of regressions as in Tables 1.6 and 1.7 with a dummy indicating that the respondent was employed in 2000 as the dependent variable (of course, in comparing overestimators to noncollege-track students, I do not condition on the type of job in 2000). Point estimates are positive in all specifications for the coefficient on overestimators, and for underestimators in specifications (3) through (5)—indicating that, if anything, both types are more likely to be employed than students not on the college track. The coefficient on overestimators is not significantly different from zero in specifications (2) through (5), and the coefficient on underestimators is not significantly different from zero in any of the specifications. (Please see Appendix 1.3 for these results).

1.5.3 The Role of Job Tenure and Educational Attainment

According to the framework outlined in Section 1.2, overestimators earn lower wages than noncollege-track individuals in the noncollege job because they have accumulated less job tenure. Table 1.8, reporting results from linear regressions of job tenure, measured in years, on the covariates discussed in Section 1.4, shows that overestimators have less job tenure than noncollege-track students conditional on holding a noncollege job in 2000. Point estimates range from -0.4 to -0.6 and are significant at less than the ten percent level in all specifications, indicating that overestimators have worked on the noncollege job for approximately 5-7 fewer months than noncollege-track students.[13]

[13] The difference in job tenure between overestimators and noncollege-track students is even more pronounced when I do not condition on job type in 2000.

Table 1.9 contains the results on educational attainment, comparing overestimators to noncollege-track students, conditional on holding a noncollege job in 2000. Educational attainment is a categorical variable taking values from one through seven (1: less than high school, 2: high school graduate, 3: some postsecondary but no degree or certificate, 4: certificate, 5: associate's degree, 6: bachelor's degree, 7: graduate degree). OLS results are reported; ordered probit results are qualitatively similar. The results accord with the predictions of Section 1.2—overestimators have significantly greater educational attainment than noncollege-track students. Point estimates range from 0.5 to 1 and are significant at the less-than-one-percent level in every specification.[14]

Now I turn to the comparison between underestimators and noncollege-track students. Table 1.10 contains the results on job tenure. Recall that underestimators are predicted to have lower job tenure even without conditioning on job type. For the most part, point estimates accord with predictions. They range from -0.4 in specification (1) (5 fewer months) to 0.1 in specification (5) (1 more month).[15] Coefficients are not significantly different from zero in specifications (2) through (5). These results are difficult to interpret meaningfully, however, because unreported F-tests of the joint significance of all covariates yield p-values greater than 0.1 in specifications (2) through (5).

Table 1.11 contains the results comparing the educational attainment of underestimators to that of noncollege-track students. Underestimators have significantly

[14] The difference in educational attainment between overestimators and noncollege-track students is even more pronounced when I do not condition on job type in 2000.
[15] When I condition on holding a noncollege job, the point estimate remains negative in specification (5).

greater educational attainment than noncollege-track students: point estimates range from 0.3 to 0.5 and are significant at the five percent level in each specification. These results lend support to the claim that underestimators earn wages no greater than those of noncollege-track students because some have given postsecondary education a try before returning to the noncollege job.

In this section, I have demonstrated that labor-market knowledge affects wages and discussed the role of job tenure and educational achievement. In the next section, I analyze school inputs that are associated with students' information capital.

1.6 School Inputs Influencing Students' Information Capital

In this section, I conduct an analysis of the school inputs that are associated with students' career aspirations and labor-market knowledge. I attempt to control for unobservable neighborhood characteristics by including a detailed set of variables describing local labor-market conditions. I obtain zip-code level data on occupation, education, income, and employment from the 1990 Census, Summary Tape File 3, and zip-code level data on industrial mix and number of business establishments from 1994 County Business Patterns (CBP) data. From the Integrated Postsecondary Education Data Center (IPEDS), I obtain the number of 2- and 4-year colleges within each high school's zip code.

Traditional education production function approaches seek to determine the effect of school inputs such as teachers, administrative methods, and pedagogical techniques (as well as school characteristics such as enrollment and grade span) on test scores (see Schwartz and Zabel, 2005, for an overview of education production functions). In this section, I perform a preliminary analysis of the relationship between school inputs and students' information-capital accumulation.

I am particularly interested in the role of guidance counselors and vocational education faculty.[16] Interaction with guidance counselors and experience in vocational education courses are directly linked to students' career aspirations and knowledge of the labor market. Crawford, Johnson, and Summers (1997) provide evidence that labor-market information provided by schools affects wages, finding that school-to-work interventions such as transmitting labor market information to students while in high school translate into higher earnings.

In order to isolate the effect of guidance and vocational faculty on students' information-capital acquisition, an ideal experiment would randomly assign students to otherwise identical schools with different numbers of vocational and guidance faculty. Since this is infeasible, one practical way to measure this effect would be to find an instrument for the number of guidance counseling and vocational education faculty employed in a school. In 1990, the Carl D. Perkins Vocational and Applied Technology Act passed, changing both the levels of federal funding for vocational education and the way these funds were allocated within states. In future work, I will investigate the effects of this act within states and determine the usefulness of changes in federal funding levels as an instrument for the number of guidance counselors and vocational faculty within a high school. One difficulty in using such an instrument in this analysis is timing: the students I study are in 12th grade during the 1991-1992 school year, when the changes

[16] A more up-to-date term for "vocational education" is "career and technical education." To be consistent with the wording of the NELS surveys, however, I use the term "vocational education."

mandated by the Perkins Act went into effect. In order to use these changes as an instrument for guidance and vocational faculty and to estimate their effect on information-capital acquisition, I would need data from a period after the changes went into effect.

Lacking such an instrument, I proceed with a correlational analysis of the relationship between the number of guidance counselors and vocational faculty and information capital. I use a multinomial logit model relating individual students' information-capital type to school inputs and other covariates. Students choose the type that yields maximum indirect utility:

$$U_{iT} = W_{iT} + \varepsilon_{iT}\ , \tag{10}$$

where $T = 1, 2, 3$, or 4 (types are defined as above), and ε_{iT} is i.i.d. Type 1 Extreme Value. Choice of type depends on student, school, and neighborhood characteristics:

$$W_{iT} = \alpha + \beta G_{iT} + \gamma V_{iT} + \varphi X_{iT} + \lambda S_{iT} + \tau Z_{iT} + \varepsilon_{iT}\ . \tag{11}$$

The choice probability, or the probability that student i chooses type T, is given by

$$P_{iT} = \frac{\exp\left(W_{iT}\right)}{\sum_{T=1}^{4} \exp\left(W_{iT}\right)}\ . \tag{12}$$

The parameters of this model are estimated using the method of maximum likelihood.

The right-hand-side variables in (11) are defined as follows. G_{iT} is the number of full-time guidance counselors in student *i*'s school, divided by tenth grade enrollment in the school. V_{iT} is similarly defined for full-time vocational education faculty.[17]

In addition to these variables of interest, I include a large number of control variables. X_{iT} contains the full set of student characteristics described in Section 1.4. S_{iT} contains the following high school characteristics: a dummy indicating that the school is a vocational school, the number of AP courses offered, the percent of the previous year's graduates attending 2- and 4-year colleges, tenth grade enrollment, student-teacher ratio, percent non-White, percent receiving free or reduced-price lunch, dummies for urban or rural location (with suburban as the omitted category), and regional dummies (North Central, South, and West, with Northeast as the omitted category).

Z_{iT} contains a wide variety of zip-code level local labor-market characteristics, and interactions with parental characteristics.[18] I obtain the following from the 1990 Census, Summary Tape File 3: percent of workers with a college job (which I interact with a dummy variable indicating that at least one of student *i*'s parents has a college job),[19] percent of those 25 and older with a B.A. or more (which I interact with a dummy indicating that at least one of student *i*'s parents has a B.A.), and per-capita income. I

[17] The tenth-grade (1990) NELS survey reports the number of guidance counselors and vocational education faculty as categorical variables with 1 = none, 2 = 1-5, 3 = 6-10, 4 = 11-15, and 5 = over 15. I assign the value 0 to category 1, 3 to category 2, 8 to category 3, 13 to category 4, and 15 to category 5.

[18] I am only able to link zip-code data to public high schools within the NELS. Of the 1,694 schools with nonmissing observations on the relevant variables, 1,404 are public.

[19] To obtain the zip code measure, I classify 1990 2-digit SOC codes into college and noncollege jobs: "Executive, administrative, and managerial occupations" and "Professional specialty occupations" are college jobs; all others are noncollege jobs. To classify parents' jobs as college or noncollege, I use the method described in Section 1.4.

also include two variables from 1994 County Business Patterns:[20] a measure of industry diversity within each zip code (computed by adding up the number of unique 2-digit SIC codes that appear in the zip code) and the total number of business establishments. My last two measures, the number of two-year and four-year colleges within the zip code, come from the Integrated Postsecondary Education Data System (IPEDS).

Table 1.12 contains the coefficients on the guidance and vocational faculty variables. Students not on the college track form the base outcome; standard errors are clustered at the school level. Appendix 1.3 contains the full set of results.

Table 1.12 gives mixed evidence on the role of guidance counselors and vocational education faculty. There appears to be no relationship between information-capital type and the number of guidance counselors. As for vocational faculty, on one hand, the table shows a negative relationship between the number of vocational faculty and the odds of choosing Type 2 over Type 1: the more vocational faculty, the less likely a student is to be an overestimator relative to being not on the college track. This is evidence that vocational education faculty can influence students' understanding of the labor market—recall that Type 2 students (overestimators) have inaccurate labor-market information, while Type 1 (noncollege-track) students share their career aspirations but have accurate labor-market information.

On the other hand, the table also suggests a negative relationship between the number of vocational education faculty and the odds of choosing Type 4 (college track) over Type 1 (noncollege track). A larger vocational education faculty might be a signal that the school has a less academic and more vocational orientation, even if it is not

[20] 1994 is the first year that zip-code level data are available in County Business Patterns.

explicitly a vocational school. Students may choose to attend a vocationally oriented school because they aspire to a noncollege job, which would bias estimates of the effect of vocational faculty on choice of information-capital type if these nonrandom attendance patterns are not addressed. In addition, students attending a vocationally oriented school may have accurate labor-market information but be less likely to aspire to a college job, either because the student has considered all postsecondary possibilities and decided that a noncollege job is the best fit, or because the student has not been exposed to college-job options. There is, in fact, a positive correlation between vocational faculty and percent of previous year's graduates attending 2-year colleges, and a negative correlation between vocational faculty and percent attending 4-year schools. Even with an extensive set of controls including these measures, I cannot claim to have included all relevant variables that affect choice of school, career aspirations, and labor-market knowledge. Additionally, though I find evidence that schools can manipulate students' career aspirations and labor-market knowledge, I cannot say that hiring more vocational faculty would be a welfare-enhancing option. Clearly, more research is needed in this area.

Apart from specification issues, it is not surprising that I have difficulty linking guidance and vocational faculty to students' information capital in light of a 2002 study by the Ferris State University Career Institute for Education and Workforce Development. This study found that more than half of the students surveyed felt that no high school personnel had been helpful in providing career or educational advice. Finding a more precise link between guidance counselors and vocational faculty and students' career aspirations and labor-market knowledge, and finding ways to strengthen this link, remain areas for further inquiry.

1.7 Conclusions and Directions for Future Research

This paper defines one measure of information capital comprising students' career aspirations and their knowledge of the labor market: 12th graders' understanding of the educational requirements of the jobs they hope to hold at 30. I develop a simple framework describing how inaccurate labor-market information leads to lower wages through decreased job tenure, driven by students entering and leaving postsecondary school as they come to an accurate understanding of the educational requirements of their chosen jobs. I find that, in similar jobs in their mid-twenties, and despite having higher grades and test scores, workers who had inaccurate labor-market information in high school earn wages no higher than students who had an accurate understanding of educational requirements. In order to determine if this effect extends past workers' early careers, repeating this analysis in a dataset like the National Longitudinal Survey of Youth, which contains information on students' educational and job aspirations in high school as well as records of labor-market outcomes throughout workers' careers, is an important next step.

I also analyze school inputs that influence information capital, paying particular attention to the role of guidance counselors and vocational education faculty. Though this is an area ripe for future research, I find preliminary evidence that schools can influence students' career aspirations and labor-market knowledge.

Information capital is both a novel output of an education production function and an important determinant of wages via educational attainment and job tenure. This paper is an early step in understanding the relationship between information capital and these outcomes, and in understanding what schools can do to improve the quality of students' information capital and prepare them for postsecondary and labor-market success.

Table 1.1: Information Capital and Career Paths

	1 (Noncollege track)
Period 1	$w^N(1+\rho)$
Period 2	$w^N(1+\rho)^2$
Period 3	$w^N(1+\rho)^3$

	2 (Overestimator)	
Period 1	In college (earn 0)	In college (earn 0)
Period 2	$w^N(1+\rho)$	In college (earn 0)
Period 3	$w^N(1+\rho)^2$	$w^C(1+\rho)$

		3 (Underestimator)	
Period 1	$w^N(1+\rho)$	$w^N(1+\rho)$	$w^N(1+\rho)$
Period 2	$w^N(1+\rho)^2$	In college (earn 0)	In college (earn 0)
Period 3	$w^N(1+\rho)^3$	$w^N(1+\rho)^2$	In college (earn 0)

	4 (College track)
Period 1	In college (earn 0)
Period 2	In college (earn 0)
Period 3	$w^C(1+\rho)$

Table 1.2: College and Noncollege Jobs

CPS jobs: Percent with B.A.	Career Goals in 12th Grade: Occupation at Age 30	Classify as "College Job"	Percent
12.87%	Office worker	No	3.24%
5.89%	Tradesperson	No	2.53%
7.26%	Farmer, farm manager	No	0.86%
9.56%	Full-time homemaker[a]	No	1.06%
3.65%	Laborer	No	0.68%
44.30%	Manager (e.g., sales manager, office manager)	No	5.28%
10.17%	Military	No	2.44%
4.04%	Operator (of machines or tools)	No	0.98%
65.94%	Professional (e.g., accountant, registered nurse, engineer)	Yes	26.71%
86.07%	Professional (e.g., dentist, doctor, lawyer)	Yes	19.65%
31.76%	Owner of a small business or restaurant, contractor	No	5.97%
13.25%	Protective service	No	3.68%
22.58%	Sales	No	1.69%
81.17%	School teacher	Yes	7.27%
4.79%	Service worker	No	2.23%
36.68%	Technical	No	5.24%
9.51%	Not planning to work[b]	No	0.26%
18.4%	Other[c]	No	10.00%
4.49%	Will be in school[d]	No	0.23%

Notes: [a] Percent of those not in the labor force because they are keeping house with a bachelor's degree or more. [b] Percent of those not in the labor force with a bachelor's degree or more. [c] Percent of population with a bachelor's degree or more. [d] Percent of those not in labor force because they are in school with a bachelor's degree or more. The total number of students in the 1992 survey with nonmissing responses to this question is 16,258.

Table 1.3: Labor-Market Knowledge Types

Labor-market knowledge type	Label	Job goal	Perceived educational requirements	Percent
1	Not on college track	Noncollege job	College degree not required	23.8%
2	Overestimator	Noncollege job	College degree required	22.0%
3	Underestimator	College job	College degree not required	5.5%
4	On college track	College job	College degree required	48.7%

Table 1.4: Means of Student and School Characteristics by Information-Capital Type

	Information-capital type				p-values of F-tests that coefficients are equal		
Variable	Not on college track	Over-estima-tors	Under-estima-tors	On college track	1, 2	1, 3	2, 3
GPA	2.621*** (0.014)	3.067*** (0.014)	2.792*** (0.025)	3.283††† (0.008)	0.000	0.000	0.000
8th grade std. test composite	46.441*** (0.199)	52.724*** (0.202)	49.031*** (0.357)	56.381††† (0.112)	0.000	0.000	0.000
Reading pro-ficiency	1.023*** (0.014)	1.304*** (0.014)	1.144*** (0.024)	1.469††† (0.008)	0.000	0.000	0.000
Math pro-ficiency	1.105*** (0.021)	1.666*** (0.022)	1.278*** (0.038)	1.927††† (0.012)	0.000	0.000	0.000
Science pro-ficiency	0.792*** (0.016)	1.077*** (0.016)	0.903*** (0.028)	1.223††† (0.009)	0.000	0.000	0.000
Take algebra	0.234*** (0.011)	0.436*** (0.011)	0.291*** (0.020)	0.532††† (0.006)	0.000	0.006	0.000
Held back a grade	0.204*** (0.007)	0.116*** (0.007)	0.117*** (0.012)	0.069††† (0.004)	0.000	0.000	0.908
Locus of control	-0.123*** (0.012)	0.085*** (0.013)	-0.073*** (0.022)	0.171††† (0.007)	0.000	0.035	0.000
Self-concept	-0.094*** (0.014)	0.083 (0.014)	-0.108*** (0.024)	0.102††† (0.008)	0.000	0.600	0.000
SES	-0.384*** (0.015)	0.096*** (0.016)	-0.339*** (0.028)	0.260††† (0.009)	0.000	0.127	0.000
Age	14.442*** (0.012)	14.316*** (0.012)	14.317*** (0.021)	14.239††† (0.007)	0.000	0.000	0.988
Female	0.432*** (0.010)	0.419*** (0.011)	0.645*** (0.019)	0.582††† (0.006)	0.324	0.000	0.000
Asian/Pa-cific Islander	0.029*** (0.005)	0.066*** (0.005)	0.037*** (0.009)	0.084††† (0.003)	0.000	0.417	0.004

Notes: This table contains results from separate regressions of each student- and school-level variable on dummies for labor-market alignment type. Type 4 is the omitted category; the regression constant gives its mean. I add the coefficients on each of Types 1-3 to the regression constant to obtain the means for Types 1-3. * indicates that the mean for Type 1, 2, or 3 is significantly different from the Type 4 mean at the 10% level, ** at the 5% level, and *** at the 1% level. ††† indicates that the Type 4 mean is significantly different from zero at the 1% level.

Table 1.4, Continued: Means of Student and School Characteristics by Information-Capital Type

Variable	Information-capital type: Not on college track	Information-capital type: Over-estima-tors	Information-capital type: Under-estima-tors	Information-capital type: On college track	p-values of F-tests that coefficients are equal: 1, 2	p-values of F-tests that coefficients are equal: 1, 3	p-values of F-tests that coefficients are equal: 2, 3
Hispanic	0.132*** (0.007)	0.111** (0.007)	0.130*** (0.012)	0.097††† (0.004)	0.008	0.908	0.122
Black	0.091* (0.006)	0.098*** (0.006)	0.096 (0.011)	0.079††† (0.003)	0.289	0.644	0.841
Native American	0.012*** (0.002)	0.009 (0.002)	0.014** (0.004)	0.006††† (0.001)	0.317	0.470	0.180
White	0.737 (0.009)	0.716* (0.010)	0.723 (0.017)	0.734††† (0.005)	0.059	0.418	0.712
Non-English dominant	0.110 (0.007)	0.110 (0.007)	0.119 (0.012)	0.108††† (0.004)	0.983	0.504	0.514
Single-parent household	0.172*** (0.008)	0.155** (0.008)	0.183*** (0.014)	0.140††† (0.004)	0.065	0.445	0.057
Discuss studies with parents	0.438*** (0.011)	0.566*** (0.011)	0.517*** (0.019)	0.620††† (0.006)	0.000	0.000	0.013
Smoke	0.067*** (0.004)	0.035 (0.004)	0.056*** (0.008)	0.030††† (0.002)	0.000	0.161	0.009
12th grade std. test composite	45.291*** (0.193)	52.177*** (0.197)	48.158*** (0.346)	56.357††† (0.110)	0.000	0.000	0.000
Guidance faculty per 10th grader	0.021 (0.001)	0.021 (0.001)	0.019*** (0.001)	0.022††† (0.000)	0.830	0.063	0.085
Vocational faculty per 10th grader	0.032*** (0.001)	0.024** (0.001)	0.028*** (0.001)	0.022††† (0.000)	0.000	0.000	0.001
Number of AP courses	4.309*** (0.134)	5.644** (0.136)	4.283*** (0.238)	5.995††† (0.075)	0.000	0.917	0.000

Notes: This table contains results from separate regressions of each student- and school-level variable on dummies for labor-market alignment type. Type 4 is the omitted category; the regression constant gives its mean. I add the coefficients on each of Types 1-3 to the regression constant to obtain the means for Types 1-3. * indicates that the mean for Type 1, 2, or 3 is significantly different from the Type 4 mean at the 10% level, ** at the 5% level, and *** at the 1% level. ††† indicates that the Type 4 mean is significantly different from zero at the 1% level.

Table 1.4, Continued: Means of Student and School Characteristics by Information-Capital Type

Variable	Information-capital type: Not on college track	Over-estima-tors	Under-estima-tors	On college track	p-values of F-tests that coefficients are equal: 1, 2	1, 3	2, 3
Percent attending 2-year	21.823*** (0.331)	19.637** (0.336)	21.197*** (0.591)	18.996††† (0.186)	0.000	0.316	0.013
Percent attending 4-year	38.067*** (0.574)	49.460*** (0.581)	39.262*** (1.021)	53.148††† (0.322)	0.000	0.269	0.000
10th grade enrollment	302.758 (5.052)	317.629 (5.128)	319.615 (9.014)	310.941††† (2.840)	0.013	0.077	0.836
Student-teacher ratio	16.212*** (0.103)	15.853* (0.105)	16.492*** (0.185)	15.672††† (0.058)	0.003	0.153	0.001
Percent non-White	26.055 (0.625)	26.286** (0.635)	26.526 (1.113)	25.029††† (0.352)	0.755	0.689	0.839
Percent free lunch	22.823*** (0.438)	17.266*** (0.445)	21.396*** (0.781)	15.617††† (0.246)	0.000	0.084	0.000
Urban	0.214*** (0.009)	0.303** (0.009)	0.232*** (0.017)	0.326††† (0.005)	0.000	0.307	0.000
Suburban	0.376*** (0.010)	0.415 (0.010)	0.421 (0.018)	0.405††† (0.006)	0.001	0.017	0.753
Rural	0.410*** (0.009)	0.282 (0.010)	0.347*** (0.017)	0.269††† (0.005)	0.000	0.000	0.000
Northeast	0.159*** (0.008)	0.207 (0.008)	0.186* (0.015)	0.214††† (0.005)	0.000	0.076	0.185
North Central	0.296*** (0.009)	0.256 (0.009)	0.284 (0.016)	0.259††† (0.005)	0.000	0.502	0.107
South	0.349 (0.010)	0.330 (0.010)	0.324 (0.017)	0.335††† (0.005)	0.082	0.165	0.766
West	0.196 (0.008)	0.207* (0.008)	0.206 (0.015)	0.192††† (0.005)	0.24	0.536	0.908

Notes: This table contains results from separate regressions of each student- and school-level variable on dummies for labor-market alignment type. Type 4 is the omitted category; the regression constant gives its mean. I add the coefficients on each of Types 1-3 to the regression constant to obtain the means for Types 1-3. * indicates that the mean for Type 1, 2, or 3 is significantly different from the Type 4 mean at the 10% level, ** at the 5% level, and *** at the 1% level. ††† indicates that the Type 4 mean is significantly different from zero at the 1% level.

