Three Essays in the Economics of Education and Labor Economics

by

F. G. Elias Walsh

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Public Policy and Economics)
in The University of Michigan
2011

Doctoral Committee:

Professor Brian A. Jacob, Co-Chair
Professor John Bound, Co-Chair
Professor Charles C. Brown
Professor John E. DiNardo

ACKNOWLEDGEMENTS

I am deeply grateful to the members of my dissertation committee – Brian Jacob, John Bound, John DiNardo and Charlie Brown – for their teaching and mentorship throughout my time in graduate school. Also, I fear to imagine what would have come of this effort without the support and friendship of Jess Goldberg, Erik Johnson and Josh Hyman.

Additionally, Justin McCrary, Sue Dynarski, Jeff Smith, Zoë McLaren, Rob Garlick, Emily Beam, Nate Schwartz and seminar participants at the University of Michigan have provided valuable suggestions about my research. A special thank you to Mary Corcoran for all of her support and encouragement. My parents, Rick and Jeanne Walsh, have been tremendously supportive, even if they had no real idea of what graduate school is really about.

Thanks also to the IPUMS folks at the Minnesota Population Center, Nancy Slavin, Raquel Saucedo, Art Kim, Joshua Garcia, Lauren McClellan, Amy Nowell, Dan Bugler, Ascencion Juarez and Angela Alonzo at the Chicago Public Schools, and to Tim Daly and Andy Sokatch at The New Teacher Project.

My research has been supported with grants from the Rackham Graduate School, and resources from the Center for Local, State and Urban Policy.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF APPENDICES

ABSTRACT

Three Essays in the Economics of Education and Labor Economics

by

F. G. Elias Walsh

Co-Chairs: Brian A. Jacob and John Bound

In the first chapter of this dissertation, I examine the role of persistent effects of past labor market conditions in explaining trends in the college-high school wage gap in the US. I document increases in the wage gap for workers since the late 1990s, which are larger than those predicted by standard explanations and are consistent with an important role for persistence in the wage gap. Using a semi-parametric estimation procedure, I show that the increases are caused by changes in age profiles in the wage gap across birth cohorts, rejecting the assumption of constant age profiles in prior work and providing evidence of persistence in the wage gap. I find that higher unemployment at the age of high school graduation leads to higher college-high school wage gaps through age 30 in the birth cohort. I identify the persistent effects of initial unemployment rates controlling flexibly for unobserved transient effects of contemporaneous conditions. The fade out of persistent effects of initial unemployment rates with age can account for over a third of the unexplained increase in the wage gap for older workers, and nearly all of the increase for younger workers. The results imply that an important component of wage inequality is driven by the luck of birth cohorts to enter the labor market when conditions are favorable. To alleviate the effects of

persistent wage inequality, policy makers should consider targeted cross-generational transfers over transfers designed to alleviate only the effects of transitory labor market conditions.

The second chapter, written jointly with Brian Jacob, examines the relationship between the formal ratings that principals give teachers and a variety of observable teacher characteristics, including proxies for productivity. Prior work has shown that principals can differentiate between more and less effective teachers, especially at the tails of the quality distribution, and that subjective evaluations of teachers are strongly correlated with subsequent student achievement. However, whereas prior work has relied on survey data, we consider formal ratings from a setting in which the stakes are reasonably high. We find that the ratings are correlated with an array of teacher qualities including experience for young teachers, education credentials, and teacher absenteeism. Our finding that principals reward qualities of teachers known to be related to student productivity provides reason to be optimistic about policies that would assign more weight to principal evaluations of teachers in career decisions and compensation.