Table 1.5: Information Capital Predicts Holding a College Job

	Specification				
Information-capital type	(1)	(2)	(3)	(4)	(5)
1	-0.252***	-0.176***	-0.163***	-0.158***	-0.155***
(Noncollege track)	(0.013)	(0.019)	(0.019)	(0.019)	(0.021)
2	-0.112***	-0.093***	-0.090***	-0.082***	-0.089***
(Overestimator)	(0.016)	(0.019)	(0.019)	(0.019)	(0.022)
3	-0.214***	-0.155***	-0.132***	-0.135***	-0.161***
(Underestimator)	(0.024)	(0.031)	(0.032)	(0.032)	(0.034)
Covariates included					
8^{th} grade ability, achievement, coursetaking, and noncognitive measures	No	Yes	Yes	Yes	Yes
Age, gender, race and ethnicity, SES	No	No	Yes	Yes	Yes
Household environment and preferences	No	No	No	Yes	Yes
12^{th} grade standardized test composite	No	No	No	No	Yes
High school FE	Yes	Yes	Yes	Yes	Yes
Regression statistics					
Number of obs.	8776	6283	6187	6076	5082
Adjusted R-squared	0.191	0.226	0.231	0.230	0.222

Notes: The dependent variable is "college job in 2000." Linear probability results reported; probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 4 (students on the college track) is the omitted category.

Table 1.6: Effect of Labor-Market Knowledge on Wages,
Comparing Overestimators to Noncollege-Track Students on Noncollege Jobs

	Specification				
Coefficient	(1)	(2)	(3)	(4)	(5)
T_2 (Overestimators)	0.048*	-0.038	-0.031	-0.028	-0.027
	(0.027)	(0.037)	(0.036)	(0.037)	(0.042)
	Covariates included				
8th grade ability, achievement, coursetaking, and noncognitive measures	No	Yes	Yes	Yes	Yes
Age, gender, race and ethnicity, SES	No	No	Yes	Yes	Yes
Household environment and preferences	No	No	No	Yes	Yes
12th grade standardized test composite	No	No	No	No	Yes
High school FE	Yes	Yes	Yes	Yes	Yes
	Regression statistics				
Number of obs.	2974	1994	1973	1928	1621
Adjusted R-squared	0.182	0.213	0.266	0.272	0.264

Notes: The dependent variable is log hourly wage in 2000, when the average age of respondents is 26. I restrict the sample to students who aspired to a noncollege job in 12th grade and who were employed in a noncollege job in 2000. OLS results reported. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Table 1.7: Effect of Labor-Market Knowledge on Wages, Comparing Underestimators to Noncollege-Track Students

	Specification				
Coefficient	(1)	(2)	(3)	(4)	(5)
T_3 (Underestimators)	-0.091**	-0.150***	-0.066	-0.061	-0.081
	(0.045)	(0.057)	(0.055)	(0.054)	(0.055)
	Covariates included				
8th grade ability, achievement, coursetaking, and noncognitive measures	No	Yes	Yes	Yes	Yes
Age, gender, race and ethnicity, SES	No	No	Yes	Yes	Yes
Household environment and preferences	No	No	No	Yes	Yes
12th grade standardized test composite	No	No	No	No	Yes
High school FE	Yes	Yes	Yes	Yes	Yes
	Regression statistics				
Number of obs.	2335	1565	1549	1514	1266
Adjusted R-squared	0.228	0.326	0.401	0.406	0.391

Notes: The dependent variable is log hourly wage in 2000, when the average age of respondents is 26. I restrict the sample to students who aspired to a noncollege job in 12th grade. OLS results reported. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Table 1.8: Labor-Market Information and Job Tenure,
Comparing Overestimators to Noncollege-Track Students on Noncollege Jobs

	Specification				
Coefficient	(1)	(2)	(3)	(4)	(5)
T_2 (Overestimators)	-0.429***	-0.609***	-0.583***	-0.530***	-0.376*
	(0.131)	(0.192)	(0.194)	(0.195)	(0.217)
Covariates included					
8th grade ability, achievement, coursetaking, and noncognitive measures	No	Yes	Yes	Yes	Yes
Age, gender, race and ethnicity, SES	No	No	Yes	Yes	Yes
Household environment and preferences	No	No	No	Yes	Yes
12th grade standardized test composite	No	No	No	No	Yes
High school FE	Yes	Yes	Yes	Yes	Yes
Regression statistics					
Number of obs.	3208	2152	2125	2076	1741
Adjusted R-squared	0.167	0.146	0.159	0.157	0.168

Notes: Job tenure is measured in years. OLS results reported. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Table 1.9: Labor-Market Knowledge and Educational Attainment, Comparing Overestimators to Noncollege-Track Students on Noncollege Jobs

Coefficient	Specification (1)	(2)	(3)	(4)	(5)
T_2 (Overestimators)	0.968***	0.630***	0.534***	0.537***	0.514***
	(0.074)	(0.102)	(0.100)	(0.103)	(0.111)
Covariates included					
8th grade ability, achievement, coursetaking, and noncognitive measures	No	Yes	Yes	Yes	Yes
Age, gender, race and ethnicity, SES	No	No	Yes	Yes	Yes
Household environment and preferences	No	No	No	Yes	Yes
12th grade standardized test composite	No	No	No	No	Yes
High school FE	Yes	Yes	Yes	Yes	Yes
Regression statistics					
Number of obs.	3180	2132	2104	2056	1722
Adjusted R-squared	0.295	0.354	0.391	0.396	0.409

Notes: Educational attainment is a categorical variable taking values from one through seven (1 = less than high school, 2 = high school graduate, 3 = some postsecondary but no degree or certificate, 4 = certificate, 5 = associate's degree, 6 = bachelor's degree, 7 = graduate degree). OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Table 1.10: Labor-Market Information and Job Tenure, Comparing Underestimators to Noncollege-Track Students

	Specification				
Coefficient	(1)	(2)	(3)	(4)	(5)
T_3 (Underestimators)	-0.359* (0.190)	-0.337 (0.283)	-0.153 (0.282)	-0.120 (0.286)	0.104 (0.313)
	Covariates included				
8th grade ability, achievement, coursetaking, and noncognitive measures	No	Yes	Yes	Yes	Yes
Age, gender, race and ethnicity, SES	No	No	Yes	Yes	Yes
Household environment and preferences	No	No	No	Yes	Yes
12th grade standardized test composite	No	No	No	No	Yes
High school FE	Yes	Yes	Yes	Yes	Yes
	Regression statistics				
Number of obs.	2510	1679	1661	1624	1359
Adjusted R-squared	0.135	0.137	0.144	0.133	0.156

Notes: Job tenure is measured in years. OLS results reported. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category. Unreported F-tests of the joint significance of all covariates yield p-values greater than 0.1 in specifications (2) through (5).

Table 1.11: Labor-Market Knowledge and Educational Attainment, Comparing Underestimators to Noncollege-Track Students

	Specification				
Coefficient	(1)	(2)	(3)	(4)	(5)
T_3 (Underestimators)	0.462*** (0.105)	0.341** (0.142)	0.362*** (0.139)	0.349** (0.139)	0.316** (0.149)
Covariates included					
8th grade ability, achievement, coursetaking, and noncognitive measures	No	Yes	Yes	Yes	Yes
Age, gender, race and ethnicity, SES	No	No	Yes	Yes	Yes
Household environment and preferences	No	No	No	Yes	Yes
12th grade standardized test composite	No	No	No	No	Yes
High school FE	Yes	Yes	Yes	Yes	Yes
Regression statistics					
Number of obs.	2497	1669	1650	1614	1347
Adjusted R-squared	0.245	0.306	0.342	0.348	0.369

Notes: Educational attainment is a categorical variable taking values from one through seven (1 = less than high school, 2 = high school graduate, 3 = some postsecondary but no degree or certificate, 4 = certificate, 5 = associate's degree, 6 = bachelor's degree, 7 = graduate degree). OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Table 1.12: The Relationship Between School Inputs and Labor-Market Knowledge

Independent variables	Overestimators	Underestimators	College-track students
Guidance faculty	2.151	-5.451	-0.311
per 10^{th} grader	(4.019)	(5.088)	(3.117)
Vocational faculty	-5.575**	-2.085	-6.380**
per 10^{th} grader	(2.609)	(4.068)	(2.668)
	Regression statistics		
Number of observations		3831	
Pseudo R-squared		0.176	

Notes: Table 11 contains the results from a multinomial logit regression of information-capital type on student, school, and zip-code characteristics. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Type 1 (noncollege-track students) is the base category. Standard errors are clustered at the school level.

Appendix 1.1: Mapping of Detailed CPS Occupations to NELS 88 Survey Jobs

1992 Mapping: CPS to NELS

	Career Goals in 12th Grade: Occupation at Age 30	
March 1992 CPS Detailed Occupation Title	Code	Label
Other admin support, inc. clerical	1	Office worker
Financial records processing	1	Office worker
Secretaries, stenographers, and typists	1	Office worker
Mail and message distributing	1	Office worker
Other precision prod., craft, & repair	2	Tradesperson
Mechanics and repairers	2	Tradesperson
Construction trades	2	Tradesperson
Farm operators and managers	3	Farmer, farm manager
Farm workers and related occupations	3	Farmer, farm manager
Forestry and fishing occs	3	Farmer, farm manager
Not in labor force because keeping house	4	Full-time homemaker
Oth handlrs,equip.cleanrs,helprs,labrrs	5	Laborer
Construction laborers	5	Laborer
Management related occupations	6	Manager (e.g., sales manager, office manager)
Officials & administrators, pub. admin.	6	Manager (e.g., sales manager, office manager)
Other executive, admin. & managerial	6	Manager (e.g., sales manager, office manager)
Supervisors, admin. support	6	Manager (e.g., sales manager, office manager)
Armed forces	7	Military
Computer equipment operators	8	Operator (of machines or tools)
Fabricatrs,assemblrs,inspectrs, samplrs	8	Operator (of machines or tools)
Motor vehicle operators	8	Operator (of machines or tools)
Machine opertrs and tenders,exc precis.	8	Operator (of machines or tools)
Freight, stock & materials handlers	8	Operator (of machines or tools)
Other transp. & material moving occs	8	Operator (of machines or tools)

1992 Mapping, Continued: CPS to NELS

	Career Goals in 12th Grade: Occupation at Age 30	
March 1992 CPS Detailed Occupation Title	Code	Label
Engineers	9	Professional (e.g., accountant, registered nurse, engineer)
Other professional specialty occs.	9	Professional (e.g., accountant, registered nurse, engineer)
Health diagnosing occs.	10	Professional (e.g., dentist, doctor, lawyer)
Lawyers and judges	10	Professional (e.g., dentist, doctor, lawyer)
Teachers, college and university	10	Professional (e.g., dentist, doctor, lawyer)
Natural Scientists	10	Professional (e.g., dentist, doctor, lawyer)
Mathematical and computer scientists	10	Professional (e.g., dentist, doctor, lawyer)
Protective service	12	Protective service
Sales reps, finance and business serv.	13	Sales
Sales reps, commodities, exc. retail	13	Sales
Supervisors and proprietors, sales occs	13	Sales
Sales related occs	13	Sales
Sales workers, retail & personal serv.	13	Sales
Teachers, except college and university	14	School teacher
Personal service	15	Service worker
Health service	15	Service worker
Food service	15	Service worker
Private household service occs	15	Service worker
Cleaning and building service	15	Service worker
Health assessment and treatment occs.	16	Technical
Technicians, exc. health,engin.&science	16	Technical
Health technologists and technicians	16	Technical
Engineering and science technicians	16	Technical
Not in labor force	17	Not planning to work

1992 Mapping, Continued: CPS to NELS

	Career Goals in 12th Grade: Occupation at Age 30	
March 1992 CPS Detailed Occupation Title	Code	Label
Not in labor force, other	18	Other
Not in labor force because in school	19	Will be in school

2000 Mapping: NELS Codes, Labels, Percent B.A. or More According to NELS, and Job Classification

NELS 2000 Occupation Code	Label	1 = College Job
1	Secretaries and receptionists	0
2	Cashiers, tellers, sales clerks	0
3	Clerks, data entry	0
4	Clerical other	0
5	Farmers, foresters, farm laborers	0
6	Personal services	0
7	Cooks, chefs, bakers, cake decorators	0
8	Laborers (other than farm)	0
9	Mechanic, repairer, service technicians	0
10	Craftsmen	0
11	Skilled operatives	0
12	Transport operatives (not pilots)	0
13	Protective services, criminal justice	0
14	Military	0
15	Business/financial support services	0
16	Financial services professionals	1
17	Sales/purchasing	0
18	Customer service	0
19	Legal professionals	1
20	Legal support	1
21	Medical practice professionals	1
22	Medical licensed professionals	0
23	Medical services	0
24	Educators-K-12 teachers	1
25	Educators-instructors other than K-12	1
26	Human services professionals	1
27	Engineers architects software engineers	1
28	Scientist, statistician professionals	1
29	Research assistants/lab technicians	1
30	Technical/professional workers, other	0
31	Computer systems/related professionals	1
32	Computer programmers	1
33	Computer/computer equipment operators	0
34	Editors, writers, reporters	1
35	Performers/artists	0
36	Managers-executive	0
37	Managers-midlevel	0

2000 Mapping, Continued: NELS Codes, Labels, Percent B.A. or More According to NELS, and Job Classification

NELS 2000 Occupation Code	Label	1 = College Job
38	Managers-supervisory, office, other Admin.	0
39	Health/recreation services	0
40	Other employed-not coded elsewhere	0

2000 Mapping: CPS to NELS and Percent B.A. or More According to CPS

March 2000 CPS Detailed Occupation Title	NELS 2000 Occupation Code	Percent B.A.
Health diagnosing occs.	21	99.36%
Lawyers and judges	19	98.75%
Teachers, college and university	25	89.23%
Natural Scientists	28	88.81%
Teachers, except college and university	24	81.53%
Engineers	27	76.66%
Other professional specialty occs.	16, 20, 34, 35	67.44%
Mathematical and computer scientists	28, 31, 32	66.32%
Health assessment and treatment occs.	22	60.64%
Management related occupations	26	54.22%
Officials & administrators, pub. admin.		51.31%
Sales reps, finance and business serv.		49.32%
Other executive, admin. & managerial	36, 37, 38	45.83%
Sales reps, commodities, exc. retail		42.53%
Supervisors and proprietors, sales occs		27.77%
Supervisors, admin. support		24.64%
Sales related occs	17	21.09%
Armed forces	14	20.29%
Computer equipment operators	33	18.88%
Health technologists and technicians	23, 39	18.65%
Engineering and science technicians	29, 30	17.81%
Protective service	13	16.31%
Farm operators and managers		16.26%
Other admin support, inc. clerical	3, 4	14.33%
Financial records processing	15	13.92%
Forestry and fishing occs	5	13.43%
Mail and message distributing		12.13%
Sales workers, retail & personal serv.	2, 18	11.04%
Personal service	6	10.01%
Secretaries, stenographers, and typists	1	9.33%
Other precision prod., craft, & repair	10,11	7.94%
Mechanics and repairers	9	6.36%
Motor vehicle operators		6.17%
Farm workers and related occupations	5	5.76%
Private household service occs		5.53%
Construction trades		5.17%
Health service		5.10%
Fabricatrs,assemblrs,inspectrs,samplrs		4.93%
Food service	7	4.16%
Machine opertrs and tenders,exc precis.		3.98%
Freight, stock & materials handlers		3.72%
Construction laborers		3.30%

2000 Mapping, Continued: CPS to NELS and Percent B.A. or More According to CPS

March 2000 CPS Detailed Occupation Title	NELS 2000 Occupation Code	Percent B.A.
Oth handlrs,equip.cleanrs,helprs,labrrs	8	3.10%
Cleaning and building service		3.02%
Other transp. & material moving occs	12	3.01%
Technicians, exc. health,engin.&science		

Appendix 1.2: Results Using Continuous Measures of Labor-Market Knowledge
Note: In this appendix, I report results for Specification (4) only.

Continuous Results from Table 1.5: Information Capital Predicts Holding a College Job

	Coefficient	Robust SE	t	P>t
Pr(Type 1)	-0.205	0.028	-7.350	0.000
Pr(Type 2)	-0.154	0.034	-4.480	0.000
Pr(Type 3)	-0.160	0.043	-3.710	0.000
GPA	0.035	0.015	2.330	0.020
Standardized test composite	0.007	0.002	4.100	0.000
Reading proficiency level 2	-0.017	0.025	-0.680	0.495
Reading proficiency level 3	-0.052	0.032	-1.640	0.100
Math proficiency level 2	-0.022	0.023	-0.950	0.341
Math proficiency level 3	-0.018	0.029	-0.620	0.538
Math proficiency level 4	-0.030	0.037	-0.830	0.408
Science proficiency level 2	0.009	0.021	0.410	0.678
Science proficiency level 3	0.027	0.026	1.020	0.309
Take algebra	0.022	0.018	1.170	0.243
Held back a grade	-0.015	0.026	-0.580	0.563

Notes: The dependent variable is "college job in 2000." Linear probability results reported; probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 4 (students on the college track) is the omitted category.

Continuous Results from Table 1.5, Continued: Information Capital Predicts Holding a College Job

	Coefficient	Robust SE	t	P>t
Locus of control	0.011	0.016	0.680	0.494
Self-concept	-0.001	0.014	-0.100	0.918
SES	0.069	0.013	5.460	0.000
Age	0.005	0.016	0.340	0.734
Female	-0.008	0.015	-0.530	0.595
Asian/Pacific Islander	-0.032	0.041	-0.790	0.431
Hispanic	-0.036	0.032	-1.140	0.254
Black	-0.005	0.035	-0.140	0.886
Native American	-0.012	0.077	-0.150	0.879
Non-English dominant	0.078	0.032	2.440	0.015
Single-parent household	0.023	0.022	1.080	0.281
Discuss studies with parents	0.037	0.015	2.470	0.013
Smoke	-0.023	0.035	-0.650	0.515
Constant	-0.190	0.240	-0.790	0.428
Regression Statistics				
Number of observations	6076			
Adjusted R-squared	0.231			

Notes: The dependent variable is "college job in 2000." Linear probability results reported; probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 4 (students on the college track) is the omitted category.

Continuous Results from Table 1.6: Effect of Labor-Market Knowledge on Wages, Comparing Overestimators to Noncollege-Track Students on the Noncollege Job

	Coefficient	Robust SE	t	P>t
Pr(Type 2)	-0.046	0.047	-0.960	0.337
GPA	0.009	0.029	0.310	0.753
Standardized test composite	0.010	0.004	2.160	0.031
Reading proficiency level 2	-0.072	0.057	-1.270	0.206
Reading proficiency level 3	-0.081	0.073	-1.120	0.262
Math proficiency level 2	0.045	0.050	0.890	0.371
Math proficiency level 3	0.006	0.073	0.080	0.934
Math proficiency level 4	0.030	0.092	0.320	0.748
Science proficiency level 2	-0.036	0.045	-0.810	0.416
Science proficiency level 3	-0.104	0.075	-1.380	0.167
Take algebra	0.063	0.043	1.480	0.140
Held back a grade	-0.084	0.063	-1.330	0.183
Locus of control	0.039	0.037	1.060	0.290
Self-concept	0.034	0.035	0.980	0.328
SES	0.029	0.032	0.880	0.380
Age	-0.006	0.040	-0.160	0.875
Female	-0.267	0.039	-6.810	0.000

Notes: The dependent variable is log hourly wage in 2000, when the average age of respondents is 26. I restrict the sample to students who aspired to a noncollege job in 12th grade and who were employed in a noncollege job in 2000. OLS results reported. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Continuous Results from Table 1.6, Continued: Effect of Labor-Market Knowledge on Wages, Comparing Overestimators to Noncollege-Track Students on the Noncollege Job

	Coefficient	Robust SE	t	P>t
Asian/Pacific Islander	-0.176	0.139	-1.270	0.206
Hispanic	-0.074	0.084	-0.870	0.383
Black	-0.102	0.082	-1.240	0.215
Native American	0.089	0.268	0.330	0.741
Non-English dominant	0.077	0.089	0.860	0.391
Single-parent household	-0.105	0.045	-2.320	0.020
Discuss studies with parents	0.007	0.044	0.150	0.880
Smoke	-0.016	0.060	-0.270	0.785
Constant	2.294	0.584	3.930	0.000
Regression Statistics				
Number of observations	1928			
Adjusted R-squared	0.272			

Notes: The dependent variable is log hourly wage in 2000, when the average age of respondents is 26. I restrict the sample to students who aspired to a noncollege job in 12th grade and who were employed in a noncollege job in 2000. OLS results reported. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Continuous Results from Table 1.7: Effect of Labor-Market Knowledge on Wages, Comparing Underestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Pr(Type 3)	-0.084	0.087	-0.960	0.336
GPA	0.017	0.039	0.450	0.656
Standardized test composite	0.007	0.005	1.250	0.213
Reading proficiency level 2	-0.088	0.066	-1.320	0.187
Reading proficiency level 3	-0.064	0.087	-0.740	0.459
Math proficiency level 2	-0.013	0.050	-0.260	0.795
Math proficiency level 3	-0.069	0.077	-0.900	0.371
Math proficiency level 4	-0.063	0.109	-0.570	0.567
Science proficiency level 2	-0.016	0.046	-0.350	0.724
Science proficiency level 3	-0.063	0.084	-0.750	0.451
Take algebra	0.110	0.050	2.230	0.026
Held back a grade	-0.138	0.079	-1.760	0.078
Locus of control	0.089	0.037	2.440	0.015
Self-concept	0.060	0.036	1.660	0.098
SES	0.057	0.039	1.460	0.144
Age	0.022	0.041	0.520	0.601
Female	-0.310	0.042	-7.480	0.000

Notes: The dependent variable is log hourly wage in 2000, when the average age of respondents is 26. I restrict the sample to students who aspired to a noncollege job in 12th grade. OLS results reported. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Continuous Results from Table 1.7, Continued: Effect of Labor-Market Knowledge on Wages, Comparing Underestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Asian/Pacific Islander	-0.289	0.199	-1.460	0.146
Hispanic	0.007	0.077	0.090	0.929
Black	-0.204	0.095	-2.150	0.032
Native American	-0.158	0.260	-0.610	0.544
Non-English dominant	-0.012	0.079	-0.150	0.883
Single-parent household	0.008	0.067	0.120	0.907
Discuss studies with parents	0.001	0.048	0.030	0.979
Smoke	-0.055	0.068	-0.810	0.418
Constant	2.088	0.633	3.300	0.001
Regression Statistics				
Number of observations		1514		
Adjusted R-squared		0.405		