In the third chapter, I ask whether the benefits of pre-school participation can account for the magnitude of the effects of school entry policies on student outcomes usually attributed to the student's age of entry into school. I develop a dynamic model of human capital accumulation, which accounts for differential pre-school experiences of children. Using the model, I show that differences in the rate of participation in pre-kindergarten activities between students assigned to enter school at different ages could lead to upward bias in prior estimates of effects of entry age on student achievement, and downward bias in estimates of the effects of time in school. I find that students assigned more time out of school are more likely to attend pre-kindergarten. Accounting for pre-kindergarten participation, I find larger effects of time in school on math scores than in prior work. To account for all of the effects of entry policies usually attributed to entry age, the benefits of pre-kindergarten would need to be implausibly large. However, using variation in the response to and compliance with school entry policies across states and time, I find evidence of large effects of pre-kindegarten attendance on math scores.

CHAPTER I

The Role of Wage Persistence in the Evolution of the College-High School Wage Gap

1.1 Introduction

By nearly any measure, cross-sectional wage inequality in the U.S. increased throughout the 1970s, 1980s, and 1990s (Card and DiNardo, 2002; Autor et al., 2008). The large increase in wage inequality is the motivation for a broad literature seeking to explain changes in the U.S. wage structure over time. The college-high school wage gap measures the component of wage inequality related to educational attainment.[1] In 1970 the average college worker earned about 30 percent more than a high school worker. By 2009 the gap had increased to 50 percent, with steep increases throughout the late 1980s and 1990s. To develop policies to address increasing wage inequality, policy makers must understand the source of changes in the wage gap.

[1] Alternative measures of wage inequality produce often subtly different pictures of trends in the U.S. wage structure. See, for example, Card and DiNardo (2002); Autor et al. (2008).

Even as the wage gap has increased, the skill-composition of the workforce has shifted toward more highly educated workers. Large increases in the relative demand for skilled (college educated) labor can account for much of the cross-sectional increase in the wage gap (see, for example, Bound and Johnson, 1992; Katz and Murphy, 1992). Other explanations for trends in the wage gap include the falling real minimum wage (Blackburn et al., 1991; Lee, 1999; Card and DiNardo, 2002) and the decline of unionization (Freeman, 1991; Card and DiNardo, 2002).

Secular changes, however, are not sufficient to explain different trends in the wage gap for young and old workers. Card and Lemieux (2001) show that increases in the wage gap during the 1980s were only present for younger workers. The presence of different trends in the wage gap by age suggests that an important component of the increase in wage inequality is due to characteristics of birth cohorts of workers. An early example, Welch (1979), finds evidence that large cohorts have smaller college-high school wage gaps. Card and Lemieux (2001) show that cohorts with a smaller proportion of college educated workers have larger college-high school wage gaps.

The effect of cohort-specific relative supply – the ratio of college to high school educated labor in a cohort – can explain most of the differential trends in the wage gap by age in the 1980s. However, as I will demonstrate in the next section, a cohort's relative supply cannot account for all of the increase in the wage gap for workers during the late 1990s and 2000s, after the period of analysis in Card and Lemieux (2001). Because the fit is worse for older workers than for younger workers, the failure of standard explanations to predict recent trends in the wage gap suggests a further role for birth cohort-specific (rather than secular) explanations.

Prior work on the role of birth cohorts in the evolution of the wage gap has considered only the effects of permanent characteristics of birth cohorts, such as cohort size or attainment, but the labor market experiences of cohorts can also generate cohort effects in the wage gap. While the effect of permanent characteristics of cohorts on the wage gap is unlikely to change with age,[2] early labor market experiences of birth cohorts can have a persistent effect on the wage gap in a cohort, which eventually fades with age. There is growing evidence that the history of labor market conditions faced by a worker is important for understanding contemporaneous wage levels (for evidence of persistence in wage levels see Beaudry and DiNardo, 1991; Baker et al., 1994; Kahn, 2010; Schmieder and von Wachter, 2010; Oyer, 2006, 2008). Differences between college and high school workers in the magnitude and duration of persistent effects of initial labor market conditions on wage levels would generate changes in the

[2]Welch (1979) considers models predicting changes across birth cohorts in age profiles in the wage gap in response to changes in cohort size. In Appendix C, I show that accounting for a related set of concerns does not affect the main results in this paper.

shape of age profiles in the wage gap across cohorts.

Some recent work has considered the interaction between skill and wage persistence. Oreopoulos et al. (2008) show that the persistent effects of graduating college when unemployment is high is larger for lower skill graduates in Canada. Genda et al. (2010) show larger effects of unemployment at the age of graduation for high school workers than for college workers in the US, but the effects for high school workers are less persistent. The evidence of differences in the persistent effects of past labor market conditions faced by college and high school workers suggests a role for wage persistence in generating changes in within-cohort age profiles in the wage gap not explained by standard models.

In this paper, I ask whether wage persistence – effects of the history of labor marker conditions faced by a worker on contemporaneous wages – can explain patterns in the college-high school wage gap unexplained by prior models. Understanding the sources of wage inequality is important in considering policies aimed at addressing wage inequality. This work has policy implications: to the extent that earnings differentials are persistent within birth cohorts, policy makers should consider targeted, cross-generational transfers over transfers designed to alleviate the effects of transitory labor market shocks.[3]

[3]Though I focus exclusively on the college-high school wage gap, the results in this paper are

I use the Current Population March Supplements (March CPS) from 1964-2009 to estimate the persistent effects of labor market conditions at the approximate age of market entry for college and high school workers on the wage gap. I demonstrate that the component of the wage gap unexplained by prior models is characterized by changes in within-cohort age profiles, consistent with an important role for wage persistence in the evolution of the wage gap. Accounting for changes in age profiles across cohorts, I find a smaller role for contemporaneous labor market conditions in the evolution of the wage gap than in prior work. Wage persistence from initial unemployment rates faced by birth cohorts accounts for over a third of the recent unexplained increase in the wage gap for older workers, and nearly all of the unexplained increase for younger workers. The effects of initial unemployment rates on the wage gap are large and persist through age 30.

This work contributes to a broad literature on the causes of trends in wage inequality. First, I consider a more flexible model of the evolution of the wage gap than in prior work by accounting for changes in age profiles across cohorts. Second, I implement a semi-parametric estimation procedure to plausibly identify within-cohort age profiles separately from contemporaneous determinants of the wage gap. Finally, this is the first paper to directly estimate the role of wage persistence in the evolution

suggestive of the role of persistence in wage inequality more generally.

of the college-high school wage gap, and extends a limited literature on the interaction between skills of workers (educational attainment) and persistence.

The remainder of the paper proceeds as follows: Section 1.2 discusses the evolution of the wage gap and documents recent unexplained changes; Section 1.3 reviews the literature on the mechanisms behind wage persistence and its effects; Section 1.4 outlines the theoretical concerns relating wage persistence and the wage gap; Section 1.5 presents the data and empirical strategy I employ to estimate the effects on the wage gap of persistence from observable labor market conditions; Section 1.6 presents the results; and Section 1.7 concludes.

1.2 The Evolution of the College-High School Wage Gap

In this section, I describe the evolution of the college-high school wage gap over time and across cohorts in detail, and evaluate the success of standard explanations in predicting recent trends in the wage gap.

1.2.1 Cross-sectional trends in the college-high school wage gap

The college-high school wage gap and the aggregate relative supply of college labor – the log ratio of the quantity of college versus high school educated workers in a given year – both increase dramatically between 1964 and 2009.[4] Figure 1.1 shows increases in the wage gap for men[5] beginning in the 1980s, with the steepest increases occurring in the late 1990s. Figure 1.2 shows a steep increase in the aggregate relative supply of college workers through the early 1980s, then a plateau until about 1990 when the aggregate supply returns to an increasing trend, though at a slower rate.