Notes: The dependent variable is log hourly wage in 2000, when the average age of respondents is 26. I restrict the sample to students who aspired to a noncollege job in 12th grade. OLS results reported. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Continuous Results from Table 1.8: Labor-Market Information and Job Tenure, Comparing Overestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Pr(Type 2)	-0.641	0.250	-2.560	0.010
GPA	-0.001	0.155	0.000	0.997
Standardized test composite	-0.005	0.020	-0.270	0.786
Reading proficiency level 2	0.626	0.301	2.080	0.038
Reading proficiency level 3	0.343	0.386	0.890	0.374
Math proficiency level 2	0.007	0.256	0.030	0.979
Math proficiency level 3	-0.209	0.320	-0.650	0.514
Math proficiency level 4	-0.352	0.437	-0.810	0.421
Science proficiency level 2	-0.212	0.223	-0.950	0.342
Science proficiency level 3	-0.055	0.299	-0.180	0.854
Take algebra	-0.058	0.229	-0.250	0.800
Held back a grade	-0.390	0.299	-1.310	0.192
Locus of control	0.147	0.169	0.870	0.385
Self-concept	0.038	0.157	0.240	0.809
SES	-0.206	0.167	-1.240	0.217
Age	0.107	0.200	0.540	0.592
Female	-0.628	0.184	-3.420	0.001

Notes: Job tenure is measured in years. OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Continuous Results from Table 1.8, Continued: Labor-Market Information and Job Tenure,Comparing Overestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Asian/Pacific Islander	0.255	0.616	0.410	0.679
Hispanic	-0.838	0.428	-1.960	0.050
Black	-0.299	0.473	-0.630	0.527
Native American	0.399	1.372	0.290	0.771
Non-English dominant	0.391	0.520	0.750	0.451
Single-parent household	-0.249	0.264	-0.940	0.346
Discuss studies with parents	0.193	0.178	1.080	0.280
Smoke	-0.479	0.440	-1.090	0.277
Constant	1.990	3.019	0.660	0.510
		Regression Statistics		
Number of observations		2076		
Adjusted R-squared		0.157		

Notes: Job tenure is measured in years. OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Continuous Results from Table 1.9: Labor-Market Knowledge and Educational Attainment, Comparing Overestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Pr(Type 2)	0.689	0.132	5.210	0.000
GPA	0.237	0.081	2.920	0.004
Standardized test composite	0.030	0.011	2.790	0.005
Reading proficiency level 2	-0.285	0.156	-1.830	0.068
Reading proficiency level 3	-0.415	0.196	-2.120	0.034
Math proficiency level 2	-0.152	0.125	-1.220	0.223
Math proficiency level 3	-0.275	0.177	-1.560	0.120
Math proficiency level 4	-0.508	0.241	-2.110	0.035
Science proficiency level 2	0.056	0.111	0.500	0.616
Science proficiency level 3	0.096	0.171	0.560	0.574
Take algebra	0.316	0.117	2.700	0.007
Held back a grade	0.087	0.146	0.600	0.552
Locus of control	-0.008	0.095	-0.080	0.937
Self-concept	0.029	0.091	0.320	0.751
SES	0.525	0.089	5.910	0.000
Age	-0.124	0.098	-1.260	0.208
Female	0.084	0.098	0.860	0.393

Notes: Educational attainment is a categorical variable taking values from one through seven (1 = less than high school, 2 = high school graduate, 3 = some postsecondary but no degree or certificate, 4 = certificate, 5 = associate's degree, 6 = bachelor's degree, 7 = graduate degree). OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Continuous Results from Table 1.9, Continued: Labor-Market Knowledge and Educational Attainment, Comparing Overestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Asian/Pacific Islander	0.019	0.266	0.070	0.942
Hispanic	0.099	0.170	0.580	0.562
Black	0.405	0.243	1.670	0.096
Native American	-0.403	0.412	-0.980	0.329
Non-English dominant	0.116	0.222	0.520	0.600
Single-parent household	-0.034	0.143	-0.240	0.814
Discuss studies with parents	0.054	0.099	0.550	0.581
Smoke	-0.315	0.186	-1.690	0.091
Constant	3.207	1.453	2.210	0.027
		Regression Statistics		
Number of observations		2056		
Adjusted R-squared		0.397		

Notes: Educational attainment is a categorical variable taking values from one through seven (1 = less than high school, 2 = high school graduate, 3 = some postsecondary but no degree or certificate, 4 = certificate, 5 = associate's degree, 6 = bachelor's degree, 7 = graduate degree). OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Continuous Results from Table 1.10: Labor-Market Information and Job Tenure, Comparing Underestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Pr(Type 3)	-0.035	0.485	-0.070	0.942
GPA	0.127	0.201	0.630	0.528
Standardized test composite	-0.002	0.027	-0.070	0.945
Reading proficiency level 2	0.095	0.342	0.280	0.781
Reading proficiency level 3	-0.054	0.482	-0.110	0.911
Math proficiency level 2	-0.052	0.283	-0.180	0.854
Math proficiency level 3	-0.119	0.397	-0.300	0.765
Math proficiency level 4	-0.064	0.596	-0.110	0.915
Science proficiency level 2	-0.160	0.250	-0.640	0.521
Science proficiency level 3	-0.046	0.365	-0.130	0.899
Take algebra	0.129	0.278	0.460	0.644
Held back a grade	-0.282	0.364	-0.770	0.440
Locus of control	0.039	0.208	0.190	0.850
Self-concept	0.100	0.184	0.540	0.588
SES	-0.249	0.200	-1.240	0.214
Age	0.223	0.211	1.060	0.290
Female	-0.574	0.220	-2.610	0.009

Notes: Job tenure is measured in years. OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Continuous Results from Table 1.10, Continued: Labor-Market Information and Job Tenure, Comparing Underestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Asian/Pacific Islander	-0.694	0.918	-0.760	0.450
Hispanic	-1.055	0.534	-1.980	0.048
Black	-0.552	0.596	-0.930	0.355
Native American	-0.852	1.527	-0.560	0.577
Non-English dominant	0.693	0.479	1.450	0.148
Single-parent household	-0.260	0.330	-0.790	0.431
Discuss studies with parents	0.070	0.222	0.310	0.753
Smoke	-0.357	0.384	-0.930	0.353
Constant	0.084	3.120	0.030	0.978
		Regression Statistics		
Number of observations		1624		
Adjusted R-squared		0.133		

Notes: Job tenure is measured in years. OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Continuous Results from Table 1.11: Labor-Market Knowledge and Educational Attainment, Comparing Underestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Pr(Type 3)	0.572	0.235	2.430	0.015
GPA	0.350	0.093	3.770	0.000
Standardized test composite	0.013	0.012	1.050	0.296
Reading proficiency level 2	-0.107	0.184	-0.580	0.559
Reading proficiency level 3	-0.217	0.226	-0.960	0.338
Math proficiency level 2	-0.112	0.132	-0.850	0.396
Math proficiency level 3	-0.174	0.195	-0.890	0.374
Math proficiency level 4	-0.624	0.282	-2.220	0.027
Science proficiency level 2	-0.031	0.121	-0.250	0.799
Science proficiency level 3	0.005	0.197	0.030	0.978
Take algebra	0.119	0.146	0.820	0.413
Held back a grade	-0.287	0.154	-1.870	0.062
Locus of control	0.033	0.099	0.340	0.737
Self-concept	-0.065	0.092	-0.700	0.482
SES	0.494	0.097	5.080	0.000
Age	0.054	0.097	0.560	0.573
Female	0.000	0.112	0.000	0.997

Notes: Educational attainment is a categorical variable taking values from one through seven (1 = less than high school, 2 = high school graduate, 3 = some postsecondary but no degree or certificate, 4 = certificate, 5 = associate's degree, 6 = bachelor's degree, 7 = graduate degree). OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Continuous Results from Table 1.11, Continued: Labor-Market Knowledge and Educational Attainment, Comparing Underestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Asian/Pacific Islander	-0.211	0.426	-0.500	0.620
Hispanic	0.000	0.209	0.000	0.999
Black	0.270	0.280	0.960	0.336
Native American	-0.409	0.362	-1.130	0.259
Non-English dominant	0.170	0.259	0.660	0.511
Single-parent household	0.128	0.159	0.810	0.421
Discuss studies with parents	0.147	0.118	1.250	0.211
Smoke	-0.430	0.196	-2.200	0.028
Constant	1.084	1.453	0.750	0.456
		Regression Statistics		
Number of observations		1614		
Adjusted R-squared		0.348		

Notes: Educational attainment is a categorical variable taking values from one through seven (1 = less than high school, 2 = high school graduate, 3 = some postsecondary but no degree or certificate, 4 = certificate, 5 = associate's degree, 6 = bachelor's degree, 7 = graduate degree). OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Appendix 1.3: Full Regression Results

Full Results from Table 1.5: Information Capital Predicts Holding a College Job

	Coefficient	Robust SE	t	P>t
Type 1 (Noncollege track)	-0.158	0.019	-8.320	0.000
Type 2 (Overestimators)	-0.082	0.019	-4.190	0.000
Type 3 (Underestimators)	-0.135	0.032	-4.230	0.000
GPA	0.035	0.015	2.340	0.019
Standardized test composite	0.007	0.002	4.100	0.000
Reading proficiency level 2	-0.015	0.025	-0.630	0.530
Reading proficiency level 3	-0.051	0.032	-1.600	0.109
Math proficiency level 2	-0.022	0.023	-0.970	0.333
Math proficiency level 3	-0.018	0.029	-0.630	0.532
Math proficiency level 4	-0.030	0.037	-0.820	0.411
Science proficiency level 2	0.009	0.021	0.430	0.671
Science proficiency level 3	0.027	0.026	1.030	0.301
Take algebra	0.021	0.018	1.150	0.251
Held back a grade	-0.016	0.026	-0.610	0.543
Locus of control	0.011	0.016	0.690	0.491
Self-concept	-0.001	0.014	-0.050	0.963
SES	0.071	0.013	5.570	0.000
Age	0.005	0.016	0.340	0.731
Female	-0.008	0.016	-0.510	0.613
Asian/Pacific Islander	-0.030	0.040	-0.750	0.451
Hispanic	-0.035	0.032	-1.090	0.274
Black	-0.005	0.035	-0.150	0.877
Native American	-0.013	0.077	-0.170	0.862
Non-English dominant	0.079	0.032	2.480	0.013
Single-parent household	0.024	0.022	1.120	0.261
Discuss studies with parents	0.038	0.015	2.520	0.012
Smoke	-0.023	0.035	-0.650	0.515
Constant	-0.231	0.239	-0.970	0.333
Regression Statistics				
Number of observations		6076		
Adjusted R-squared		0.230		

Notes: The dependent variable is "college job in 2000." Linear probability results reported; probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 4 (students on the college track) is the omitted category.

Full Results from Table 1.6: Effect of Labor-Market Knowledge on Wages, Comparing Overestimators to Noncollege-Track Students on the Noncollege Job

	Coefficient	Robust SE	t	P>t
Type 2 (Overestimators)	-0.028	0.037	-0.760	0.450
GPA	0.008	0.029	0.270	0.786
Standardized test composite	0.010	0.004	2.170	0.030
Reading proficiency level 2	-0.073	0.057	-1.270	0.206
Reading proficiency level 3	-0.081	0.073	-1.120	0.262
Math proficiency level 2	0.044	0.050	0.880	0.378
Math proficiency level 3	0.006	0.073	0.080	0.935
Math proficiency level 4	0.029	0.092	0.320	0.751
Science proficiency level 2	-0.036	0.045	-0.810	0.418
Science proficiency level 3	-0.104	0.075	-1.380	0.167
Take algebra	0.062	0.043	1.460	0.146
Held back a grade	-0.084	0.063	-1.330	0.185
Locus of control	0.039	0.037	1.040	0.296
Self-concept	0.034	0.035	0.970	0.331
SES	0.028	0.033	0.860	0.390
Age	-0.006	0.040	-0.150	0.882
Female	-0.267	0.039	-6.820	0.000
Asian/Pacific Islander	-0.178	0.139	-1.280	0.203
Hispanic	-0.074	0.084	-0.880	0.380
Black	-0.101	0.082	-1.230	0.218
Native American	0.085	0.269	0.320	0.751
Non-English dominant	0.077	0.089	0.860	0.391
Single-parent household	-0.106	0.045	-2.320	0.020
Discuss studies with parents	0.007	0.044	0.150	0.878
Smoke	-0.017	0.060	-0.280	0.779
Constant	2.289	0.585	3.910	0.000
Regression Statistics				
Number of observations		1928		
Adjusted R-squared		0.272		

Notes: The dependent variable is log hourly wage in 2000, when the average age of respondents is 26. I restrict the sample to students who aspired to a noncollege job in 12th grade and who were employed in a noncollege job in 2000. OLS results reported. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Full Results from Table 1.7: Effect of Labor-Market Knowledge on Wages, Comparing Underestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Type 3 (Underestimators)	-0.061	0.054	-1.130	0.258
GPA	0.019	0.039	0.480	0.635
Standardized test composite	0.007	0.005	1.270	0.206
Reading proficiency level 2	-0.087	0.066	-1.310	0.191
Reading proficiency level 3	-0.067	0.087	-0.770	0.440
Math proficiency level 2	-0.014	0.050	-0.280	0.782
Math proficiency level 3	-0.068	0.077	-0.890	0.375
Math proficiency level 4	-0.061	0.109	-0.560	0.575
Science proficiency level 2	-0.016	0.046	-0.340	0.731
Science proficiency level 3	-0.062	0.084	-0.740	0.462
Take algebra	0.109	0.049	2.210	0.028
Held back a grade	-0.142	0.078	-1.820	0.070
Locus of control	0.089	0.037	2.450	0.015
Self-concept	0.059	0.036	1.640	0.102
SES	0.057	0.039	1.450	0.147
Age	0.022	0.041	0.530	0.594
Female	-0.307	0.042	-7.390	0.000
Asian/Pacific Islander	-0.291	0.198	-1.470	0.143
Hispanic	0.010	0.077	0.130	0.897
Black	-0.205	0.095	-2.150	0.032
Native American	-0.152	0.258	-0.590	0.555
Non-English dominant	-0.011	0.079	-0.140	0.892
Single-parent household	0.012	0.065	0.180	0.859
Discuss studies with parents	0.002	0.048	0.040	0.970
Smoke	-0.055	0.068	-0.810	0.418
Constant	2.058	0.634	3.250	0.001
Regression Statistics				
Number of observations		1514		
Adjusted R-squared		0.406		

Notes: The dependent variable is log hourly wage in 2000, when the average age of respondents is 26. I restrict the sample to students who aspired to a noncollege job in 12th grade. OLS results reported. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Full Results on Employment Status in 2000: Comparing Overestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Type 3 (Overestimators)	0.022	0.020	1.130	0.258
GPA	-0.018	0.018	-1.010	0.313
Standardized test composite	0.001	0.002	0.420	0.674
Reading proficiency level 2	0.002	0.031	0.070	0.944
Reading proficiency level 3	0.008	0.040	0.190	0.848
Math proficiency level 2	0.024	0.027	0.890	0.376
Math proficiency level 3	0.011	0.035	0.320	0.748
Math proficiency level 4	0.022	0.047	0.470	0.636
Science proficiency level 2	-0.036	0.024	-1.490	0.137
Science proficiency level 3	-0.040	0.031	-1.290	0.197
Take algebra	-0.012	0.022	-0.530	0.594
Held back a grade	-0.060	0.034	-1.780	0.075
Locus of control	0.013	0.020	0.640	0.525
Self-concept	0.020	0.018	1.110	0.269
SES	0.007	0.016	0.450	0.655
Age	-0.018	0.020	-0.930	0.353
Female	-0.138	0.019	-7.390	0.000
Asian/Pacific Islander	0.066	0.080	0.830	0.408
Hispanic	0.076	0.037	2.080	0.038
Black	-0.026	0.050	-0.520	0.606
Native American	0.041	0.078	0.520	0.605
Non-English dominant	-0.102	0.048	-2.140	0.033
Single-parent household	-0.018	0.029	-0.630	0.530
Discuss studies with parents	-0.001	0.019	-0.030	0.975
Smoke	0.074	0.037	2.000	0.046
Constant	1.226	0.302	4.060	0.000
Regression Statistics				
Number of observations		2566		
Adjusted R-squared		0.235		

Notes: The dependent variable is "employed in 2000." Linear probability results reported; probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Full Results on Employment Status in 2000: Comparing Underestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Type 3 (Underestimators)	0.024	0.033	0.740	0.460
GPA	-0.022	0.027	-0.800	0.423
Standardized test composite	-0.001	0.003	-0.420	0.672
Reading proficiency level 2	-0.019	0.042	-0.460	0.644
Reading proficiency level 3	0.027	0.056	0.490	0.626
Math proficiency level 2	-0.007	0.032	-0.210	0.832
Math proficiency level 3	-0.062	0.046	-1.360	0.175
Math proficiency level 4	-0.015	0.063	-0.240	0.811
Science proficiency level 2	-0.027	0.028	-0.970	0.332
Science proficiency level 3	0.001	0.044	0.030	0.975
Take algebra	0.030	0.031	0.950	0.343
Held back a grade	-0.029	0.053	-0.550	0.580
Locus of control	0.010	0.027	0.360	0.719
Self-concept	0.020	0.025	0.810	0.418
SES	0.025	0.022	1.130	0.261
Age	-0.024	0.026	-0.900	0.366
Female	-0.159	0.028	-5.690	0.000
Asian/Pacific Islander	0.103	0.067	1.540	0.124
Hispanic	0.087	0.060	1.460	0.146
Black	0.015	0.073	0.200	0.841
Native American	0.158	0.106	1.480	0.138
Non-English dominant	-0.018	0.067	-0.270	0.784
Single-parent household	0.046	0.043	1.070	0.284
Discuss studies with parents	-0.015	0.026	-0.580	0.561
Smoke	0.067	0.055	1.220	0.225
Constant	1.437	0.386	3.730	0.000
Regression Statistics				
Number of observations		1640		
Adjusted R-squared		0.286		

Notes: The dependent variable is "employed in 2000." Linear probability results reported; probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Full Results from Table 1.8: Labor-Market Information and Job Tenure, Comparing Overestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Type 2 (Overestimators)	-0.530	0.195	-2.720	0.007
GPA	0.005	0.155	0.030	0.975
Standardized test composite	-0.005	0.020	-0.230	0.822
Reading proficiency level 2	0.630	0.301	2.090	0.037
Reading proficiency level 3	0.344	0.386	0.890	0.373
Math proficiency level 2	0.002	0.256	0.010	0.993
Math proficiency level 3	-0.219	0.320	-0.680	0.494
Math proficiency level 4	-0.359	0.435	-0.820	0.410
Science proficiency level 2	-0.215	0.223	-0.960	0.335
Science proficiency level 3	-0.054	0.299	-0.180	0.857
Take algebra	-0.059	0.229	-0.260	0.795
Held back a grade	-0.391	0.298	-1.310	0.190
Locus of control	0.142	0.169	0.850	0.398
Self-concept	0.040	0.157	0.250	0.799
SES	-0.198	0.167	-1.180	0.237
Age	0.105	0.200	0.530	0.598
Female	-0.628	0.184	-3.420	0.001
Asian/Pacific Islander	0.263	0.617	0.430	0.670
Hispanic	-0.836	0.427	-1.960	0.051
Black	-0.287	0.473	-0.610	0.543
Native American	0.380	1.367	0.280	0.781
Non-English dominant	0.390	0.517	0.750	0.451
Single-parent household	-0.246	0.264	-0.930	0.352
Discuss studies with parents	0.199	0.178	1.120	0.265
Smoke	-0.481	0.439	-1.090	0.274
Constant	1.975	3.018	0.650	0.513
Regression Statistics				
Number of observations		2076		
Adjusted R-squared		0.158		

Notes: Job tenure is measured in years. OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Full Results from Table 1.9: Labor-Market Knowledge and Educational Attainment, Comparing Overestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Type 2 (Overestimators)	0.537	0.103	5.220	0.000
GPA	0.236	0.081	2.910	0.004
Standardized test composite	0.029	0.011	2.710	0.007
Reading proficiency level 2	-0.289	0.156	-1.850	0.064
Reading proficiency level 3	-0.417	0.196	-2.130	0.034
Math proficiency level 2	-0.147	0.124	-1.190	0.236
Math proficiency level 3	-0.268	0.177	-1.510	0.130
Math proficiency level 4	-0.500	0.241	-2.080	0.038
Science proficiency level 2	0.057	0.111	0.510	0.610
Science proficiency level 3	0.094	0.171	0.550	0.582
Take algebra	0.318	0.117	2.710	0.007
Held back a grade	0.087	0.146	0.590	0.553
Locus of control	-0.002	0.095	-0.020	0.984
Self-concept	0.028	0.091	0.310	0.755
SES	0.519	0.089	5.800	0.000
Age	-0.124	0.099	-1.250	0.210
Female	0.084	0.098	0.850	0.394
Asian/Pacific Islander	0.017	0.266	0.070	0.948
Hispanic	0.100	0.169	0.590	0.553
Black	0.395	0.242	1.630	0.103
Native American	-0.376	0.396	-0.950	0.343
Non-English dominant	0.116	0.223	0.520	0.602
Single-parent household	-0.036	0.144	-0.250	0.801
Discuss studies with parents	0.049	0.099	0.500	0.618
Smoke	-0.312	0.187	-1.670	0.095
Constant	3.248	1.463	2.220	0.027
Regression Statistics				
Number of observations		2056		
Adjusted R-squared		0.396		

Notes: Educational attainment is a categorical variable taking values from one through seven (1 = less than high school, 2 = high school graduate, 3 = some postsecondary but no degree or certificate, 4 = certificate, 5 = associate's degree, 6 = bachelor's degree, 7 = graduate degree). OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 4 (students on the college track) is the omitted category.

Full Results from Table 1.10: Labor-Market Information and Job Tenure, Comparing Underestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Type 3 (Underestimators)	-0.120	0.286	-0.420	0.675
GPA	0.134	0.201	0.670	0.504
Standardized test composite	-0.002	0.027	-0.060	0.950
Reading proficiency level 2	0.103	0.343	0.300	0.765
Reading proficiency level 3	-0.055	0.483	-0.110	0.909
Math proficiency level 2	-0.052	0.283	-0.180	0.856
Math proficiency level 3	-0.115	0.397	-0.290	0.772
Math proficiency level 4	-0.053	0.596	-0.090	0.929
Science proficiency level 2	-0.156	0.250	-0.620	0.532
Science proficiency level 3	-0.041	0.365	-0.110	0.910
Take algebra	0.124	0.279	0.440	0.657
Held back a grade	-0.288	0.364	-0.790	0.428
Locus of control	0.042	0.208	0.200	0.841
Self-concept	0.098	0.184	0.530	0.597
SES	-0.248	0.200	-1.240	0.216
Age	0.220	0.211	1.040	0.299
Female	-0.558	0.220	-2.530	0.011
Asian/Pacific Islander	-0.680	0.918	-0.740	0.459
Hispanic	-1.042	0.533	-1.950	0.051
Black	-0.544	0.595	-0.910	0.361
Native American	-0.806	1.512	-0.530	0.594
Non-English dominant	0.707	0.479	1.480	0.140
Single-parent household	-0.245	0.327	-0.750	0.455
Discuss studies with parents	0.075	0.222	0.340	0.736
Smoke	-0.344	0.383	-0.900	0.369
Constant	0.096	3.128	0.030	0.976
Regression Statistics				
Number of observations		1624		
Adjusted R-squared		0.133		

Notes: Job tenure is measured in years. OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 1 (students not on the college track) is the omitted category.

Full Results from Table 1.11: Labor-Market Knowledge and Educational Attainment, Comparing Underestimators to Noncollege-Track Students

	Coefficient	Robust SE	t	P>t
Type 3 (Underestimators)	0.349	0.139	2.510	0.012
GPA	0.345	0.094	3.690	0.000
Standardized test composite	0.012	0.012	1.000	0.319
Reading proficiency level 2	-0.110	0.182	-0.600	0.547
Reading proficiency level 3	-0.202	0.226	-0.890	0.372
Math proficiency level 2	-0.106	0.131	-0.800	0.421
Math proficiency level 3	-0.173	0.195	-0.890	0.375
Math proficiency level 4	-0.631	0.281	-2.240	0.025
Science proficiency level 2	-0.031	0.121	-0.250	0.799
Science proficiency level 3	0.001	0.197	0.010	0.996
Take algebra	0.120	0.145	0.820	0.410
Held back a grade	-0.275	0.154	-1.780	0.076
Locus of control	0.034	0.099	0.340	0.731
Self-concept	-0.063	0.092	-0.680	0.498
SES	0.497	0.097	5.110	0.000
Age	0.052	0.096	0.540	0.587
Female	-0.007	0.112	-0.060	0.949
Asian/Pacific Islander	-0.200	0.416	-0.480	0.631
Hispanic	-0.012	0.205	-0.060	0.952
Black	0.277	0.278	1.000	0.319
Native American	-0.423	0.348	-1.210	0.225
Non-English dominant	0.178	0.257	0.690	0.489
Single-parent household	0.117	0.158	0.740	0.459
Discuss studies with parents	0.146	0.117	1.240	0.214
Smoke	-0.427	0.195	-2.190	0.029
Constant	1.255	1.443	0.870	0.385
	Regression Statistics			
Number of observations		1614		
Adjusted R-squared		0.348		

Notes: Educational attainment is a categorical variable taking values from one through seven (1 = less than high school, 2 = high school graduate, 3 = some postsecondary but no degree or certificate, 4 = certificate, 5 = associate's degree, 6 = bachelor's degree, 7 = graduate degree). OLS results reported; ordered probit results show the same signs and significance levels. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level. Robust standard errors and appropriate panel weights are used. Type 4 (students on the college track) is the omitted category.

Full Set of School-Level Results from Table 1.12: The Relationship between School Inputs and Labor-Market Knowledge

	Coefficient	Robust SE	t	P>t
		Overestimators		
Guidance faculty per 10th grader	2.151	4.019	0.540	0.593
Vocational faculty per 10th grader	-5.575	2.609	-2.140	0.033
GPA	0.595	0.098	6.060	0.000
8th grade std. test composite	0.008	0.015	0.530	0.596
Reading proficiency level 2	0.022	0.200	0.110	0.911
Reading proficiency level 3	-0.120	0.244	-0.490	0.624
Math proficiency level 2	-0.072	0.169	-0.430	0.668
Math proficiency level 3	-0.210	0.214	-0.980	0.325
Math proficiency level 4	-0.041	0.274	-0.150	0.882
Science proficiency level 2	-0.146	0.143	-1.020	0.306
Science proficiency level 3	-0.121	0.187	-0.650	0.517
Take algebra	0.041	0.114	0.360	0.718
Held back a grade	-0.093	0.192	-0.480	0.628
Locus of control	0.074	0.114	0.650	0.517
Self-concept	0.177	0.102	1.730	0.083
SES	0.482	0.126	3.820	0.000
Age	0.190	0.112	1.690	0.091
Female	-0.064	0.116	-0.560	0.578

Notes: Table contains the results from a multinomial logit regression. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Type 1 (noncollege-track students) is the base category. Standard errors are clustered at the school level.

Full Set of School-Level Results from Table 1.12, Continued: The Relationship between School Inputs and Labor-Market Knowledge

	Coefficient	Robust SE	t	P>t
Asian/Pacific Islander	0.819	0.333	2.460	0.014
Hispanic	0.437	0.239	1.830	0.067
Black	0.612	0.252	2.430	0.015
Native American	1.247	0.727	1.720	0.086
Non-English dominant	0.139	0.238	0.580	0.559
Single-parent household	0.135	0.158	0.850	0.394
Discuss studies with parents	0.030	0.113	0.270	0.787
Smoke	-0.444	0.302	-1.470	0.141
12th grade std. test composite	0.060	0.010	5.930	0.000
Number of AP courses	-0.005	0.012	-0.440	0.659
Percent attending 2-year	0.001	0.004	0.150	0.881
Percent attending 4-year	0.000	0.004	-0.080	0.933
10th grade enrollment	-0.001	0.000	-1.650	0.099
Student-teacher ratio	0.017	0.021	0.780	0.433
Percent non-White	-0.001	0.003	-0.500	0.616
Percent free lunch	0.005	0.004	1.210	0.226
Urban	0.106	0.169	0.630	0.529
Rural	-0.165	0.147	-1.130	0.260
North Central	0.059	0.181	0.320	0.747
South	0.196	0.201	0.970	0.330
West	-0.122	0.236	-0.520	0.604
College graduate in HH	0.390	0.317	1.230	0.219

Notes: Table contains the results from a multinomial logit regression. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Type 1 (noncollege-track students) is the base category. Standard errors are clustered at the school level.

Full Set of School-Level Results from Table 1.12, Continued: The Relationship between School Inputs and Labor-Market Knowledge

	Coefficient	Robust SE	t	P>t
Pct. B.A. in zip code	-0.039	0.021	-1.800	0.072
College graduate in HH*Pct. B.A. in zip code	0.003	0.013	0.240	0.808
Pct. unemployed in zip code	0.035	0.033	1.070	0.285
Per-capita income in zip code	0.000	0.000	1.700	0.090
College job in HH	-0.401	0.487	-0.820	0.410
Pct. college job in zip code	6.083	2.696	2.260	0.024
College job in HH*Pct. college job in zip code	1.350	2.033	0.660	0.507
Num. 4-year colleges in zip code	0.231	0.115	2.020	0.044
Num. 2-year colleges in zip code	-0.123	0.086	-1.430	0.154
Num. 2-digit SIC industries in zip code	-0.086	0.071	-1.220	0.223
Num. business est. in zip code	0.000	0.000	1.610	0.108
Constant	-8.723	1.993	-4.380	0.000

Notes: Table contains the results from a multinomial logit regression. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Type 1 (noncollege-track students) is the base category. Standard errors are clustered at the school level.

Full Set of School-Level Results from Table 1.12, Continued: The Relationship between School Inputs and Labor-Market Knowledge

	Coefficient	Robust SE	t	P>t
Underestimators				
Guidance faculty per 10th grader	-5.451	5.088	-1.070	0.284
Vocational faculty per 10th grader	-2.085	4.068	-0.510	0.608
GPA	0.149	0.137	1.090	0.275
8th grade std. test composite	0.011	0.019	0.570	0.570
Reading proficiency level 2	0.243	0.306	0.800	0.426
Reading proficiency level 3	0.090	0.366	0.240	0.807
Math proficiency level 2	0.095	0.247	0.380	0.701
Math proficiency level 3	-0.042	0.290	-0.140	0.886
Math proficiency level 4	0.128	0.385	0.330	0.739
Science proficiency level 2	0.268	0.209	1.290	0.198
Science proficiency level 3	0.432	0.260	1.660	0.097
Take algebra	-0.039	0.195	-0.200	0.841
Held back a grade	0.262	0.290	0.900	0.366
Locus of control	0.196	0.170	1.150	0.248
Self-concept	-0.147	0.129	-1.140	0.254
SES	-0.462	0.172	-2.680	0.007

Notes: Table contains the results from a multinomial logit regression. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Type 1 (noncollege-track students) is the base category. Standard errors are clustered at the school level.

Full Set of School-Level Results from Table 1.12, Continued: The Relationship between School Inputs and Labor-Market Knowledge

	Coefficient	Robust SE	t	P>t
Age	-0.044	0.183	-0.240	0.810
Female	0.804	0.157	5.110	0.000
Asian/Pacific Islander	0.215	0.532	0.400	0.686
Hispanic	0.537	0.325	1.650	0.098
Black	0.346	0.349	0.990	0.321
Native American	0.625	0.848	0.740	0.461
Non-English dominant	0.011	0.327	0.030	0.974
Single-parent household	0.346	0.210	1.650	0.099
Discuss studies with parents	0.086	0.154	0.560	0.578
Smoke	-0.169	0.402	-0.420	0.674
12^{th} grade std. test composite	0.050	0.014	3.500	0.000
Number of AP courses	-0.026	0.017	-1.530	0.126
Percent attending 2-year	-0.007	0.007	-1.070	0.285
Percent attending 4-year	-0.006	0.006	-0.960	0.336
10^{th} grade enrollment	0.000	0.001	-0.550	0.582
Student-teacher ratio	0.017	0.030	0.570	0.566
Percent non-White	-0.002	0.005	-0.420	0.672
Percent free lunch	-0.002	0.007	-0.360	0.718
Urban	0.080	0.259	0.310	0.756
Rural	-0.078	0.206	-0.380	0.706
North Central	-0.538	0.234	-2.300	0.021
South	-0.313	0.275	-1.140	0.255
West	-0.174	0.338	-0.510	0.607

Notes: Table contains the results from a multinomial logit regression. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Type 1 (noncollege-track students) is the base category. Standard errors are clustered at the school level.

Full Set of School-Level Results from Table 1.12, Continued: The Relationship between School Inputs and Labor-Market Knowledge

	Coefficient	Robust SE	t	P>t
College graduate in HH	0.344	0.520	0.660	0.509
Pct. B.A. in zip code	-0.031	0.023	-1.360	0.175
College graduate in HH*Pct. B.A. in zip code	0.008	0.020	0.400	0.690
Pct. unemployed in zip code	-0.058	0.041	-1.420	0.155
Per-capita income in zip code	0.000	0.000	-1.100	0.272
College job in HH	0.022	0.807	0.030	0.978
Pct. college job in zip code	4.960	3.290	1.510	0.132
College job in HH*Pct. college job in zip code	-0.636	3.313	-0.190	0.848
Num. 4-year colleges in zip code	0.040	0.195	0.200	0.838
Num. 2-year colleges in zip code	-0.118	0.164	-0.720	0.472
Num. 2-digit SIC industries in zip code	-0.078	0.097	-0.800	0.422
Num. business est. in zip code	0.000	0.000	1.320	0.187
Constant	-3.649	3.091	-1.180	0.238

Notes: Table contains the results from a multinomial logit regression. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Type 1 (noncollege-track students) is the base category. Standard errors are clustered at the school level.

Full Set of School-Level Results from Table 1.12, Continued: The Relationship between School Inputs and Labor-Market Knowledge

	Coefficient	Robust SE	t	P>t
		College-track students		
Guidance faculty per 10th grader	-0.311	3.157	-0.100	0.921
Vocational faculty per 10th grader	-6.380	2.556	-2.500	0.013
GPA	0.803	0.095	8.420	0.000
8th grade std. test composite	0.005	0.013	0.390	0.695
Reading proficiency level 2	-0.013	0.201	-0.070	0.948
Reading proficiency level 3	0.048	0.242	0.200	0.844
Math proficiency level 2	-0.088	0.156	-0.570	0.570
Math proficiency level 3	-0.257	0.198	-1.290	0.196
Math proficiency level 4	-0.031	0.253	-0.120	0.903
Science proficiency level 2	-0.075	0.140	-0.530	0.595
Science proficiency level 3	-0.081	0.179	-0.450	0.652
Take algebra	0.037	0.108	0.340	0.732
Held back a grade	-0.350	0.204	-1.720	0.086
Locus of control	0.235	0.102	2.300	0.022
Self-concept	0.081	0.094	0.860	0.390
SES	0.533	0.115	4.630	0.000

Notes: Table contains the results from a multinomial logit regression. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Type 1 (noncollege-track students) is the base category. Standard errors are clustered at the school level.

Full Set of School-Level Results from Table 1.12, Continued: The Relationship between School Inputs and Labor-Market Knowledge

	Coefficient	Robust SE	t	P>t
Age	0.183	0.110	1.670	0.095
Female	0.697	0.101	6.930	0.000
Asian/Pacific Islander	0.902	0.318	2.830	0.005
Hispanic	0.686	0.232	2.960	0.003
Black	0.754	0.237	3.180	0.001
Native American	0.721	0.731	0.990	0.324
Non-English dominant	0.009	0.230	0.040	0.969
Single-parent household	0.282	0.145	1.940	0.052
Discuss studies with parents	0.067	0.102	0.660	0.508
Smoke	0.026	0.294	0.090	0.928
12^{th} grade std. test composite	0.110	0.010	10.790	0.000
Number of AP courses	-0.009	0.009	-0.970	0.330
Percent attending 2-year	0.003	0.004	0.670	0.506
Percent attending 4-year	-0.002	0.004	-0.410	0.682
10^{th} grade enrollment	-0.001	0.000	-3.080	0.002
Student-teacher ratio	0.007	0.022	0.320	0.750
Percent non-White	0.000	0.003	0.000	0.996
Percent free lunch	0.002	0.005	0.380	0.708
Urban	0.201	0.179	1.120	0.261
Rural	0.017	0.146	0.110	0.910
North Central	0.083	0.172	0.490	0.627
South	0.164	0.181	0.900	0.366
West	-0.217	0.229	-0.950	0.342

Notes: Table contains the results from a multinomial logit regression. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Type 1 (noncollege-track students) is the base category. Standard errors are clustered at the school level.

Full Set of School-Level Results from Table 1.12, Continued: The Relationship between School Inputs and Labor-Market Knowledge

	Coefficient	Robust SE	t	P>t
College graduate in HH	0.598	0.309	1.930	0.053
Pct. B.A. in zip code	-0.044	0.017	-2.530	0.011
College graduate in HH*Pct. B.A. in zip code	0.004	0.013	0.310	0.758
Pct. unemployed in zip code	0.065	0.036	1.790	0.073
Per-capita income in zip code	0.000	0.000	1.540	0.125
College job in HH	-0.562	0.465	-1.210	0.227
Pct. college job in zip code	6.556	2.457	2.670	0.008
College job in HH*Pct. college job in zip code	2.095	1.935	1.080	0.279
Num. 4-year colleges in zip code	0.057	0.138	0.420	0.678
Num. 2-year colleges in zip code	0.060	0.083	0.720	0.474
Num. 2-digit SIC industries in zip code	-0.057	0.065	-0.880	0.379
Num. business est. in zip code	0.000	0.000	2.500	0.012
Constant	-11.815	1.823	-6.480	0.000
Pseudo R-squared		0.176		

Notes: Table contains the results from a multinomial logit regression. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Type 1 (noncollege-track students) is the base category. Standard errors are clustered at the school level.

References

Akerlof, George A. and Rachel E. Kranton. 2002. "Identity and Schooling: Some Lessons for the Economics of Education." *Journal of Economic Literature*, 40: 1167-1201.

Altonji, Joseph G. 1993. "The Demand for and Return to Education when Education Outcomes are Uncertain." *Journal of Labor Economics*, 11(1): 48-83.

________. 1995. "The Effects of High School Curriculum on Education and Labor Market Outcomes." *The Journal of Human Resources*, 30(3): 409-438.

Altonji, Joseph G. and Thomas A. Dunn. 1996. "Using Siblings to Estimate the Effect of School Quality on Wages." *The Review of Economics and Statistics*, 78(4): 665-671.

Anderberg, Dan and Fredrik Andersson. 2007. "Stratification, Social Networks in the Labour Market, and Intergenerational Mobility." *The Economic Journal*, 117: 782-812.

Becker, Gary S. 1993. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Chicago: University of Chicago Press.

Betts, Julian R. 1995. "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth." *The Review of Economics and Statistics*, 77(2): 231-250.

________. 1996. "What Do Students Know about Wages? Evidence from a Survey of Undergraduates." *The Journal of Human Resources*, 31(1): 27-56.

Blackburn, McKinley and David Neumark. 1992. "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials." *The Quarterly Journal of Economics*, 107(4): 1421-1436.

Blau, Francine D. and Marianne A. Ferber. 1991. "Career Plans and Expectations of Young Women and Men: The Earnings Gap and Labor Force Participation." *The Journal of Human Resources*, 26(4): 581-607.

Botelho, Anabela and Ligia Costa Pinto. 2004. "Students' Expectations of the Economic Returns to College Education: Results of a Controlled Experiment." *Economics of Education Review*, 23:645-653.

Boyle, Richard P. 1966. "The Effect of the High School on Students' Aspirations." *The American Journal of Sociology*, 71(6): 628-639.

Card, David. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics Volume 3*, ed. Orley Ashenfelter and David Card, 2439-2483. New York: Elsevier.

Card, David and Alan B. Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *The Journal of Political Economy*, 100(1): 1-40.

Card, David and A. Abigail Payne. 2002. "School Finance Reform, the Distribution of School Spending, and the Distribution of Student Test Scores." *Journal of Public Economics*, 83: 49-82.

Carvajal, Manuel J. David Bendana, Alireza Bozorgmanesh, Miguel A. Castillo, Katayoun Pourmasiha, Priya Rao, and Juan A. Torres. 2000. "Inter-gender Differentials between College Students' Earnings Expectations and the Experience of Recent Graduates." *Economics of Education Review*, 19: 229-243.

Chen, Stacey H. 2008. "Estimating the Variance of Wages in the Presence of Selection and Unobserved Heterogeneity." *The Review of Economics and Statistics*, 90(2): 275-289.

Crawford, David L., Amy W. Johnson, and Anita A. Summers. 1997. "Schools and Labor Market Outcomes." *Economics of Education Review*, 16(3): 255-269.

Cunha, Flavio, James Heckman, and Salvador Navarro. 2005. "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." *Oxford Economic Papers*, 57: 191-261.

Dearden, Lorraine, Javier Ferri, and Costas Meghir. 2002. "The Effect of School Quality on Educational Attainment and Wages." *The Review of Economics and Statistics*, 84(1): 1-20.

Dolton, P.J. and A. Vignoles. 2002. "Is a Broader Curriculum Better?" *Economics of Education Review*, 21: 415-429.

Dominitz, Jeff and Charles F. Manski. 1996. "Eliciting Student Expectations of the Returns to Schooling." *The Journal of Human Resources*, 31(1): 1-26.

Farber, Henry S. 1999. "Mobility and Stability: The Dynamics of Job Change in Labor Markets." In *Handbook of Labor Economics Volume 3*, ed. Orley Ashenfelter and David Card, 2439-2483. New York: Elsevier.

Hurley, Dan and Jim Thorp, eds. 2002. *Decisions without Direction: Career Guidance and Decision-making among American Youth*. Big Rapids, MI: Ferris State University Career Institute for Education and Workforce Development.

Fersterer, Josef and Rudolf Winter-Ebmer. 2002. "Smoking, Discount Rates, and Returns to Education." *Economics of Education Review*, 22: 561-566.

Rivkin, Steven G., Erik A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica*, 73(2): 417-458.

Heckman, James, Anne Layne-Farrar, and Petra Todd. 1996. "Human Capital Pricing Equations with an Application to Estimating the Effect of School Quality on Earnings." *The Review of Economics and Statistics*, 78(4): 562-610.

Heckman, James, Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics*, 24(3): 411-482.

Hvide, Hans K. 2003. "Education and the Allocation of Talent." *Journal of Labor Economics*, 21(4): 945-976.

Ida, Takanori and Rei Goto. 2009. "Interdependency Among Addictive Behaviours and Time/Risk Preferences: Discrete Choice Model Analysis of Smoking, Drinking, and Gambling." *Journal of Economic Psychology*, 30: 608-621.

Jackson, Gregory A. 1982. "Public Efficiency and Private Choice in Higher Education." *Educational Evaluation and Policy Analysis*, 4(2): 237-247.

Jensen, Robert. Forthcoming. "The Perceived Returns to Education and the Demand for Schooling." *The Quarterly Journal of Economics*.

Kane, Thomas J. and Cecilia Elena Rouse. 1995. "Labor-Market Returns to Two- and Four-Year College." *American Economic Review*, 85(3): 600-614.

Kaufmann, Katja and Orazio Attanasio. 2009. "Educational Choices, Subjective Expectations, and Credit Constraints." National Bureau of Economic Research Working Paper 15087.

Leigh, Duane E. and Andrew M. Gill. 2004. "The Effect of Community Colleges on Changing Students' Educational Aspirations." *Economics of Education Review*, 23: 95-102.

Ludwig, Jens. 1999. "Information and Inner-city Educational Attainment." *Economics of Education Review*, 19: 17-30.

Manski, Charles F. 1989. "Schooling as Experimentation: A Reappraisal of the Postsecondary Dropout Phenomenon." *Economics of Education Review*, 8(4): 305-312.

Manski, Charles F. and David A. Wise. 1983. *College Choice in America*. Cambridge, MA: Harvard University Press.

Mishel, Lawrence and Richard Rothstein, eds. 2002. *The Class Size Debate*. Washington, D.C.: Economic Policy Institute.

Montmarquette, Claude, Sophie Mahseredjian, and Rachel Houle. 2001. "The Determinants of University Dropouts: A Bivariate Probability Model with Sample Selection." *Economics of Education Review*, 20: 475-484.

Nguyen, Trang. 2008. "Information, Role Models, and Perceived Returns to Education: Experimental Evidence from Madagascar." PhD diss. Massachusetts Institute of Technology.

Polachek, Solomon W. and John Robst. 1998. "Employee Labor Market Information: Comparing Direct World of Work Measures of Workers' Knowledge to Stochastic Frontier Estimates." *Labour Economics*, 5: 231-242.

Rose, Heather and Julian R. Betts. 2004. "The Effect of High School Courses on Earnings." *The Review of Economics and Statistics*, 86(2): 497-513.

Rosenbaum, James. 1998. "College-For-All: Do Students Understand What College Demands?" *Social Psychology of Education*, 2: 55-80.

Rouse, Cecilia Elena. 2004. "Low-Income Students and College Attendance: An Exploration of Income Expectations." *Social Science Quarterly*, 85(5): 1299-1317.

Smith, Herbert L. and Brian Powell. 1990. "Great Expectations: Variations in Income Expectations among College Seniors." *Sociology of Education*, 63(3): 194-207.