The increasing trends in Figures 1.1 and 1.2 indicate a positive correlation between the wage gap and aggregate relative supply. As a result of diminishing marginal returns, a market with more college workers, *ceteris paribus*, implies a lower marginal product of college labor. If workers are paid their marginal product, aggregate relative supply and the wage gap should negatively covary. However, the negative relationship must hold empirically only if the relative demand for skilled labor remains constant. By regressing the wage gap on the aggregate relative supply and a linear time trend (representing the increase in demand for skilled labor), Katz and Murphy (1992) show that aggregate relative supply and the wage gap do negatively covary.[6]

The simple Katz and Murphy (1992) regression model is excessively parsimonious

[4] 1964 and 2009 represent the earliest and latest years of the March CPS available at the time of this writing.

[5] I exclude women to avoid challenges in modeling changes in female labor supply over the sample period that are beyond the scope of this paper

[6] Katz and Murphy (1992) estimate an elasticity of substitution between college and high school workers of 1.2 (men and women combined) and a 3.3 percent increase in the wage gap per year as a result of increasing demand. Using the 1964-2009 March CPS sample and the Katz and Murphy (1992) specification, I estimate an elasticity of substitution between college and high school workers of 2.6 (for men only) and a 1.6 percent increase per year. The larger elasticity for men and the smaller time trend is consistent with more recent work (Card and Lemieux, 2001, for example).

for several reasons. First, it does not account for other time-varying predictors of the wage gap such as the falling real minimum wage (Blackburn et al., 1991; Lee, 1999; Card and DiNardo, 2002) and the decline of unionization (Freeman, 1991; Card and DiNardo, 2002). Second, it does not permit a role for changes in the wage gap as a result of characteristics of birth cohorts or persistent effects of early labor market experiences of cohorts. Card and Lemieux (2001) show that the trends in the wage gap for young and old workers appear very different, inconsistent with a model in which only contemporaneous labor market conditions affect wages.[7]

1.2.2 Birth cohorts and trends in the college-high school wage gap

Card and Lemieux (2001) develop a model that predicts the presence of cohort effects in the wage gap. They argue that imperfect substitution between age groups (conditional on educational attainment) leads to the prediction that a cohort's relative supply of college labor will be negatively related to the wage gap. I will return to the model below, but the intuition is that if employers could perfectly substitute across age groups, then a cohort's own relative supply will be irrelevant except in-so-far as it affects the aggregate ratio. However, if an employer desires both young and old "skills," then college workers in a cohort with low relative supply will benefit.

[7]Welch (1979) and Bound and Johnson (1992) also consider models predicting different trends in the wage gap by age.

Figure 1.3 updates a similar figure in Card and Lemieux (2001) showing the trends in the wage gap by age group. The dashed vertical line at 1997 indicates the last year of the sample in Card and Lemieux (2001). In the 1980s the increase in the wage gap is only seen among the younger workers (solid line), while the increases in the wage gap after the late 1990s are most pronounced for older workers (dashed line). Though for simplicity the figure only includes workers aged 26-30 and 46-50, the trend for workers age 31-45 generally lies between the two lines, and the wage gap for workers age 51-60 appears similar to the trend for the older workers. Differences in trends in the wage gap by age are suggestive of birth cohort effects. For example, cohorts that are young in the 1980s will be older workers in the 2000s. That the older workers experience increases in the wage gap roughly 20 years after the young workers is indicative of birth cohort effects in the wage gap.[8]

Consistent with the imperfect age substitution explanation, changes in the cohort-specific relative supply roughly correspond to the trends in the wage gap by age seen in Figure 1.3. Figure 1.4 shows the trend in the cohort-specific relative supply – the log ratio of college to high school workers in a birth cohort – for the 1920 through

[8]In fact, the increases for the older workers seem to begin in the late 1990s, ahead of the prediction of a model with cohort effects. This is not necessarily evidence against cohort effects since contemporaneous market conditions, which affect both young and old workers simultaneously, could obscure the cohort component of the wage gap. I will more precisely distinguish between cohort and time effects in the wage gap in the empirical work.