Steifel, Leanna, Amy Ellen Schwartz, Ross Robenstein, and Jeffrey Zabel. 2005. *Measuring Student Performance and Efficiency: Implications for Practice and Research*. Larchmont, NY: Eye on Education..

Streufert, Peter. 2000. "The Effect of Underclass Social Isolation on Schooling Choice." *Journal of Public Economic Theory*, 2(4): 461-482.

Topel, Robert H. and Michael P. Ward. 1992. "Job Mobility and the Careers of Young Men." *The Quarterly Journal of Economics*, 107(2): 439-479.

Willis, Robert J. and Sherwin Rosen. 1979. "Education and Self-Selection." *The Journal of Political Economy*, 87(5): S7-S36.

CHAPTER 2

COMBINATION CLASSES AND EDUCATIONAL ACHIEVEMENT

Abstract: This paper determines the effect of membership in a combination class on student achievement in first grade. I address the selection biases that may arise from implementing combination classes. In order to control for any systematic differences between schools that offer combination classes and those that do not, I conduct a within-school analysis using school fixed effects. I find little evidence of meaningful nonrandom assignment of teachers to combination classes. There is, however, evidence that first graders in 1-2 combinations are positively selected based on ability. Using a rich set of covariates, I am able to control for the variables influencing selection. Estimates of the effect of combination class membership in first grade on reading and general knowledge test scores are not significantly different from zero. The estimate of the effect on math scores for first graders in 1-2 combinations is positive and significant, indicating that they can be expected to outperform single-grade students by one-seventh of a standard deviation. This result is not sensitive to functional-form assumptions. In addition, I find no evidence that first graders in schools offering combination classes perform worse than first graders in schools that do not offer such classes. Therefore, I conclude that combination classes may be a Pareto-improving option for school administrators.

2.1 Introduction

The combination class, in which students from two adjacent grades are grouped within one classroom under one teacher, is an increasingly common method of classroom organization, yet has received little attention in the literature. The nationwide trend toward class-size reduction suggests that combination classes will only become more prevalent since they can be used to attain class-size goals by smoothing enrollment across grades. They are a cost-saving option, allowing schools to use fewer teachers and classrooms. If combination-class membership has a nonnegative effect on student outcomes, offering such classes is an attractive strategy for schools looking to save money without sacrificing educational quality.

Combination classes also offer another avenue besides age at school entry to assess the effect of relative age on student performance. The age-at-school entry literature focuses on students in similar learning environments and assesses the effect of relative and absolute age on student achievement and other outcomes. Relatively older students are consistently shown to perform better on reading and math tests.[21] Rather than being a relative age effect, more recent research has established that this is likely to be an absolute age effect, and that being relatively younger might actually lead to higher test scores.[22] This paper compares students of the same absolute age who are placed in different learning environments—single-grade and combination classes—in which curricula and teaching methods differ along with students' relative ages. I am asking a

[21] See Stipek (2002) for a detailed literature review.
[22] See, for example, Bedard and Dhuey (2006), Datar (2006), Black, Devereux, and Salvanes (2008), and Elder and Lubotsky (2006).

different question, but my results support the recent findings in the age-at-school-entry literature.

This paper also contributes to the small body of literature that directly addresses the effect of combination classes on student achievement. Within this literature, there is little consensus.[23] In addition to differential success in dealing with nonrandom selection, prior studies have not distinguished between relatively older students (those in the higher grade of the combination class) and relatively younger students (those in the lower grade), even though treatment systematically differs along this dimension.

I seek to determine the effects of membership in K-1 and 1-2 combination classes on student achievement in first grade, as measured by test scores from the spring of the first-grade year. I address the selection biases that may arise when implementing combination classes. In order to control for any systematic differences between schools that offer combination classes and those that do not, I conduct a within-school analysis of schools offering combination classes using school fixed effects. I find little evidence of meaningful nonrandom assignment of teachers to combination classes, indicating that differences in outcomes are not due to differences in teacher quality. There is some evidence that first graders in 1-2 classes are positively selected based on ability. Using a rich set of covariates, however, I am able to effectively control for the variables influencing selection. Estimates of the effect of combination class membership in first grade on reading and general knowledge test scores are not significantly different from zero. The estimate of the effect on math scores for first graders in 1-2 combination

[23] See Veenman (1995) for a meta-analysis, and Hill and Rowe (1998) and Sims (2008) for more recent studies.

classes is positive and significant, indicating that 1-2 students—that is, students who are young relative to their classmates—can be expected to outperform their single-grade peers by one-seventh of a standard deviation. In addition, I find no evidence that first graders in schools offering combination classes perform worse than first graders in schools that do not offer such classes, indicating that offering combination classes may be a Pareto-improving option for school administrators.

This paper proceeds as follows. Section 2.2 discusses the conceptual framework underlying the estimation of combination-class effects. Section 2.3 describes the rich data set used in the analysis. Section 2.4 documents the selection issues that arise at the school, classroom, and student level when a school chooses to offer combination classes. Section 2.5 presents the main results of the paper and discusses some robustness checks. Section 2.6 demonstrates that first graders in schools offering combination classes do not seem to perform worse than first graders in schools that do not offer combination classes, and Section 2.7 concludes.

2.2 Conceptual Framework

2.2.1 Class Type as an Input in an Education Production Function

Student achievement as measured by test scores is a function of many variables. A child's performance in elementary school depends on the characteristics of the child's home, such as socioeconomic status (SES) and parental involvement. School performance also depends on the child's own characteristics, such as scholastic ability, past educational experience, behavior, and motivation. In addition, performance hinges on attributes of the student's school, such as demographics, school resources, and calendar type (year-round or traditional nine-month calendar). Finally, class characteristics such as teacher quality, curriculum, and classroom organization influence student achievement.

In this paper, I am interested in the effect of classroom organization in first grade on test scores—specifically, whether the class is K-1 combination, a 1-2 combination, or a single-grade first-grade class. If schools randomly chose to offer combination classes, and if teachers and students were randomly assigned to combination classes, a simple linear regression of first-grade test scores on dummy variables for K-1 and 1-2 combination-class membership would yield estimates of combination-class treatment effects that could be interpreted causally.

In the following two subsections, I discuss what exactly the combination-class treatment entails and the obstacles to the causal interpretation of the coefficients on combination-class dummy variables that emerge under nonrandom assignment.

2.2.2 The Combination-Class Treatment

Combination classes differ from single-grade classes on several dimensions. Some of these differences are inherent to combination classes and would exist even if schools randomly decided to offer combination classes. It is the effect of these inherent characteristics that I would like to isolate.

First, the age span within a combination class is wider than within a single-grade class. For instance, if a kindergarten class contains five- and six-year-olds, and a first grade class contains six- and seven-year-olds, a K-1 combination would contain children aged five to seven. Evidence on the effect of age diversity within a classroom is inconclusive.[24]

A related but separate characteristic of the combination-class treatment is that students are systematically placed within this wider age range so that they end up as relatively older or relatively younger than their classmates. First graders in K-1 classes are relatively old, and first graders in 1-2 classes are relatively young. Because of the wider age span in combination classes, these relative age differences are more pronounced than in single-grade classes. Elder and Lubotsky (2006) show that having older classmates tends to raise reading and math achievement, conditional on the student's own age. This is in contrast to earlier findings within the age-at-school entry literature that relatively older students do better. This literature considers students within one type of class whose ages are different because of school entry cutoff dates. Instead, I am looking at students who are the same age in different types of classrooms, where relative age depends on the type of combination class in which the student is placed.

In addition to their relative age differences, first graders in K-1 and in 1-2 combination classes are likely to experience different teaching methods and curricula than students in single-grade classes. In a survey of 35 combination-class teachers in California, Mason, Burns, and Armesto (1993) find that teachers tend to use a mixed approach in combination classes, in which the teacher separates students by grade level for certain subjects such as math and reading and uses large-group instruction for subjects such as science and social studies. We can assume that the large-group curriculum in a K-1 combination class will be aimed at a lower level than the large-group curriculum in a

[24] See, for example, Miller (1995).

1-2 combination class. In this way, the combination class effect will differ depending on a student's relative grade level within the class. At first glance, the mixed approach would seem to have a positive effect on first graders in a 1-2 combination and a negative effect on those in a K-1 combination, relative to the performance of single-grade first graders. It is not inconceivable, however, that first graders in a K-1 combination would benefit from the review of kindergarten concepts and do better in a K-1 combination class than they would have in a single-grade class.

2.2.3 Confounding Factors Resulting from Nonrandom Assignment of Students and Teachers to Combination Classes

Combination classes differ from single-grade classes in other ways if there is nonrandom assignment to combination classes. First, schools that decide to offer combination classes may be systematically different from those that do not. For example, multi-track year-round schools may have a small number of students per grade level and choose to offer combination classes in order to use fewer classrooms. Calendar type may have an effect on student achievement apart from its association with combination classes. If a year-round calendar has a negative effect on student achievement, as shown in Graves (2007), the combination-class effect would be biased downwards.

Second, the resources available to students in combination classes may be different from those available to single-grade students. If combination classes are systematically larger, for example, students may be adversely affected. Teaching quality may also differ by class type. In a survey of 72 school principals in California, Burns, Mason, and Demiranda (1993) find that many select only the best teachers for combination classes. If this is indeed the general selection criterion, the positive effect of the teachers' skill will bias estimates of the combination-class effect upwards.

Finally, students are assigned to combination classes. The main reason for nonrandom assignment is to make these classes more attractive to teachers (Mason et al., 1993). Generally, the goal is to make student ability more homogeneous than it would be under random assignment, or to populate the class with independent workers. First graders placed in combination classes are likely to be positively selected on behavior in all cases. Selection on ability is likely to be positive for 1-2 placement and negative for K-1 placement. I document the selection that occurs on all three levels and discuss how I deal with selection at each level in Section 2.4. In the next section, I discuss the data source used in this paper.

2.3 Data

The Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K) Restricted-Use Data Set is an ongoing study focusing on children's early school experiences. It has a rich set of student-, classroom-, and school-level variables, allowing me to determine what factors influence a school's decision to offer combination classes and to analyze the teacher and student characteristics that influence assignment, as well as measure the effect of combination class membership on test scores.

In this study, I use data collected in the spring of the children's kindergarten year and the spring of their first-grade year. Spring first-grade reading, math, and general knowledge standardized test scores are the outcome variables of interest. Spring kindergarten scores are prior test score controls. I use a variety of child-, classroom-, and school-level controls: child characteristics (gender, race and ethnicity, age), family background variables (SES, home language), teacher characteristics (race and ethnicity, education, experience), classroom characteristics (demographics, student performance, classroom activities, age distribution, class size), and school characteristics (location, calendar type, percent minority students, percent of students eligible for free lunch). In addition, the ECLS-K contains behavior measures that are typically unobservable to the econometrician. Finally, I match schools to the National Center for Education Statistics (NCES) Common Core of Data (CCD) to obtain data on enrollment by grade level for the 1999-2000 school year.

I restrict the sample to public-school first graders, and only include students who were first-time kindergarteners in the 1998-1999 school year and remain in the dataset as first graders in the 1999-2000 school year, resulting in a sample of 10,640 students. I further restrict the sample to students whose first-grade class type can be accurately identified, resulting in a final sample of 9,339 individuals.

2.4 Selection Issues

In order to identify the causal effect of combination-class membership, one must address the selection that occurs when combination classes are offered. First, I discuss school-level selection, then teacher- and student-level selection.

2.4.1 School-Level Selection

Are schools that offer combination classes systematically different from those that do not? 17 percent of the public elementary schools sampled in the ECLS-K offer some type of combination (K-1 or 1-2) or multi-grade (K-1-2, 1-2-3, K-1-2-3, etc.) class. Table 2.1 contains a breakdown of schools by the types of first-grade classes they offer. Schools that offer some type of combination or multi-grade class fall into two broad categories. 92 schools offer single grade first-grade classes and one or both combination classes (K-1, 1-2, or K-1 and 1-2), or the two combination classes only. I will call these "combination-class schools."

76 schools offer K-1 and 1-2 classes only (that is, no single-grade classes), or offer first and K-1 or 1-2 along with some other type of multi-grade class. These "multi-grade schools" appear to have so few students per grade level that their only option is to combine grades, as in a one-room schoolhouse. In the analysis that follows, I drop students from multi-grade schools and consider the sample of students attending combination-class schools only. In doing so, I am able to divide students cleanly into three groups of first graders: those in a single-grade class, those in a K-1 combination, and those in a 1-2 combination. In addition, I am able to focus on schools in which the only first-grade class options come from this set of possible class types.

Table 2.2 compares combination-class schools to those offering only single-grade classes. Combination-class schools seem to base the decision to offer these classes on classroom constraints (i.e., crowding) and school calendar type. Indeed, these 92 schools are more likely to have a year-round calendar than the schools offering only single-grade classes. Classroom constraints may be a function of school calendar type, especially if the school operates on a multi-track year-round calendar. In this type of school, the student body and staff are divided into three to five tracks. At any one time, all but one of the tracks is attending school and the last track is on vacation.

The ECLS-K does not reveal if a school is a single- or a multi-track year-round school. However, multi-track year-round schools are fairly common among year-round schools. In California, for example, which is home to 44 percent of year-round schools nationwide (National Association for Year-Round Education, 2007), 48 percent of year-round schools use a multi-track calendar (California Department of Education, Statistical Summary of Year-Round Programs, 2005-2006). Burns, Mason, and Demiranda (1993) find that multi-track principals are constrained in their assignment of students to different types of classes since there are relatively low numbers of students in each grade level per track, and principals may have little choice but to combine adjacent grades into a combination class.

Combination-class schools differ on other dimensions as well. They are more likely to be in the West. They have a higher percentage of minorities and larger average enrollments in grades K through two. They also have a significantly lower number of full-time equivalent (FTE) teachers per student. This could be both a cause and an effect of combination classes. An overcrowded school is more likely to switch to a multi-track year-round calendar, which in turn may lead to the adoption of combination classes. On the other hand, one of the intended results of combination classes is that students from two grade levels are combined into one class, necessitating one teacher instead of two and lowering the teacher-pupil ratio.

Combination-class schools appear to be more disadvantaged than single-grade schools, which could bias estimates of the combination-class effect downwards if these differences are not addressed. Sims (2008) finds that second and third graders in schools with a higher percentage of students in combination classes perform worse than second and third graders in schools with fewer combination-class students. He uses an instrumental variables technique to account for the school's decision to offer combination classes, but shows that his instrument is correlated with observable school characteristics. Because it may also be correlated with unobservable school characteristics, the estimates in Sims (2008) may be biased downwards. In order to avoid any school-level bias, I make my sample of schools as homogeneous as possible by considering only combination-class schools. I address any additional systematic, school-level differences by focusing on within-school differences between combination- and single-grade classes using school fixed effects.

2.4.2 Classroom-Level Selection

Within combination-class schools, are K-1 or 1-2 teachers systematically different from single-grade teachers? Do K-1 or 1-2 classes systematically differ from single-grade classes? This section answers these questions.

The teacher-level variables I analyze are as follows: gender, race and ethnicity, experience, education, job satisfaction (enjoys present teaching job, believes teacher makes a difference in children's lives, and would choose teaching again),[25] and paid and unpaid preparation hours per week. Table 2.3 contains the means of each of these variables by class type. The sample is restricted to teachers within combination-class schools. Of these teachers, 293 teach single-grade first, 47 teach K-1 combinations, and 99 teach 1-2 combinations.

[25] The job satisfaction variables contain teachers' responses on a five-point Likert scale in which one = "strongly disagree" and 5 = "strongly agree."

Comparing K-1 teachers to single-grade teachers, K-1 teachers are slightly less likely to be white but are similar to single-grade teachers in other respects. Comparing 1-2 teachers to single-grade teachers, 1-2 teachers are less likely to be male, more likely to be white, less likely to be Hispanic and more likely to be Asian. 1-2 teachers also appear to be happier with their career choice than single-grade teachers, answering more positively to the question of whether they would choose teaching again. This could be either a result of their experience teaching a 1-2 combination, or a reason for their assignment to such a class. The latter would result in an upward bias in estimating the effect of 1-2 membership on test scores. Because of the direction-of-causality problem and since the difference in means is only significant at the ten percent level, I ignore this possibility in the analysis that follows.

A more complete indication of nonrandom selection of teachers is to see if the teacher-level variables, taken together, influence assignment to combination classes within schools. I model the selection of teachers using a simple linear model that includes school fixed effects.[26] All of the job satisfaction variables have a direction-of-causality problem, so I do not include them in the following model:

$$\begin{aligned} class\_type_i &= \beta_1 male_i + \beta_2 black_i + \beta_3 hispanic_i + \beta_4 asian_i + \beta_5 other_i \\ &\quad + \beta_6 yrs\_teach_i + \beta_7 some\_grad_i + \beta_8 grad\_dgr_i + \sum_{j=1}^{J} \delta_j s_j + \varepsilon_i \end{aligned} \quad (1)$$

Teachers are indexed by *i*, schools by *j*. I run two separate regressions. In the first, $class\_type_i$ equals one if teacher *i* teaches a K-1 class. In this regression, I restrict my sample to teachers within combination-class schools offering only single-grade and K-1 combination classes. Recall from Table 2.1 that this is the second most common

[26] Probit results produce similar marginal effects.

type of combination-class school. Thus, $class\_type_i$ equals zero if teacher i teaches a single-grade class within this type of school.

In the second regression, $class\_type_i$ equals one if teacher i teachers a 1-2 class, and the sample is restricted to teachers within combination-class schools offering only single-grade and 1-2 combination classes. This is the most common type of combination-class school. The independent variables in both regressions are the observed teacher characteristics: gender, ethnicity, experience, and education, as well as school fixed effects.

Table 2.4 contains the results. Though K-1 teachers are less likely to be male and more likely to be Hispanic or Black, and 1-2 teachers are more likely to be Asian, there appears to be no evidence of meaningful selection on observables—the coefficients on years of teaching experience and the education dummies are not significant individually or jointly in either regression.[27] This lack of evidence on selection based on experience and education suggests that nonrandom assignment of teachers is not a source of bias in the outcome regressions in Section 2.5.

In addition to comparing teachers by class type, I compare the following classroom characteristics: size, percent boys, percent minority, percent gifted, percent limited English proficiency, percent below grade level in reading and math, age distribution, and teaching methods (use of whole-class, small-group, or individual activities).

---

[27] $F(3,82) = 0.90$, p-value = 0.446 in the K-1 regression; $F(3,176) = 0.67$, p-value = 0.574 in the 1-2 regression.

Table 2.5 contains the means of classroom-level variables obtained from regressions on dummies for K-1 class and 1-2 class and school fixed effects. Single-grade classes form the base case. We observe the obvious differences in age distribution: K-1 classes are younger and 1-2 classes are older than single-grade classes. There is debate about the effect of class size on student achievement,[28] but in any case, combination classes do not differ from single-grade classes along this dimension. This lack of variation could be due to the fact that a plurality of the students in the sample (31%) lives in California, which implemented its Class Size Reduction Act in the 1996-1997 school year, giving financial rewards to schools that reduced class size in grades K-3 to 20 students or fewer. By the 1999-2000 school year, 99% of first graders were in classes of 20 or fewer students (California Department of Education, 2009).

Teaching methods also differ according to Table 2.5. Teacher-directed whole-class and individual activities are less common in K-1 than in single-grade classes. Child-selected activities are more common in K-1 classes than they are in single-grade or 1-2 classes. Differences in teaching methods are part of the combination-class treatment effect that I want to estimate. Class composition, however, also differs, and this is a result of nonrandom selection—a confounding factor that could bias estimates of the combination-class effect. 1-2 classes contain more gifted students, which points to the possibility of positive peer selection (though these classroom-level data do not specify if the gifted students are first- or second-graders). Positive selection of peers into combination classes will bias estimates upwards. In order to address this source of bias, I

[28] See, for example, Hoxby (2000) or Mishel and Rothstein (2002).

run outcome regressions in Section 2.5 including average peer ability as measured by kindergarten test scores as a partial control.

2.4.3 Student-Level Selection

In this section, I analyze student-level variables to determine if there is positive or negative selection into combination classes. The student-level variables are as follows: sex, age, ethnicity, home language, SES, kindergarten behavior measures, and kindergarten and first-grade math, reading, and general knowledge standardized test scores.

Behavior is a typically unobservable determinant of student achievement, but the ECLS-K contains several behavior measures. Students' kindergarten teachers rated their behavior along five dimensions. The Approaches to Learning Scale measures behaviors that affect the ease with which children can benefit from the learning environment. The Self-Control Scale has four items that indicate the child's ability to control behavior. The five Interpersonal Skills items rate the child's ability to get along with others. The Externalizing Problem Behaviors scale rates the frequency with which a child acts out, and the Internalizing Problem Behavior Scale asks about the apparent presence of anxiety, loneliness, low self-esteem, and sadness.

Table 2.6 contains means of student-level variables by class type. K-1 students are more likely to be Hispanic than single-grade students, more likely to internalize problem behaviors, and have lower kindergarten and first-grade reading scores. 1-2 students are more likely to be white and less likely to be black or speak a language other than English than single-grade students. In addition, they appear to be positively selected on behavior and prior test scores—they are better behaved and have higher kindergarten test scores than K-1 or single-grade students. They also have higher first-grade test scores, which could be a result of a 1-2 treatment effect or of positive selection. Note that Table 2.6 does not take school fixed effects into account—there may be systematic differences across schools that offer 1-2 classes and schools that offer K-1 classes that are being picked up in these average child characteristics.

In order to consider the joint effect of these variables on the assignment to combination classes within schools, I model student selection using school fixed effects. The model is as follows:

$$\begin{aligned} class\_type1_i = \beta_0 + \beta_1 male_i + \beta_2 black_i + \beta_3 hispanic_i + \beta_4 asian_i + \beta_5 other_i \\ + \beta_6 non\_eng_i + \beta_7 ses_i + \beta_8 learnK_i + \beta_9 controlK_i + \beta_{10} personalK_i \quad (2) \\ + \beta_{11} externK_i + \beta_{12} internK_i + \beta_{13} readK_i + \beta_{14} mathK_i + \beta_{15} genK_i \\ + \sum_{j=1}^{J} \delta_j s_j + \varepsilon_i . \end{aligned}$$

Students are indexed by i, schools by j; s_j is a school fixed effect. Background characteristics and kindergarten test scores and behavior measures are used as predictors of first-grade class type I run two separate regressions—one for schools offering only single-grade and K-1 classes, and one for schools offering only single-grade and 1-2 classes, as in the previous subsection. In the first, $class\_type1_i$ equals one if the student is in a K-1 combination; in the second, $class\_type1_i$ equals one if the student is in a 1-2 combination.

Regression results are contained in Table 2.7. There is little evidence for selection into K-1 classes. K-1 students are more likely to internalize problem behaviors than their single-grade counterparts, but F-tests of the joint significance of kindergarten test scores and behavior measures fail to reject the null hypothesis.[29]

The table gives mixed evidence for selection into 1-2 classes. 1-2 students have significantly higher kindergarten math scores but appear to be less well behaved than single grade students—are they placed into 1-2 classes because they would be bored in a single-grade class? Considering the results of F-tests of joint significance of the test score and behavior measures, however, we see strong evidence that high-achieving first graders are assigned to 1-2 classes. An F-test of kindergarten behavior measures alone fails to reject the null hypothesis, but F-tests of kindergarten test scores alone and with the behavior measures show that these variables are jointly significant.[30]

This positive selection will bias estimates of the combination-class effect upwards unless I can control for the variables influencing class assignment. Including prior-year test scores and behavior measures in the outcome regressions, discussed below, seems to accomplish this and allows me to estimate a coefficient that can be interpreted causally.

2.5 Results

In this section, I discuss the results from four outcome-regression models. The dependent variables are first-grade reading, math, and general knowledge test scores. I run one regression per test score for a total of three regressions per model. The independent variables differ by model, but all include school fixed effects. Model 1

[29] Testing the joint significance of kindergarten behavior measures, I obtain $F(5, 235) = 1.20$, p-value = 0.312. Testing kindergarten test scores, I obtain $F(3, 235) = 0.42$, p-value = 0.70. Testing behavior measures and test scores, I obtain $F(8, 235) = 1.05$, p-value = 0.402.
[30] Testing the joint significance of kindergarten social rating scores, I obtain $F(5, 521) = 1.30$, p-value = 0.261. Testing kindergarten test scores, I obtain $F(3, 521) = 3.76$, p-value = 0.011. Testing social rating and test scores, I obtain $F(8, 521) = 2.26$, p-value = 0.022.

contains only dummies for class type, with single-grade classes being the omitted category. Model 2 contains class-type dummies as well as the student background characteristics sex, age, ethnicity, home language, and SES. Model 3 contains combination-class dummies, background characteristics, and kindergarten test scores. Finally, Model 4 contains class-type dummies, background characteristics, kindergarten test scores, and kindergarten behavior measures. This information is summarized in Table 2.8.

Table 2.9 contains the coefficients on the K-1 and 1-2 dummies from each of the three regressions in each of the four models. As controls are added, the coefficient on K-1 membership moves from negative to positive but is insignificant at the 5% level in all cases (it is significant at the 10% level in the Model 2 regression of first-grade reading test scores on combination-class dummies and student characteristics). The coefficient on 1-2 membership shrinks as controls are added, but retains significance for math scores when the full set of controls is used.

The 1-2 coefficient of 1.3 in the Model 4 regression of first-grade math scores on the combination class dummy, student characteristics, kindergarten test scores, and kindergarten behavior measures indicates that 1-2 membership is associated with nearly a two-percentile-point gain in first-grade math test scores relative to single-grade students. This is approximately one-seventh of a standard deviation. Interpreting this causally could be problematic due to the positive selection of 1-2 students documented in the previous section. Note, however, that the signs, magnitudes, and significance levels of the class-type dummies do not change much between Model 3 and Model 4 in the math test score regressions. Nor do the adjusted R-squared values change substantially between these two models. The difference between them is that the Model 4 contains child social rating scores that proxy for qualities such as behavior and motivation that are usually unobservable to the econometrician. That signs, magnitudes, significance levels, and adjusted R-squared values do not change substantially between Models 3 and 4 is one indication that kindergarten test scores are a good proxy for ability and other usually unobservable characteristics such as behavior and motivation, and allows me to conclude that I have adequately controlled for selection bias.[31]

In order to address the possible upward bias from possible peer selection, I run the Model 4 regression including average peer test scores as independent variables. The signs, magnitudes, and significance levels of the coefficients are nearly identical,[32] indicating that peer effects are not a source of bias.

In the models discussed above, I have imposed a linear relationship on student characteristics and test scores. In order to determine whether my results are sensitive to this linear structure, I re-estimate the combination-class treatment effect using propensity score matching. I find that none of these estimates is significant. The coefficient of interest, however, (the effect of being in a 1-2 combination class on first-grade math scores) is 1.7, slightly larger than the OLS estimate,[33] indicating that the linear model does not overstate the effect of being in a 1-2 class on math scores.

[31] In addition, I tried several different specifications of these models in which I included behavior measures first, then added test scores. (Please see Appendix 2.1 for detailed results). In the model with background characteristics and behavior measures but no test scores, the coefficient on the 1-2 dummy in the reading regression was 1.372 and significant at the 10 percent level; otherwise, estimates were qualitatively similar to Model 3, above. The adjusted R-squared values were approximately 0.4, however—much smaller than in Model 3, indicating that including test scores without behavior measures adds more information than including behavior measures without test scores.
[32] Appendix 2.1 contains these results.
[33] These results are contained in Appendix 2.1.

2.6 Overall Impact

Sims (2008) finds that children in schools with a higher percentage of students in combination classes perform worse than children in schools with fewer combination-class students. This could be because, once combination classes are implemented, single-grade students do worse than they would have if the school had not implemented combination classes, perhaps because resources are diverted to the combination classes and away from single-grade classes. In this case, my finding that first graders in 1-2 classes outperform their single-grade peers could be explained by single-grade students doing worse than they would have had the school not implemented combination classes.

Addressing the question of whether 1-2 students benefit at the expense of other first graders is difficult, however, because schools that offer combination classes are quite different from schools that do not. One way to address this would be to regress test score outcomes on a dummy for whether the school offers combination classes, kindergarten test scores, kindergarten behavior measures, student background characteristics, as well as school-level controls. If the school-level controls accounted for all the relevant differences between schools that choose to offer combination classes and those that do not, the coefficient on the school-type dummy could be interpreted as the causal effect of offering combination classes on first-grade test scores.

As a rudimentary check that, overall, first graders are not harmed by a school's decision to offer combination classes, I compare single-grade schools to schools offering first grade and a 1-2 combination by regressing first-grade test scores on a dummy indicating that the school offers single-grade and 1-2 classes. As other independent variables, I include the student-level variables from Model 4, as well as the following school-level covariates: indicators for region, community size and year-round school, average grade-level enrollment, standard deviation of enrollment across grades, full-time equivalent teachers per student, percent minority, and percent eligible for free lunch. Table 2.10 contains the coefficients of interest (please see Appendix 2.1 for full regression results).

None of the coefficients is significant, though the point estimates are negative for reading and math test scores. As discussed above, schools that offer combination classes tend to be larger, have a higher percentage of minority students, and have fewer teachers per student than schools that do not offer combination classes. That is, combination-class schools tend to be more disadvantaged than single-grade schools. To the extent that this is true for unobservable school characteristics influencing the choice to offer combination classes and student outcomes, we can assume that these coefficients are biased downwards. This reinforces the conclusion that, overall, the decision to offer combination classes, at least at the 1-2 level, does not harm first graders overall. Thus, combination classes may be a Pareto-improving option for school administrators.

2.7 Conclusion

In this paper, I document the selection issues that arise at the school, classroom, and student level when a school chooses to offer combination classes. To address school-level selection, I limit the sample to combination-class schools and use school fixed effects in the outcome regressions. To address teacher-level selection, I model teacher assignment to combination classes and find little evidence for meaningful nonrandom selection. To address student-level selection, I model student assignment to combination classes and find evidence for positive selection into 1-2 classes. I therefore use a rich set of control variables, including behavior measures that are usually unavailable to the econometrician, to more plausibly assume ignorability of treatment and estimate the causal effect of combination-class membership in first-grade on first-grade test scores.

I find that there is no effect on reading or general knowledge scores for students in either type of class, but that 1-2 combination-class membership is associated with an increase of one-seventh of a standard deviation on math test scores relative to single-grade students. This result is not sensitive to functional form assumptions. In addition, I find little evidence that 1-2 students benefit at the expense of other first graders.

These results indicate that combination-class membership in first grade has at worst, no effect and at best, a small positive effect on student achievement as measured by test scores. I conclude that combination classes may be a Pareto-improving option for school administrators. Given that more and more states are implementing class-size reduction initiatives, and that combination classes conserve scarce resources by allowing schools to use fewer teachers and classrooms, it is more important than ever for school administrators to find ways to reduce class size in the least costly manner. Combination classes allow school administrators to reduce class size within one grade while smoothing class size across grades, and should be considered a viable means of classroom organization.

It should be acknowledged, however, that implementing combination classes is problematic for other reasons. Teachers do not like them (Mason, Burns, and Armesto, 1993), though some of the cost savings could be used to compensate teachers for this. In addition, parents may not want their children to be placed in the higher grade of a combination class because they perceive this as a signal that their children are low achievers, even though the data I have presented here indicate that K-1 students are statistically indistinguishable from their single-grade counterparts.

This paper shows that 1-2 (i.e., lower-grade) students benefit from combination-class membership. These students are relatively young compared to their classmates, and this result supports recent findings in the age-at-school entry literature that relatively younger students benefit from having older peers. An interesting direction for future research would be to determine if lower-grade students in other combination classes (e.g., third graders in a 3-4 class) also benefit, and if these benefits persist over time.

Chapter 2 has been submitted for publication of the material as it may appear in the *Economics of Education Review*. The dissertation author was the sole author of this paper.

Table 2.1: Types of First-Grade Classes Offered

Combination and/or multi-grade class offering	No. of schools	Pct. of schools
Single-grade first grade class only	845	0.834
Any type of combination or multi-grade class	168	0.166
Combination-class schools	92	0.091
First, 1-2	55	0.054
First, K-1	26	0.026
First, K-1, 1-2	9	0.009
K-1, 1-2	2	0.002
Multi-grade schools	76	0.075
First, other	39	0.038
K-1 only	11	0.011
1-2 only	8	0.008
Other only	8	0.008
First, 1-2, other	6	0.006
1-2, other	2	0.002
First, K-1, 1-2, other	1	0.001
First, K-1, other	1	0.001
K-1, 1-2, other	0	0.000
K-1, other	0	0.000

Table 2.2: Comparison of Combination-Class Schools to Schools Offering Single-Grade Classes Only

School characteristic	Schools offering single-grade classes only	Combination-class schools
West	0.206 (0.014)	0.446 (0.052)***
FTE teachers per student	0.061 (0.000)	0.057 (0.001)***
Northeast	0.187 (0.013)	0.076 (0.028)***
Midwest	0.239 (0.015)	0.130 (0.035)**
Percent minority	41.722 (1.209)	50.244 (3.406)**
Year-round	0.041 (0.009)	0.095 (0.032)**
Average grade-level enrollment over grades K-2	90.111 (1.566)	98.417 (4.661)*
Std. dev. of grade-level enrollment over grades K-2	10.871 (0.327)	9.210 (0.581)*
Total enrollment	558.743 (8.929)	598.761 (27.487)
Average grade-level enrollment over all grades	87.356 (1.457)	93.739 (4.173)
Suburb	0.398 (0.017)	0.337 (0.050)
City	0.396 (0.017)	0.457 (0.052)
Town	0.086 (0.010)	0.109 (0.033)
Rural	0.120 (0.011)	0.098 (0.031)
South	0.368 (0.017)	0.348 (0.050)
Standard deviation of grade-level enrollment over all grades	18.626 (0.568)	19.005 (1.877)
Percent of students eligible for free lunch	33.705 (1.246)	33.623 (4.042)

Note: This table contains the results of a two-sample Student's t-test assuming equal variances. * denotes that the means are significantly different at the 10% level, ** at the 5% level, and *** at the 10% level.

Table 2.3: Means of Teacher Characteristics by Class Type

Teacher characteristic	Single-grade 1st mean	K-1 mean	1-2 mean
Male	0.049 (0.013)	0 (0)	0.010 (0.010)*
White	0.761 (0.025)	0.638 (0.071)*	0.835 (0.038)†††
Black	0.035 (0.011)	0.085 (0.041)	0.010 (0.010)††
Hispanic	0.165 (0.022)	0.213 (0.060)	0.082 (0.028)**,††
Asian	0.021 (0.009)	0.043 (0.030)	0.062 (0.025)**
Other	0.018 (0.008)	0.021 (0.021)	0.010 (0.010)
Years teaching	11.846 (0.564)	13.553 (1.431)	12.402 (0.953)
B.A. or less	0.226 (0.025)	0.25 (0.066)	0.245 (0.045)
Some graduate school	0.373 (0.029)	0.295 (0.070)	0.394 (0.051)
Graduate degree	0.401 (0.029)	0.455 (0.076)	0.362 (0.050)
Enjoys present teaching job	4.358 (0.044)	4.447 (0.109)	4.371 (0.094)
Makes a difference	4.503 (0.036)	4.574 (0.073)	4.618 (0.056)
Would choose teaching again	4.292 (0.057)	4.319 (0.140)	4.484 (0.090)*
Paid prep hours per week	1.906 (0.049)	1.804 (0.115)	1.889 (0.070)
Unpaid prep hours per week	3.613 (0.060)	3.362 (0.123)	3.793 (0.098)†††

Notes: I consider only teachers in combination-class schools. Of these, 293 teach single-grade first, 47 teach K-1 combinations, and 99 teach 1-2 combinations. Standard errors are in parentheses. * denotes that the K-1 or the 1-2 mean is different from the single-grade mean at the 10% level, ** at the 5% level, and *** at the 1% level. † denotes that the 1-2 mean is different from the K-1 mean at the 10% level, †† at the 5% level, and ††† at the 1% level.

Table 2.4: Modeling Teacher Selection

Teacher characteristic	Regression 1: K-1 combination dummy as dependent variable	Regression 2: 1-2 combination dummy as dependent variable
Male	-0.365 (0.158)**	-0.303 (0.189)
Black	0.536 (0.186)***	-0.153 (0.216)
Hispanic	0.232 (0.138)*	-0.094 (0.135)
Asian	0.458 (0.360)	0.482 (0.260)*
Other	0.451 (0.395)	0.237 (0.341)
Years teaching	0.006 (0.005)	0.005 (0.197)
Some graduate school	-0.151 (0.155)	-0.049 (0.676)
Graduate degree	-0.107 (0.148)	-0.116 (0.335)
Constant	0.202 (0.137)	0.285 (0.088)**
Regression statistics	Regression 1	Regression 2
Number of obs.	116	239
p-value of F statistic	0.025	0.264
Adj. R-squared	-0.061	-0.132

Notes: Table 2.4 contains the results of two linear regressions of class-type dummies on teacher characteristics. Both regressions include school fixed effects. In Regression 1, the sample is restricted to the 26 schools offering only single-grade first and K-1 classes. In Regression 2, the sample is restricted to the 55 schools offering only single-grade first and 1-2 classes. Robust standard errors are in parentheses.
a) F(8, 82) for the K-1 regression; F(8, 176) for the 1-2 regression.

Table 2.5: Within-school Means of Classroom Characteristics by Class Type

Classroom characteristic	Single-grade 1st mean	K-1 mean	1-2 mean
Class size	20.800 (0.169)	20.288 (0.733)	21.338 (0.669)
Percent boys	0.512 (0.006)	0.513 (0.026)	0.509 (0.016)
Percent minority	53.949 (0.822)	52.996 (2.596)	51.663 (2.149)
Percent gifted	0.016 (0.006)	0.053 (0.024)	0.082 (0.026)**
Percent limited English proficiency	0.395 (0.029)	0.315 (0.056)	0.283 (0.070)
Percent reading below grade level	0.269 (0.011)	0.315 (0.046)	0.252 (0.025)
Percent math below grade level	0.185 (0.009)	0.238 (0.040)	0.193 (0.020)
Teacher-directed whole class activity	3.966 (0.044)	3.561 (0.151)***	3.795 (0.112)
Teacher-directed small group activities	3.550 (0.050)	3.670 (0.140)	3.654 (0.108)
Teacher-directed individual activities	2.794 (0.052)	2.458 (0.139)**	2.689 (0.120)
Child-selected activities	2.483 (0.042)	2.873 (0.114)***	2.589 (0.091)†
Percent 5 years or younger	0.001 (0.003)	0.150 (0.025)***	0.0001 (0.009)†††
Percent 6 years old	0.406 (0.011)	0.495 (0.027)***	0.201 (0.025)***,†††
Percent 7 years old	0.558 (0.011)	0.352 (0.033)***	0.475 (0.027)***,†††
Percent 8 years old	0.035 (0.005)	0.002 (0.013)**	0.299 (0.022)***,†††
Percent 9 years old	0.0001 (0.0001)	-0.0001 (0.004)	0.024 (0.007)***,†††
Percent 10 years or older	0.0001 (0.0001)	0.0001 (0.0002)	0.001 (0.001)

Notes: Table 2.5 contains the results of regressions of each of the classroom-level variables on dummies for K-1 class and 1-2 class and school fixed effects. Single-grade first grade classes form the base case. Standard errors are in parentheses. * denotes that the K-1 or the 1-2 mean is different from the single-grade mean at the 10% level, ** at the 5% level, and *** at the 1% level. † denotes that the 1-2 mean is different from the K-1 mean at the 10% level, †† at the 5% level, and ††† at the 1% level.

Table 2.6: Means of Student Characteristics by Class Type

Student characteristic	Single-grade 1st mean	K-1 mean	1-2 mean
Male	0.513	0.458	0.489
	(0.016)	(0.051)	(0.033)
Age in months (Spring 1st)	85.874	86.066	86.103
	(0.135)	(0.419)	(0.264)
White	0.460	0.417	0.529
	(0.016)	(0.051)	(0.033)*,†
Black	0.130	0.094	0.08
	(0.011)	(0.030)	(0.018)**
Hispanic	0.290	0.375	0.258
	(0.015)	(0.050)*	(0.029)††
Asian	0.063	0.031	0.076
	(0.008)	(0.018)	(0.018)
Other	0.057	0.083	0.058
	(0.008)	(0.028)	(0.016)
Language other than English spoken at home	0.222	0.161	0.146
	(0.014)	(0.038)	(0.024)**
SES	-0.119	-0.189	-0.048
	(0.027)	(0.079)	(0.051)
Approaches to learning (Spring K)	3.097	3.058	3.215
	(0.022)	(0.068)	(0.043)**,†
Self-control (Spring K)	3.165	3.152	3.271
	(0.021)	(0.072)	(0.039)**
Interpersonal (Spring K)	3.108	3.097	3.228
	(0.022)	(0.071)	(0.039)**,†
Externalizing problem behaviors (Spring K)	1.684	1.615	1.535
	(0.022)	(0.079)	(0.039)***
Internalizing problem behaviors (Spring K)	1.534	1.656	1.524
	(0.016)	(0.064)**	(0.032)††

Notes: 931 students in single-grade first; 96 in K-1; 225 in 1-2. Standard errors in parentheses. * denotes that the K-1 or the 1-2 mean is different from the single-grade mean at the 10% level; **, the 5% level; ***, the 1% level. † denotes that the 1-2 mean is different from the K-1 mean at the 10% level; ††, the 5% level, †††, the 1% level.

Table 2.6, Continued: Means of Student Characteristics by Class Type

Student characteristic	Single-grade 1st mean	K-1 mean	1-2 mean
Reading test score (Spring K)	50.862 (0.315)	48.629 (1.251)**	52.694 (0.695)**,†††
Math test score (Spring K)	49.763 (0.308)	48.424 (1.222)	51.520 (0.651)**,††
General Knowledge test score (Spring K)	49.724 (0.335)	48.085 (1.166)	51.322 (0.644)** ,††
Reading test score (Spring 1st)	50.388 (0.283)	48.643 (1.034)*	51.737 (0.669)** ,††
Math test score (Spring 1st)	49.908 (0.297)	48.685 (1.313)	52.052 (0.596)*** ,†††
General Knowledge test score (Spring 1st)	49.588 (0.321)	48.403 (1.130)	51.043 (0.618)** ,††

Notes: 931 students in single-grade first; 96 in K-1; 225 in 1-2. Standard errors in parentheses. * denotes that the K-1 or the 1-2 mean is different from the single-grade mean at the 10% level; **, the 5% level; ***, the 1% level. † denotes that the 1-2 mean is different from the K-1 mean at the 10% level; ††, the 5% level, †††, the 1% level.

Table 2.7: Modeling Student Selection

Student characteristic	Regression 1: K-1 combination dummy as dependent variable	Regression 2: 1-2 combination dummy as dependent variable
Male	-0.005	-0.023
	(0.052)	(0.037)
Age in months (Spring 1st)	0.009	0.004
	(0.007)	(0.005)
Black	-0.243	-0.125
	(0.090)***	(0.069)*
Hispanic	0.080	0.021
	(0.104)	(0.061)
Asian	0.015	-0.025
	(0.111)	(0.082)
Other	0.071	0.005
	(0.115)	(0.089)
Language other than English spoken at home	-0.013	0.023
	(0.107)	(0.071)
SES	-0.008	0.041
	(0.040)	(0.030)
Approaches to learning	0.015	-0.086
	(0.064)	(0.050)*
Self-control	0.017	-0.027
	(0.078)	(0.066)
Interpersonal	-0.030	0.079
	(0.073)	(0.053)
Externalizing problem behaviors	-0.058	-0.058
	(0.055)	(0.044)
Internalizing problem behaviors	0.105	0.007
	(0.054)	(0.044)
Reading test score (Spring K)	-0.0002	0.007
	(0.004)	(0.003)**
Math test score (Spring K)	-0.005	0.002
	(0.005)	(0.003)
General Knowledge test score (Spring K)	0.001	0.0004
	(0.004)	(0.003)
Constant	-0.486	-0.350
	(0.647)	(0.471)
Regression statistics	Regression 1	Regression 2
Number of obs.	277	591
p-value of F statistic	0.205	0.001
Adj. R-squared	0.085	0.103

Notes: Table 2.7 contains the results of two linear regressions of class-type dummies on student characteristics. Both regressions include school fixed effects. In Regression 1, the sample is restricted to the 26 schools offering only single-grade first and K-1 classes. In Regression 2, the sample is restricted to the 55 schools offering only single-grade first and 1-2 classes. Robust standard errors are in parentheses.

Table 2.8: Four Outcome-Regression Models

Independent variables	Model 1	Model 2	Model 3	Model 4
K-1 and 1-2 dummies	X	X	X	X
Student characteristics		X	X	X
Kindergarten test scores			X	X
Kindergarten social rating scores				X
School fixed effects	X	X	X	X

Table 2.9: Coefficients of Interest from Four Outcome Regressions

	K-1 combination	1-2 combination	
Dependent variable: first-grade reading test score			
	Coeff.	Coeff.	Adjusted R-squared
Model 1	-1.270 (1.326)	2.319 (0.866)***	0.177
Model 2	-2.224 (1.294)*	1.331 (0.865)	0.265
Model 3	-0.186 (0.768)	-0.071 (0.466)	0.707
Model 4	0.211 (0.784)	0.186 (0.478)	0.708
Dependent variable: first-grade math test score			
	K-1 combination	1-2 combination	
	Coeff.	Coeff.	Adjusted R-squared
Model 1	-2.642 (1.553)*	3.700 (0.814)***	0.183
Model 2	-3.205 (1.477)**	2.637 (0.779)***	0.275
Model 3	0.442 (1.007)	1.286 (0.525)**	0.642
Model 4	0.640 (1.051)	1.333 (0.551)**	0.646
Dependent variable: first-grade general knowledge test score			
	K-1 combination	1-2 combination	
	Coeff.	Coeff.	Adjusted R-squared
Model 1	-1.008 (1.396)	1.975 (0.813)**	0.208
Model 2	-2.100 (1.302)	0.555 (0.720)	0.393
Model 3	0.232 (0.831)	-0.316 (0.471)	0.712
Model 4	0.253 (0.891)	-0.381 (0.481)	0.714

Note: Robust standard errors are in parentheses. Longitudinal weights are used; results are not sensitive to the inclusion of weights or to clustering at the class level. Please see Appendix 2.1.