1979 birth cohorts. The large increase in the wage gap for young workers in the 1980s corresponds to the dip in the aggregate relative supply. For example, the 1955 birth cohort, whose members are age 30 in 1985, is located both in the middle of the dip in the cohort-specific relative supply and in the middle of the steep increase in the wage gap for young workers.[9]

The Card and Lemieux (2001) model with imperfect substitution between age groups predicts that the increases in the relative wage for the baby boom cohorts when young will carry through to their late 40s in the late 1990s and 2000s. The fact that older workers (dashed line) see a large increase in the college-high school wage gap relative to young workers (solid line) after 1997 is consistent with the general predictions of the imperfect age substitution model

Card and Lemieux (2001) evaluate the predictive power of the model using the Katz and Murphy (1992) regression of the wage gap on a linear time trend and aggregate relative supply, but also including the cohort-specific relative supply. They estimate an elasticity of substitution between age groups of 5.0, an elasticity of substitution between college and high school labor of 1.6, and a linear time trend in the wage gap of 1.7 percent per year.[10] These results are consistent with an important

[9]The large size of the 1950-1960 "baby boom" cohorts could be responsible for much of the drop in rates of educational attainment (Card and Lemieux, 2000).

[10]I obtain similar results on the same sample, but excluding the 1960 Census. I estimate an elasticity of substitution between age groups of 3.0, an elasticity of substitution between education

role for imperfect substitution between ages in the evolution of the wage gap, but given that the birth cohorts characterized by declining rates of educational attainment were only observed through their mid-40s in the Card and Lemieux (2001) data, it is important to assess the ability of the model to predict more recent trends in the wage gap.

1.2.3 What is left unexplained in the college-high school wage gap?

The Card and Lemieux (2001) model with imperfect substitution between ages fails to explain recent trends in the wage gap, though is reasonably successful in explaining earlier trends. I apply the estimates of Card and Lemieux (2001) to generate predicted wage gaps over the entire post-1970 sample period.[11] Figure 1.5 shows the actual wage gaps and the predicted trends. The predicted trends (the bold solid and dashed lines) fall well below the actual wage gaps (the fainter lines) after 2000, especially for the older workers. For older workers, the predicted wage gap is nearly 20 log points below the actual level in 2009. The Card and Lemieux (2001) model cannot explain the magnitude of the increase in the college-high school wage gap for all workers, nor the extent of the divergence seen in the wage gap between young and

groups of 2.1, and a linear time trend of 1.7 percent per year. The differences are likely driven by dropping the 1960 Census, since my results do not change significantly if I use the wage gaps reported by Card and Lemieux (2001).

[11] Results using parameters from my replication results, or from parameters estimated on the entire sample period fit the pre-1990 data somewhat better, but do not improve the fit for the later years.

old workers since the late 1990s.

What could be responsible for the failure of the model to explain the divergence between the wage gaps for young and old workers? Unobserved time trends (such as non-linear changes in relative demand) would affect both young and old workers identically, and so can't resolve the unexplained divergence between young and old workers. Permanent effects of unobserved characteristics of cohorts would affect workers both when young and old, so any improvements in the fit for later years will also worsen the fit in the earlier years, given the successful pre-1997 fit of the model. For example, an increase in the 1955 birth cohort's wage gap will improve the fit for 50-year old workers in 2005, but will worsen the fit for 30-year old workers in 1985.

I will later show that the solution must come in the form of cohort effects that change with age.[12] Changes in the effects of cohort-specific determinants of the wage gap with age can explain why the increases in the wage gap for young workers appear even steeper as the cohorts age. Persistent (but not permanent) effects of past labor market conditions on the wage gap generate changes in wage gap age profiles across cohorts. For example, if high school workers born in the 1950s face above average labor market conditions at the age of high school graduation, then these initial labor

[12]In this discussion, I ignore the possibility that contemporaneous conditions could have different effects on workers of different ages, as there are few compelling theoretical grounds for its consideration.