Table 2.10: Do 1-2 Students Benefit at the Expense of Other First Graders? Results from a Regression of Test Scores on a School-type Dummy and School Characteristics

	Independent variable: dummy for school offering first and 1-2 classes
Dependent variable	Coefficient
Reading test score	-0.342
(Spring 1st)	(0.414)
Math test score	-0.111
(Spring 1st)	(0.812)
General Knowledge	0.088
test score (Spring 1st)	(0.786)

Note: Table 2.10 contains the results of regressing first grade test scores on a dummy indicating that the school offers single-grade and 1-2 classes, as well as student-level variables from Model 4 and following school-level covariates: indicators for region, community size and year-round school, average class size, standard deviation of enrollment across grades, full-time equivalent teachers per student, percent minority, and percent eligible for free lunch. Robust standard errors are in parentheses.

Appendix 2.1: Full Regression Results

Table 2.9 Full Results:
Model 1

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
	Regression statistics		
Number of obs	1131	1169	1129
F	4.32	11.59	3.32
Prob > F	0.014	0.000	0.037
R-squared	0.243	0.247	0.272
Adj R-squared	0.177	0.183	0.208
Root MSE	7.891	8.380	8.431
	Coefficients		
K-1 dummy	-1.270	-2.642	-1.008
	(1.326)	(1.553)*	(1.396)
1-2 dummy	2.319	3.700	1.975
	(0.866)***	(0.814)***	(0.813)**
Constant	49.890	49.386	49.129
	(0.294)***	(0.303)***	(0.340)***

Note: In Models 1 through 4, robust standard errors are in parentheses. Longitudinal weights are used; results are not sensitive to the inclusion of weights or to clustering at the class level.

Model 2

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
	Regression statistics		
Number of obs	1061	1097	1060
F	14.54	14.06	31.33
Prob > F	0.000	0.000	0.000
R-squared	0.334	0.340	0.449
Adj R-squared	0.265	0.275	0.393
Root MSE	7.382	7.820	7.320
	Coefficients		
K-1 dummy	-2.224 (1.294)*	-3.205 (1.477)**	-2.100 (1.302)
1-2 dummy	1.331 (0.865)	2.637 (0.779)***	0.555 (0.720)
Male	-2.806 (0.525)***	-0.401 (0.527)	0.592 (0.504)
Age in months (Spring 1st)	0.217 (0.066)***	0.342 (0.070)***	0.498 (0.070)***
Black	-2.388 (1.102)**	-4.698 (0.954)***	-6.710 (1.033)***
Hispanic	-0.289 (0.794)	-1.180 (0.875)	-1.109 (0.832)
Asian	3.201 (1.396)**	0.961 (1.343)	-0.559 (1.381)
Other	0.347 (1.410)	-0.495 (1.424)	0.086 (1.295)
Language other than English spoken at home	-2.484 (0.889)***	-2.131 (0.978)**	-4.519 (0.983)***
SES	2.922 (0.428)***	2.742 (0.439)***	3.878 (0.426)***
Constant	34.148 (5.686)***	22.355 (6.045)***	8.993 (6.054)

Note: In Models 1 through 4, robust standard errors are in parentheses. Longitudinal weights are used; results are not sensitive to the inclusion of weights or to clustering at the class level.

Model 3

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
	Regression statistics		
Number of obs	1015	1015	1014
F	110.93	76.17	112.29
Prob > F	0.000	0.000	0.000
R-squared	0.736	0.678	0.741
Adj R-squared	0.707	0.642	0.712
Root MSE	4.604	5.203	4.786
	Coefficients		
K-1 dummy	0.186 (0.768)	0.442 (1.007)	0.232 (0.831)
1-2 dummy	-0.071 (0.466)	1.286 (0.525)**	-0.316 (0.471)
Male	-0.887 (0.336)***	0.507 (0.388)	0.915 (0.357)**
Age in months (Spring 1st)	-0.081 (0.047)*	-0.021 (0.051)	0.043 (0.049)
Black	0.693 (0.593)	-0.880 (0.633)	-2.435 (0.722)***
Hispanic	0.526 (0.541)	-0.411 (0.606)	-0.592 (0.546)
Asian	2.547 (1.002)**	-0.702 (1.086)	0.208 (0.903)
Other	0.977 (0.922)	0.304 (0.952)	0.747 (0.942)
Language other than English spoken at home	-0.093 (0.681)	0.763 (0.728)	0.313 (0.672)
SES	0.096 (0.279)	-0.020 (0.303)	0.714 (0.309)**
Reading test score (Spring K)	0.569 (0.029)***	0.066 (0.033)**	0.041 (0.033)
Math test score (Spring K)	0.195 (0.030)***	0.597 (0.035)***	0.126 (0.034)***

Note: In Models 1 through 4, robust standard errors are in parentheses. Longitudinal weights are used; results are not sensitive to the inclusion of weights or to clustering at the class level.

Model 3, Continued

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
General Knowledge test score (Spring K)	0.034 (0.026)	0.113 (0.031)***	0.617 (0.029)***
Constant	17.419 (3.797)***	13.012 (4.220)***	7.532 (3.952)*

Note: In Models 1 through 4, robust standard errors are in parentheses. Longitudinal weights are used; results are not sensitive to the inclusion of weights or to clustering at the class level.

Model 4

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
	Regression statistics		
Number of obs	968	968	967
F	80.76	56.53	79.38
Prob > F	0.000	0.000	0.000
R-squared	0.740	0.685	0.745
Adj R-squared	0.708	0.646	0.714
Root MSE	4.540	5.179	4.780
	Coefficients		
K-1 dummy	0.211 (0.784)	0.640 (1.051)	0.253 (0.891)
1-2 dummy	0.186 (0.478)	1.333 (0.551)**	-0.381 (0.481)
Male	-0.727 (0.345)**	0.671 (0.429)	1.088 (0.381)***
Age in months (Spring 1st)	-0.078 (0.047)*	-0.042 (0.052)	0.054 (0.050)
Black	0.816 (0.604)	-0.826 (0.638)	-2.527 (0.755)***
Hispanic	0.685 (0.548)	-0.536 (0.621)	-0.788 (0.558)
Asian	2.535 (0.979)**	-0.982 (1.110)	-0.230 (0.952)
Other	0.714 (0.937)	-0.116 (1.006)	0.982 (0.988)
Language other than English spoken at home	-0.677 (0.699)	0.414 (0.787)	0.450 (0.734)
SES	0.229 (0.285)	-0.026 (0.310)	0.578 (0.314)*
Approaches to learning	1.529 (0.404)***	2.140 (0.476)***	-0.107 (0.454)
Self-control	-1.033 (0.489)**	-1.023 (0.580)*	0.402 (0.519)
Interpersonal	-0.047 (0.472)	-0.435 (0.545)	0.635 (0.483)

Note: In Models 1 through 4, robust standard errors are in parentheses. Longitudinal weights are used; results are not sensitive to the inclusion of weights or to clustering at the class level.

Model 4, Continued

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
Externalizing problem behaviors	0.003 (0.399)	-0.157 (0.421)	0.538 (0.370)
Internalizing problem behaviors	0.142 (0.433)	0.006 (0.470)	-0.612 (0.403)
Reading test score (Spring K)	0.535 (0.029)***	0.024 (0.034)	0.030 (0.034)
Math test score (Spring K)	0.182 (0.031)***	0.569 (0.037)***	0.118 (0.037)***
General Knowledge test score (Spring K)	0.026 (0.026)	0.114 (0.031)***	0.616 (0.029)***
Constant	18.245 (4.378)***	16.457 (4.679)***	4.595 (4.697)

Note: In Models 1 through 4, robust standard errors are in parentheses. Longitudinal weights are used; results are not sensitive to the inclusion of weights or to clustering at the class level.

Alternative Model 3 Regression Results: Regressors Include Class-type Dummies, Child Characteristics, and Kindergarten Behavior Measures

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
	Regression statistics		
Number of obs	1010	1042	1009
F	23.32	25.68	28.98
Prob > F	0.000	0.000	0.000
R-squared	0.473	0.470	0.505
Adj R-squared	0.412	0.411	0.448
Root MSE	6.536	7.081	6.975
	Coefficients		
K-1 dummy	-0.665 (1.117)	-1.736 (1.451)	-0.865 (1.271)
1-2 dummy	1.372 (0.755)*	2.552 (0.701)***	0.667 (0.683)
Male	-1.468 (0.468)***	1.082 (0.527)**	1.706 (0.535)***
Age in months (Spring 1st)	0.071 (0.061)	0.176 (0.066)***	0.393 (0.070)***
Black	-1.082 (0.984)	-3.450 (0.884)***	-5.769 (1.059)***
Hispanic	-0.583 (0.720)	-1.494 (0.783)*	-1.633 (0.817)**
Asian	2.070 (1.229)*	-0.300 (1.278)	-2.275 (1.336)*
Other	-0.326 (1.100)	-1.238 (1.211)	0.193 (1.181)
Language other than English spoken at home	-3.434 (0.864)***	-3.227 (0.956)***	-4.531 (0.977)***
SES	2.111 (0.378)***	1.773 (0.400)***	3.075 (0.424)***

Note: In Models 1 through 4, robust standard errors are in parentheses. Longitudinal weights are used; results are not sensitive to the inclusion of weights or to clustering at the class level.

Alternative Model 3 Regression Results, Continued: Regressors Include Class-type Dummies, Child Characteristics, and Kindergarten Behavior Measures

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
Approaches to learning	6.613 (0.572)***	6.777 (0.563)***	3.492 (0.602)***
Self-control	-2.489 (0.765)***	-2.092 (0.814)**	-1.115 (0.811)
Interpersonal	-0.453 (0.729)	-0.713 (0.714)	0.933 (0.718)
Externalizing problem behaviors	0.197 (0.608)	0.116 (0.608)	0.527 (0.595)
Internalizing problem behaviors	-0.700 (0.581)	-0.762 (0.641)	-1.402 (0.612)**
Constant	35.382 (5.871)***	24.634 (6.442)***	8.434 (6.899)

Note: In Models 1 through 4, robust standard errors are in parentheses. Longitudinal weights are used; results are not sensitive to the inclusion of weights or to clustering at the class level.

Robustness Check: Model 4 Regression Including Average Peer Test Scores as Independent Variables

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
	Regression statistics		
Number of obs	968	968	967
F	70.91	48.63	69.45
Prob > F	0.000	0.000	0.000
R-squared	0.742	0.685	0.748
Adj R-squared	0.709	0.645	0.716
Root MSE	4.527	5.187	4.764
	Coefficients		
K-1 dummy	0.865 (0.852)	0.749 (1.066)	0.693 (0.909)
1-2 dummy	0.401 (0.486)	1.390 (0.587)**	-0.384 (0.451)
Male	-0.748 (0.342)**	0.667 (0.430)	1.071 (0.378)***
Age in months (Spring 1st)	-0.081 (0.046)*	-0.044 (0.052)	0.056 (0.049)
Black	0.785 (0.612)	-0.838 (0.641)	-2.513 (0.757)***
Hispanic	0.599 (0.550)	-0.531 (0.623)	-0.942 (0.553)*
Asian	2.515 (0.961)***	-0.928 (1.098)	-0.468 (0.960)
Other	0.711 (0.919)	-0.110 (1.004)	0.972 (0.963)
Language other than English spoken at home	-0.761 (0.719)	0.328 (0.801)	0.650 (0.731)
SES	0.173 (0.287)	-0.041 (0.309)	0.568 (0.313)*
Approaches to learning	1.460 (0.396)***	2.119 (0.471)***	-0.107 (0.458)
Self-control	-0.916 (0.497)*	-0.980 (0.581)*	0.378 (0.518)
Interpersonal	-0.036 (0.468)	-0.430 (0.544)	0.626 (0.477)

Robustness Check, Continued: Model 4 Regression Including Average Peer Test Scores as Independent Variables

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
Externalizing problem behaviors	0.076 (0.395)	-0.133 (0.421)	0.519 (0.375)
Internalizing problem behaviors	0.137 (0.434)	0.013 (0.470)	-0.657 (0.401)
Reading test score (Spring K)	0.542 (0.029)***	0.025 (0.034)	0.039 (0.034)
Math test score (Spring K)	0.181 (0.031)***	0.570 (0.038)***	0.111 (0.037)***
General Knowledge test score (Spring K)	0.022 (0.026)	0.114 (0.031)***	0.611 (0.029)***
Average Spring K Reading test score	0.002 (0.014)	-0.007 (0.016)	0.039 (0.016)**
Average Spring K Math test score	0.015 (0.011)	0.008 (0.013)	-0.012 (0.014)
Average Spring K General Knowledge test score	-0.011 (0.013)	0.000 (0.014)	-0.025 (0.012)**
Constant	17.246 (4.438)***	16.234 (4.680)***	4.318 (4.691)

Robustness Check 2: Propensity Score Matching Estimate of the Effect of 1-2 Class Membership

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
	Regression statistics		
Number of treated observations	178	178	178
Number of control observations	199	225	199
	Coefficients		
1-2 dummy	1.467 (1.112)	1.721 (1.204)	-0.669 (1.248)

Note: The sample is restricted to students in schools offering first and 1-2 classes only.

Table 2.10 Full Results

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
	Regression statistics		
Number of obs	4489	4488	4486
F	123.94	144.15	349.71
Prob > F	0.000	0.000	0.000
R-squared	0.599	0.622	0.706
Root MSE	5.120	5.174	4.793
	Coefficients		
School offers single-grade and 1-2	-0.342 (0.414)	-0.111 (0.465)	0.089 (0.326)
Male	-0.170 (0.192)	0.906 (0.174)***	0.942 (0.161)***
Age in months (Spring 1st)	-0.094 (0.027)***	-0.019 (0.026)	-0.011 (0.021)
Black	0.843 (0.386)**	-1.229 (0.347)***	-1.162 (0.340)***
Hispanic	0.990 (0.379)***	0.214 (0.376)	-0.589 (0.341)*
Asian	1.006 (0.389)**	-0.433 (0.730)	-0.679 (0.465)
Other	0.927 (0.504)*	-0.496 (0.426)	-0.621 (0.407)
Language other than English spoken at home	0.071 (0.457)	0.640 (0.487)	-0.191 (0.442)
SES	0.434 (0.152)***	0.378 (0.161)**	0.522 (0.136)***
Approaches to learning	1.356 (0.252)***	1.522 (0.244)***	-0.169 (0.197)
Self-control	-0.396 (0.293)	-0.510 (0.342)	-0.026 (0.260)
Interpersonal	-0.336 (0.276)	-0.105 (0.277)	0.225 (0.249)

Note: Table contains the results of regressing first grade test scores on a dummy indicating that the school offers single-grade and 1-2 classes, as well as student-level variables from Model 4 and following school-level covariates: indicators for region, community size and year-round school, average class size, standard deviation of enrollment across grades, full-time equivalent teachers per student, percent minority, and percent eligible for free lunch. Robust standard errors are in parentheses.

Table 2.10 Full Results, Continued

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
Externalizing problem behaviors	-0.181 (0.201)	-0.230 (0.244)	0.187 (0.239)
Internalizing problem behaviors	0.276 (0.220)	0.273 (0.216)	-0.112 (0.189)
Reading test score (Spring K)	0.496 (0.016)***	0.064 (0.016)***	0.061 (0.015)***
Math test score (Spring K)	0.149 (0.018)***	0.517 (0.017)***	0.109 (0.014)***
General Knowledge test score (Spring K)	0.038 (0.014)***	0.137 (0.016)***	0.634 (0.014)***
Midwest	-0.303 (0.438)	0.346 (0.351)	0.341 (0.319)
South	-0.075 (0.406)	0.679 (0.382)*	0.416 (0.325)
West	-0.241 (0.448)	0.303 (0.429)	0.669 (0.364)*
Suburb	-0.013 (0.330)	-0.181 (0.322)	-0.306 (0.255)
Town	-0.254 (0.449)	-0.314 (0.418)	-0.008 (0.400)
Rural	-0.078 (0.404)	-0.266 (0.363)	-0.089 (0.293)
Average grade-level enrollment across grades K-2	-0.006 (0.003)*	0.000 (0.003)	-0.004 (0.003)

Note: Table contains the results of regressing first grade test scores on a dummy indicating that the school offers single-grade and 1-2 classes, as well as student-level variables from Model 4 and following school-level covariates: indicators for region, community size and year-round school, average class size, standard deviation of enrollment across grades, full-time equivalent teachers per student, percent minority, and percent eligible for free lunch. Robust standard errors are in parentheses.

Table 2.10 Full Results, Continued

Top row: dependent variable	First grade reading score	First grade math score	First grade general knowledge score
Std. dev. of grade-level enrollment across grades K-2	0.019 (0.019)	0.027 (0.019)	0.038 (0.019)**
Year-round	0.632 (0.510)	0.116 (0.653)	-0.532 (0.417)
FTE teachers per student	7.964 (10.739)	13.641 (9.144)	13.449 (9.462)
Pct. minority	-0.016 (0.005)***	-0.001 (0.005)	-0.010 (0.004)**
Pct. eligible for free lunch	-0.009 (0.006)	-0.005 (0.006)	-0.006 (0.005)
Constant	23.389 (2.380)***	12.037 (2.366)***	10.410 (2.113)***

Note: Table contains the results of regressing first grade test scores on a dummy indicating that the school offers single-grade and 1-2 classes, as well as student-level variables from Model 4 and following school-level covariates: indicators for region, community size and year-round school, average class size, standard deviation of enrollment across grades, full-time equivalent teachers per student, percent minority, and percent eligible for free lunch. Robust standard errors are in parentheses.

References

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113: 151-184.

Bedard, Kelly and Elizabeth Dhuey. 2006. "The Persistence of Early Childhood Maturity: International Evidence of Long-run Age Effects." *Quarterly Journal of Economics*, 121:1437-1472.

Betts, Julian. 1996. "Is There a Link Between School Inputs and Earnings? Fresh Scrutiny of an Old Literature." University of California at San Diego Economics Working Paper Series 96-09.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2008. "Too Young to Leave the Nest? The Effects of School Starting Age." National Bureau of Economic Research Working Paper 13969.

Burns, Robert, DeWayne Mason, and Michael A. Demiranda. 1993. "How Elementary Principals Assign Teachers and Students to Combination Classes," California Educational Research Cooperative, University of California, Riverside.

Statistical Summary of Year-round Programs. 2006. California Department of Education. www.cde.ca.gov/ls/fa/yr/guide.asp (accessed May 1, 2010).

Fingertip Facts: K-3 Class Size Reduction Program History and Summarized Data. 2009. California Department of Education . www.cde.ca.gov/ls/cs/k3/facts.asp (accessed May 1, 2010).

Datar, Ashlesha (2006). "Does Delaying Kindergarten Entrance Give Children a Head Start?" *Economics of Education Review*, 25: 43-62.

Elder, Todd E. and Darren H. Lubotsky. 2006. "Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers." www.rc.rand.org/labor/seminars/adp/pdfs/2007_lubotsky.pdf.

Graves, Jennifer. 2007. "The Effect of Year Round School Calendars on Academic Performance." PhD diss. University of California, Irvine.

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

Hill, Peter W. and Kenneth J. Rowe. 1998. "Modeling Student Progress in Studies of Educational Effectiveness." *Social Effectiveness and School Improvement*, 9: 310-333.

Hoxby, Caroline M. 2000. “The Effects of Class Size on Student Achievement: New Evidence from Population Variation.” *Quarterly Journal of Economics*, 115: 1239-1285.

Mason, DeWayne A. and Robert B. Burns. 1994. “A Review of the Literature on Combination Classes,” California Educational Research Cooperative, University of California, Riverside.

________.1995a. “Organizational Constraints on the Formation of Elementary School Classes.” *American Journal of Education*, 103(2): 185-212.

________. 1995b. “Teachers’ Views of Combination Classes.” *Journal of Educational Research*, 89(1): 36–45.

________. 1996. “‘Simply No Worse and Simply No Better’ May Simply Be Wrong: A Critique of Veenman’s Conclusion about Multigrade Classes.” *Review of Educational Research*, 66(3): 307-322.

Mason, DeWayne A. and Roland Doepner. 1998. “Principals’ Views of Combination Classes” *Journal of Educational Research*, 91(3): 160–172.

Mason, DeWayne A., Robert Burns, and Jorge Armesto. 1993. “Teachers’ Views about Combination Classes.” California Educational Research Cooperative, University of California, Riverside.

Mason, DeWayne A. and Janet Stimson. 1994. “A National Survey of Combination and Nongraded Classes.” California Educational Research Cooperative, University of California, Riverside.

Miller, William. 1995. “Are Multi-Age Grouping Practices a Missing Link in the Educational Reform Debate?” National Association of Secondary School Principals Bulletin 1995.

Mishel, Lawrence and Richard Rothstein. 2002. *The Class Size Debate*, Washington, D.C.: Economic Policy Institute.

Statistical Summaries of Year-round Education Programs. 2007. National Association for Year-round Education. www.nayre.org/STATISTICAL%20SUMMARIES%20OF%20YRE%202007.pdf (accessed May 1, 2010).

Veenman, Simon. 1995. “Cognitive and Noncognitive Effects of Multigrade and Multi-Age Classes: A Best-Evidence Synthesis.” *Review of Educational Research*, 65(4) 319-381.

Veenman, Simon. 1996. "Effects of Multigrade and Multi-Age Classes Reconsidered." *Review of Educational Research* 66(3): 323-340.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press.

Stipek, Deborah. 2002. "At What Age Should Children Enter Kindergarten? A Question for Policy Makers and Parents." *Social Policy Report: a Publication of the Society for Research in Child Development*, 16: 3-16.

Sims, David. 2008. "A Strategic Response to Class Size Reduction: Combination Classes and Student Achievement in California." *Journal of Policy Analysis and Management*, 27: 457-478.

CHAPTER 3

NEIGHBORHOOD DEMOGRAPHICS, SCHOOL EFFECTIVENESS, AND RESIDENTIAL LOCATION CHOICE

Abstract: In this paper, I investigate how neighborhood demographics and school effectiveness influence the residential location decisions of parents of different income levels. I find that low-income parents in the San Francisco Bay Area respond more strongly to school effectiveness than to neighborhood demographics, but that the reverse is true for high-income parents.