market conditions could lead to lower wage gaps for workers in these birth cohorts when young. If the persistent effects of initial labor market conditions on the wage gap fade as the cohorts age, then the wage gap will grow with age, producing the unexplained increase in the wage gap for workers born in the 1950s when they are older in the late 1990s and 2000s.[13]

1.3 Prior Literature on Wage Persistence

1.3.1 Persistent effects of past labor market conditions on wage levels

There is growing evidence of large persistent effects of the history of labor marker conditions faced by a worker on contemporaneous wage levels. Baker et al. (1994) use within-firm data to show that variation in starting wages across entry cohorts has long-term effects on career paths and future wages.[14] Beaudry and DiNardo (1991) show that the contemporaneous unemployment rate is not significantly related to contemporaneous wages after conditioning on the history of unemployment rates during a job spell, indicating that past labor market conditions are important determinants of wage levels.[15] Oreopoulos et al. (2008), Kahn (2010), and Genda et al. (2010)

[13]The fade-out of persistent effects explains the change in the wage gap over time within cohorts, but does not explain the level of the wage gaps by age group in Figure 1.5. The level of the prediction in the figure is fixed to the mean wage gap in the Card and Lemieux (2001) sample period.

[14]Cohort effects in Baker et al. (1994) need not be linked to labor market conditions, as they could result from any unobserved characteristics of cohorts.

[15]Using similar specifications, Devereux and Hart (2007) do not find evidence of wage persistence in Britain though the results are robust to replication with alternative data in the US (Schmieder and von Wachter, 2010).

estimate the effects of the unemployment rate at the time of college graduation on future wages over time and each find similar results: the initial effects are large (a one percentage point decrease in the unemployment rate leads to a one to six percent increase in the wage level of college graduates), and diminish after ten years.[16]

I am aware of just two papers that explicitly consider the role of skill or education in the effects of wage persistence. Oreopoulos et al. (2008) find evidence that graduating college into higher unemployment rates in Canada leads to larger negative persistent effects on wages for lower skilled graduates. Genda et al. (2010) find that workers with a high school education or less in the US experience three percent lower wages in response to a one percentage point increase in the unemployment rate at market entry, but only for their first three years in the labor market. Persistent effects for workers with more than a high school education are smaller, but somewhat more persistent. The differential timing and size of the effects of graduating in a recession by education groups suggest a role for persistence in the college-high school wage gap.[17]

[16]Also, Jacobson et al. (1993) find persistent effects of job loss on wages, Hershbein (2010) finds persistent effects of initial unemployment rates on labor supply for women, and Oyer (2006, 2008) find initial market conditions affect career outcomes for economists and investment bankers.

[17]Persistent effects on wages by education group in Genda et al. (2010) are broken down by years of potential experience, whereas the wage gap is usually tracked by age. A re-interpretation of their results suggests that an increase in the unemployment rate at college graduation should lead to negative effects on the wage gap for young workers, while the rate at high school graduation should not be relevant. My results do not confirm this, and as I will discuss below, the difference is most likely a result of insufficient controls for correlated contemporaneous labor market conditions in Genda et al. (2010) and others.

1.3.2 Wage persistence mechanisms

The prior literature has identified three likely mechanisms behind the phenomenon of persistent wages. The first, implicit contracts, implies deviations in the wage from a worker's contemporaneous marginal product. The other two, job search with match-specific productivities and human capital accumulation, are fully consistent with a competitive spot wage market – a market in which all workers are paid a wage equal to their marginal product. Any or all of these mechanisms could generate persistence in the wage gap. Though in the empirical work I am only able to draw tentative conclusions about which of the mechanisms are responsible for persistence in the wage gap, a brief overview of the three mechanisms provides useful context.