3.1 Introduction

School choice has become an important and controversial issue in recent years, with popular discussion of the issue focusing on systems such as vouchers and open enrollment plans that allow students to attend public schools other than their neighborhood schools (Barrow, 2002; Hoxby, 2003). Parents, however, have long exercised school choice through residential location decisions in what is sometimes called Tiebout sorting, in reference to Tiebout's (1956) paper arguing that individuals will reveal their preferences for local public goods by voting with their feet.

This paper seeks to answer two questions. First, what do parents care about when choosing a place to live—neighborhood demographics or school effectiveness? Second, how does the influence of neighborhood demographics and school effectiveness on residential location choice vary by household income? In order to capture neighborhood demographics, I use test scores predicted on the basis of neighborhood and student attributes. I capture school effectiveness by using a type of value-added measure.

I find a monotonic relationship with predicted test scores and income: predicted scores have a negative but insignificant association with location choice for the poorest parents but are increasingly positive and significant as income increases. In general, lower- and middle-income parents respond more strongly to school effectiveness than higher-income parents, but its relationship to income is not monotonic: school effectiveness grows in importance as income increases from the first to the third quintile but decreases in importance to parents in the fourth and fifth quintiles. Parents in the first quintile of the income distribution respond significantly more strongly to school effectiveness than to predicted test scores. The reverse is true for parents in the fourth and fifth quintiles of the income distribution. This evidence suggests that lower-income parents place a high value on school effectiveness and that higher-income parents place more weight on neighborhood demographics when choosing a place to live.

This paper proceeds as follows. Section 3.2 contains a brief literature review. Section 3.3 describes the data sets used in my analysis, the sample of households I consider, and the construction of my school-quality variables. Section 3.4 describes the estimation method and results and Section 3.5 concludes.

3.2 Literature Review

Whether parents care about school effectiveness or only about neighborhood demographics has important implications for the school-choice debate. Rothstein (2006) hypothesizes that schools in markets with more choice should be more effective than schools in markets with less choice if parents value school effectiveness, but that this need not be the case if parents instead value other attributes such as peer and neighborhood quality. He finds that choice does not have strong effects on school effectiveness and presents several plausible explanations: parents may place a low weight on effectiveness, they may value effectiveness but lack the information necessary to identify effective schools, or variation in effectiveness is responsible for only a small share of cross-school variation in student outcomes.

Hastings and Weinstein (2008) address the important finding that low-income parents place less weight on academics than other groups. They ask if this is because they expect lower returns to education for their children, or because these parents find gathering information more costly. When they provide lower-income parents with improved information on test scores, they find that parents choose higher-scoring schools for their children. Their results imply that information is important—better information leads parents to exercise school choice.

If information is important to parents, what type of information do they care about? One way to assess the importance of different school-quality measures is to measure their effects on housing prices in what are known as hedonic house price models. Authors have attempted to measure the effect of test score levels, test score gains, school value-added, school letter grades, and other school characteristics on housing prices. By looking within school districts at houses located on attendance district boundaries, Black (1999) attempts to remove the variation in nonschool neighborhood characteristics and finds that parents are willing to pay 2.5 percent more for a five-percent increase in test scores, a smaller effect than most estimates. Clapp, Nanda, and Ross (2008) also find a small but positive effect of test scores on property values. Figlio and Lucas (2004) find that Florida housing markets respond strongly to the initial assignment of school letter grades, but that as more information is provided, grades are capitalized less and less.

Some authors consider test score levels a naïve measure of school quality. Dills (2004) considers instead test score gains and finds little to no relation between changes in test scores and changes in total housing value in a district.[34] Brasington and Haurin (2006) use a value-added measure like that of Hayes and Taylor (1996) and similar to that used in this paper. Value-added indicators measure school performance by isolating the contribution of schools from all of the nonschool factors that also contribute to

[34] She does, however, find a positive effect of college entrance exam performance on total housing value.

student achievement. While Hayes and Taylor (1996) find a positive and significant effect of school value-added on home prices, Brasington and Haurin (2006) find no relationship. Summarizing the literature, Dills (2004) finds that most commonly, house prices capitalize average proficiency test scores and claims that “researchers typically attribute the lack of house price response to value-added measures as a lack of sophistication” on the part of home buyers (p. 2).

One disadvantage of hedonic house price models is that, though individual household characteristics such as number of children influence households’ willingness to pay for school quality, such characteristics are not included in reduced-form hedonic regressions of log price on house and neighborhood characteristics. Hedonic models do not consider the location decisions of individual households; as such, they cannot include individual household characteristics in the analysis.

Discrete choice models of household location allow the researcher to examine how the effect of school quality on location choice differs with household characteristics. Bayer, Ferreira, and McMillan (2007) estimate household preferences over a broad range of housing and neighborhood characteristics. They employ a two-part model consisting of the household residential location decision problem and a market-clearing condition. They find that households in the San Francisco Bay area are willing to pay an additional one percent in house prices when the average performance of the local school is increased by five percent. Hastings, Kane, and Staiger (2005) find that parents value proximity highly, and that the value attached to a school’s mean test score increases with a student’s income and own academic ability.

Barrow (2002) uses SAT scores to estimate the effect of school quality on households' residential location choices. Her identification strategy is to compare the location choices of households with children to households without children, reasoning that nonschool neighborhood characteristics will affect both types similarly, while households with children will care more about school quality. In other words, she assumes that unobservable neighborhood attributes are unlikely to be correlated with school quality interacted with household child status. She finds that White households with children appear to exercise school choice through residential location decisions. Willingness to pay varies positively with wealth, education, and age of household head. Among African-American households, however, she does not find evidence that households with children locate in areas with higher school quality than childless households, perhaps because some African-American households encounter restricted neighborhood choice sets.

The contributions of this paper are to decompose test scores into a neighborhood-demographics component and a school-effectiveness component, and to examine how the importance of these components varies by household income, using a unique subsample of households that have undergone a move-inducing shock. In the next section, I describe the data sets used in my analysis, the sample of households I consider, and the construction of my school-quality variables.

3.3 Data, Household Sample, and School-Quality Variables

The data on household characteristics are obtained from the 2000 Census via the University of Minnesota's Integrated Public Use Microdata Series, Version 5.0. I restrict my sample to households with heads who are employed and who had moved from out of state to the San Francisco Bay Area within the past five years. My sample consists of 8,702 households in six counties: Alameda, Contra Costa, Marin, San Francisco, San Mateo, and Santa Clara.

I restrict my sample to an area within California because school spending is not directly related to local property tax rates as it is in many other states. The State Supreme Court's decision in the case *Serrano v. Priest (1971)* mitigates the problem of controlling adequately for effective tax rates as they relate to school spending across regions. I further restrict my sample to the San Francisco Bay Area because this area is densely populated and contains many different local jurisdictions. In addition, very few commutes originating in this area end up outside the area, and few commutes ending up in this area originate from outside the area (Bayer, Ferreira, and McMillan, 2007).

I focus on out-of-state movers for two reasons. One, they are less subject to Proposition 13 lock-in. California's Proposition 13, passed in 1978, capped property tax rates at one percent of a home's assessed value. It also limited the rate of growth of property taxes. Since then, housing values have increased considerably in California, so homeowners who have owned a house for many years in California have a disincentive to move as they would have to pay property taxes on the new home's higher assessed value.

The second reason I focus on out-of-state movers is that I assume they have undergone a move-inducing shock. It is important to address the following question: if communities are in Tiebout equilibrium to begin with, why do people move? Kane, Staiger, and Riegg (2006) and Figlio and Lucas (2004) analyze different regime changes (the redrawing of attendance district boundaries and the introduction of state-assigned school grades, respectively) that can be considered as shocks that might induce families to re-sort. I look instead at households choosing to move to California after experiencing a shock in their home state; for example, job relocation. According to 2000 CPS data (obtained from the University of Minnesota's Integrated Public Use Microdata Series, Current Population Survey, Version 2.0), 45 percent of individuals moving from outside of California to California within the previous year moved for job-related reasons, while only 11 percent of within-California movers cited job-related reasons for moving.

One weakness of the public use Census data is that household location is identified at the Public Use Microdata Area (PUMA) level. PUMAs are areas of approximately 100,000 people. The six counties I consider contain 47 PUMAs, none of which crosses a county line. In my analysis, neighborhood is synonymous with PUMAs, though PUMAs do not tend to line up with school district or attendance zone boundaries, and several districts may be contained within one PUMA. I obtain neighborhood (PUMA) characteristics from the 2000 Census, extracted via the Missouri Census Data Center's Dexter Data Extractor.

The data for the school-quality measures are obtained from the California Department of Education. I use fifth-grade STAR test scores from the 1997-1998 and the 1998-1999 school year. I compute a reading and math composite score for each school year, then average these composites across the two school years in order to reduce year-to-year noise. I obtain the 97/98-98/99 average of school demographic characteristics from the National Center for Education Statistics' Common Core of Data.

In order to compute predicted scores and the school effectiveness measure, I regress the average composite score for each school on the following school-level variables, using Ordinary Least Squares: fifth grade enrollment, percent Asian/Pacific Islander, percent Hispanic, percent Other, and percent eligible for free lunch (percent White is omitted). I also use the following PUMA-level covariates: percent of adults 25 or over with less than a high school degree and with a college degree (percent with a high school degree is omitted), percent unemployed, and percent urban (percent rural is omitted). Table 3.1 contains the results of this regression. Enrollment is negatively associated with test scores, as are percent Black, percent Hispanic, and percent Other (relative to percent White). Relative to percent with a high school degree, percent less than high school and percent college graduates are positively associated with test scores.

Predicted test scores are the fitted values from this regression and school effectiveness is the regression residual. The fitted values capture the contribution of student and neighborhood attributes to test scores. The regression residual captures the contribution of schools to test scores given these attributes. Though in spirit this is a value-added measure, it differs from measures such as those discussed in Hayes and Taylor (1996) and Brasington and Haurin (2006) in that the dependent variable is a two-year average of test scores from the same grade, whereas the dependent variable in their models is test score gains. After obtaining these measures in a school-level regression, I aggregate to the PUMA level.

3.4 Estimation Method and Results

Barrow (2002) estimates a multinomial logit model of household location choice. Other authors refer to this type of model as a random utility model or a conditional logit model. In addition, since the model was introduced by McFadden (1974), some refer to it as McFadden's choice model. I will follow the convention of Wooldridge (2002) and use the term conditional logit. According to Wooldridge (2002), the conditional logit model is intended for problems in which consumer choices are at least partly based on observable attributes of the alternatives under consideration, as they are in my analysis, while the multinomial logit model is appropriate for problems where characteristics of the alternatives are not important, or if data on these attributes are not available (p. 501).

According to the conditional logit model, households maximize indirect utility:

$$U_{hj} = V_{hj} + \varepsilon_{hj}, \tag{1}$$

where *h* indexes households, *j* indexes neighborhoods, and

$$V_{hj} = X'_{hj}\beta . \tag{2}$$

ε_{hj} is independently and identically distributed as type 1 extreme value. X'_{hj} contains neighborhood characteristics, including school quality, and their interactions with household characteristics.

The choice probability, or the probability that household *h* chooses community *j*, is given by

$$P_{hj} = \frac{\exp(V_{hj})}{\sum_{k=1}^{K}\exp(V_{hk})}, \tag{3}$$

where *K* denotes the total number of neighborhoods under consideration. The parameters of this model are estimated using the method of maximum likelihood.

A problematic restriction of this model is that assuming that the error term is i.i.d., type 1 extreme value gives rise to the independence from irrelevant alternatives (IIA) assumption. According to the IIA assumption, the odds of choosing *j* over *j'* are independent of the presence or characteristics of a third alternative. This assumption is questionable in the context of household location choice, but it allows me to consider a subset of the full set of neighborhood choices and focus on the San Francisco Bay Area. In addition, McFadden (1984) shows that even in cases where the IIA assumption is implausible, the conditional logit model is robust as measured by goodness-of-fit or prediction accuracy.

Bayer, Ferreira, and McMillan (2007) model residential location choice using a conditional logit model. In addition, they include an equilibrium condition which is a set of residential location choices and prices such that the housing market clears and each household makes its optimal choice given the decisions of all other households. They also utilize the school attendance zone boundary fixed effect method pioneered by Black (1999) in order to address the correlation between school quality and unobserved neighborhood characteristics. Instead of positing that the San Francisco Bay Area housing market is in equilibrium, I restrict my sample to households that have undergone a move-inducing shock. I am unable to employ the boundary fixed effect method of Black (1999) because I use public-use Census data. Like Barrow (2002), therefore, I compare households with children to households without children in order to address correlation between school quality and unobserved neighborhood characteristics, reasoning that unobserved neighborhood characteristics affect both types similarly, while school quality has a greater impact on households with children.

Tables 3.2 and 3.3 contain the results of two different specifications of conditional logit model of residential location choice. In both specifications, the dependent variable is the choice of PUMA. As additional controls for unobservable PUMA characteristics that affect parents and nonparents similarly, I include PUMA fixed effects in both specifications. The other covariates in the first specification are interactions between predicted test scores and a parent dummy, and between the school efficiency measure and the parent dummy. The parent dummy indicates that the household head has at least one of his or her own children under 18 living in the household. In the second specification, I include three additional interactions. I include an interaction between the parent dummy and percent of PUMA residents who live in urban areas because parents may have different preferences than other employed movers on this dimension. The other interactions are between a dummy indicating that the household head is White and percent of PUMA residents who are White, and between a dummy indicating that the household head has a college degree and the percent of PUMA residents with a college degree. I include these to capture any preferences on the part of individuals for being with others like them. I split the sample into five household income quintile groups and run separate regressions for each group, for a total of ten regressions.

Overall, the patterns are the same. In both specifications, the parent-predicted score interaction exhibits a monotonic relationship with income. The parent-effectiveness interaction exhibits a concave relationship, increasing in importance as income increases from the first to the third quintile, and decreasing in importance for the fourth and fifth quintiles.

Though the patterns are similar, the second specification is preferred. Akaike's and Schwarz's information criteria are lower in value in the second specification at every income level. In the second specification, the interaction between parent and predicted score has a negative but statistically insignificant coefficient for parents in the first quintile of the income distribution. The point estimate is positive and significant for parents in the second quintile and continues to increase for parents in the third, fourth, and fifth quintiles. The interaction between parent and school effectiveness is positive and statistically significant for parents in the first quintile of the income distribution. The point estimate increases for the second and third quintiles, then decreases for the fourth and fifth quintiles.

According to F-tests of the equality of the coefficients, parents in the first quintile of the income distribution respond significantly more strongly to school effectiveness than to predicted test scores. The reverse is true for parents in the fourth and fifth quintiles of the income distribution. These patterns suggest that lower-income parents place a high value on school effectiveness and that higher-income parents place more weight on neighborhood demographics when choosing a place to live.

3.5 Conclusion

Parents exercise school choice through residential location decisions. Lower-income parents appear to place a high value on school effectiveness, and higher-income parents place more weight on neighborhood demographics when choosing a place to live. Authors such as Dills (2004) consider school effectiveness to be a more sophisticated measure of school quality than average test scores. This paper suggests that lower-income parents who move to the San Francisco Bay Area from out of state exercise school choice in a sophisticated way.

An important area for further research is to determine how these parents learn about school effectiveness, since it is a more difficult statistic to obtain than average test scores, and also more difficult to observe than the demographic characteristics of a particular neighborhood.

Table 3.1: Ordinary Least Squares Regression Results

Dependent Variable: 1998-1999 Average Reading and Math Composite Score	
	Coefficient (Standard Error)
Enrollment	-0.008*** (0.002)
Percent Asian/Pacific Islander	0.019 (0.026)
Percent Black	-0.374*** (0.036)
Percent Hispanic	-0.382*** (0.035)
Percent Other	-0.391* (0.210)
Percent Free Lunch Eligible	-0.384*** (0.029)
Percent Less than High School	0.532*** (0.110)
Percent College Graduate	0.578*** (0.044)
Percent Unemployed	0.622 (0.457)
Percent Urban	-0.112 (0.110)
Number of Observations	793
Adjusted R-squared	0.832

Notes: * indicates significance at the ten percent level, ** at the five percent level, and *** at the one percent level. Percent White and Percent High School Graduate are omitted categories.

Table 3.2: Conditional Logit Regression Results, Specification 1

Dependent Variable: Choice of PUMA					
		Coefficient (Standard Error)	p-value that coefficients are equal	AIC	BIC
Income Quintile 1	Parent*Predicted Score	-0.085 (0.054)	0.001	12646.2	13093.2
	Parent*Effective-ness	0.175*** (0.049)			
Income Quintile 2	Parent*Predicted Score	0.044 (0.050)	0.049	12966.5	13414.1
	Parent*Effective-ness	0.203*** (0.053)			
Income Quintile 3	Parent*Predicted Score	0.089* (0.052)	0.065	12395.3	12841.6
	Parent*Effective-ness	0.248*** (0.057)			
Income Quintile 4	Parent*Predicted Score	0.247*** (0.052)	0.120	12531.9	12978.9
	Parent*Effective-ness	0.120** (0.052)			
Income Quintile 5	Parent*Predicted Score	0.338*** (0.049)	0.001	12042.4	12489.4
	Parent*Effective-ness	0.085* (0.049)			

Notes: * indicates significance at the ten percent level, ** at the five percent level, and *** at the one percent level. PUMA fixed effects are included in the regression but their coefficients are not reported.

Table 3.3: Conditional Logit Regression Results, Specification 2

Dependent Variable: Choice of PUMA					
		Coefficient (Standard Error)	p-value that coefficients are equal	AIC	BIC
Income Quintile 1	Parent*Predicted Score	-0.013 (0.056)	0.042	12345.9	12820.9
	Parent*Effective-ness	0.157*** (0.050)			
	Parent*Percent Urban	-0.109*** (0.024)			
	White*Percent White	0.028*** (0.003)			
	College Graduate*Pct. College Graduate	0.045*** (0.004)			
Income Quintile 2	Parent*Predicted Score	0.127** (0.052)	0.529	12739.1	13214.6
	Parent*Effective-ness	0.179*** (0.054)			
	Parent*Percent Urban	-0.091*** (0.022)			
	White*Percent White	0.023*** (0.003)			
	College Graduate*Pct. College Graduate	0.045*** (0.004)			
Income Quintile 3	Parent*Predicted Score	0.128** (0.054)	0.277	12158.6	12632.8
	Parent*Effective-ness	0.223*** (0.057)			
	Parent*Percent Urban	-0.102*** (0.023)			
	White*Percent White	0.035*** (0.003)			
	College Graduate*Pct. College Graduate	0.047*** (0.005)			

Notes: * indicates significance at the ten percent level, ** at the five percent level, and *** at the one percent level. PUMA fixed effects are included in the regression but their coefficients are not reported.

Table 3.3, Continued: Conditional Logit Regression Results, Specification 2

Dependent Variable: Choice of PUMA					
		Coefficient (Standard Error)	p-value that coefficients are equal	AIC	BIC
Income Quintile 4	Parent*Predicted Score	0.280*** (0.054)	0.027	12326.7	12801.6
	Parent*Effective-ness	0.096* (0.053)			
	Parent*Percent Urban	-0.129*** (0.054)			
	White*Percent White	0.032*** (0.003)			
	College Graduate*Pct. College Graduate	0.049*** (0.005)			
Income Quintile 5	Parent*Predicted Score	0.349*** (0.051)	0.001	11850.2	12325.1
	Parent*Effective-ness	0.095* (0.050)			
	Parent*Percent Urban	-0.154*** (0.022)			
	White*Percent White	0.041*** (0.004)			
	College Graduate*Pct. College Graduate	0.036*** (0.007)			

Notes: * indicates significance at the ten percent level, ** at the five percent level, and *** at the one percent level. PUMA fixed effects are included in the regression but their coefficients are not reported.

References

Barrow, Lisa. 2002. "School Choice through Relocation: Evidence from the Washington, D.C. Area." *Journal of Public Economics*, 86: 155-189.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences across Schools and Neighborhoods." *Journal of Political Economy*, 115(4): 588-638.

Black, Sandra. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *The Quarterly Journal of Economics*, 114(2): 577-599.

Borjas, Geroge J., Stephen G. Bronars, and Stephen J. Trejo. 1992. "Self-Selection and Internal Migration in the United States." National Bureau of Economic Research Working Paper 4002.

Brasington, David and Donald R. Haurin. 2006. "Educational Outcomes and House Values: A Test of the Value Added Approach." *Journal of Regional Science*, 46(2): 245-268.

Brunner, Eric J. and Jon Sonstelie. 2006. "California's School Finance Reform: An Experiment in Fiscal Federalism." University of Connecticut Department of Economics Working Papers 06-09.

Coleman, James S. 1966. "Equality of Educational Opportunity Study." U.S. Department of Education.

Clapp, John M., Anupam Nanda, and Stephen L. Ross. 2008. "Which School Attributes Matter? The Influence of School District Performance and Demographic Composition on Property Values." *Journal of Urban Economics*, 63: 451-466.

Dills, Angela K. 2004. "Do Parents Value Changes in Test Scores? High Stakes Testing in Texas." *Contributions to Economic Analysis and Policy*, 3(1): 1-32.

Figlio, David N. and Maurice E. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *American Economic Review*, 94(3): 591-604.

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

Hastings, Justine S. and Jeffrey M. Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *The Quarterly Journal of Economics*, 123(4): 1373-1414.

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

________. 2003. "School Choice and School Productivity: Could School Choice Be a Tide that Lifts All Boats?" In *The Economics of School Choice*, ed. Caroline Hoxby, 287-342. Chicago: University of Chicago Press.

Hayes, Kathy J. and Lori L. Taylor. 1996. "Neighborhood School Characteristics: What Signals Quality to Homebuyers?" *Federal Reserve Bank of Dallas Economic Review*, 3:2-9.

Kane, Thomas J. and Douglas O. Staiger. 2002. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *The Journal of Economic Perspectives* 16(4): 91-114.

Kane, Thomas J., Douglas O. Staiger, and Stephanie K. Riegg. 2006. "School Quality, Neighborhoods, and Housing Prices." *American Law and Economics Review*, 8(2): 183-212.

McFadden, Donald (1978). "Modeling the Choice of Residential Location." In *Spatial Interaction Theory and Planning Models*, ed. Anders Karlgvist, Lars Lundquist, Folke Snickers, and Jorgen Weibull, 75-96. Amsterdam: North Holland Publishing Company.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *The Journal of Political Economy,* 82(1): 34-35.

Rothstein, Jesse. 2006. "Good Principals or Good Peers: Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions." *The American Economic Review*, 96(4): 1333-1350.

Taylor, Lori L. 2005. "Revealed-Preference Measures of School Quality." In *Measuring School Performance and Efficiency: Implications for Practice and Research*, ed. Leanna Steifel, Amy Ellen Schwartz, Ross Robenstein, and Jeffrey Zabel, 163-184. Larchmont, NY: Eye on Education.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *The Journal of Political Economy*, 64(5): 416-424.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press.