Implicit contracts "signed" at the start of a job spell lead to wage persistence by governing the level of a worker's current and future wages (Harris and Holmstrom, 1982; Beaudry and DiNardo, 1991). The usual motivation for the presence of implicit contracts is as insurance offered by the firm to protect workers against unexpected fluctuations in marginal products. The contracted wage in any future period will be a function of expectations about future labor market conditions, conditional on market conditions at the time of hire. As a result, the contemporaneous wage paid to a worker will not generally equal the worker's contemporaneous marginal product to the firm, and will depend on labor market conditions at the time of hire.

Implicit contracts imply the presence of *ex post* rents for firms or workers, which provide incentives for cancellation. A usual refinement is to allow job mobility for workers, but to assume full commitment to contracts by firms. Then contracts will be "downwardly rigid" in the sense that the wage will rise as the marginal product rises, but will not change as a result of a drop in the worker's marginal product. The main result in Beaudry and DiNardo (1991) that the minimum unemployment rate faced is related to future wages is consistent with the presence of downwardly rigid contracts,

The remaining two mechanisms behind wage persistence are consistent with a spot wage market in which workers are paid their marginal products in all periods. Hagedorn and Manovskii (2009) and Oreopoulos et al. (2008) provide models with on-the-job-search in which workers who enter a labor market characterized by poor economic conditions receive fewer job offers and thus obtain worse job matches on average. The idiosyncratic match quality component of the worker's marginal product is persistent over the duration of the match, which could end as a result of exogenous or endogenous job destruction or if the worker obtains a better match through on-the-job search.[18]

[18]Oreopoulos et al. (2008) assume no job destruction, though this is not critical for the presence of

The predictions of match quality models of wage persistence are similar to those of implicit contract models: the effects of initial labor market conditions will have persistent effects on wages that will end with the dissolution of the match. Because improved labor market conditions will lead to better matches through on-the-job search, like implicit contracts, persistence from match quality implies that contemporaneous wages should be related to the best conditions faced since entry.

This discussion of match quality has ignored a potentially important interaction with human capital growth. If match quality affects the rate of (transferrable) human capital growth, then wage effects of initial conditions will persist and evolve across job spells (Oreopoulos et al., 2008). However, match-specific productivities are not required for human capital accumulation to result in wage persistence. Gibbons and Waldman (2006) argue that features of a labor market that affect the type of jobs that workers will accept can lead to wage persistence consistent with spot wages. If initial market conditions affect the distribution of available jobs, by shifting down the distribution of available wages, for example, then workers will be more willing to accept lower human capital growth in return for higher initial wages.

If accumulated human capital is not all firm-specific, the effects of human capi-

persistent effects of past market conditions. As Hagedorn and Manovskii (2009) show, job destruction will tend to reduce, but not eliminate persistence from job match quality.

tal accumulation will follow workers across job spells. While, in match quality and implicit contracting models, the effects of initial labor market conditions are erased once conditions improve, this is not the case in a human capital accumulation model. To the extent that early human capital investment decisions dominate those made in response to later changes in market conditions, persistence from human capital accumulation will be related to initial conditions, but not changes in labor market conditions. The distinctions between the predictions of the match and contracting models and the human capital model suggest a way to distinguish between the models, which I consider in the empirical work.

1.4 Theory

I begin this section by presenting the framework in Card and Lemieux (2001) to derive a prediction for the evolution of the wage gap over time and across birth cohorts in a setting with no wage persistence. Then I expand the model to account for wage persistence from implicit contracts, on-the-job search, and human accumulation. Finally, I discuss the implications of the model for persistent effects on the wage gap from changes in unemployment rates.

1.4.1 Production with two types of labor and imperfect age substitution

I begin with a standard production function with two types of labor (high school and college educated). I include skill-biased efficiency parameters and allow for imperfect substitution between age groups. Define a CES aggregate production function and two (cross-sectional) CES subaggregates of college and high school labor:

$$y_t = (\theta_{ht} H_t^{\rho} + \theta_{ct} C_t^{\rho})^{1/\rho} \tag{1.1}$$

$$C_t = \left[\sum_a (\alpha_{ca} \tilde{C}^{\eta}_{k=t-a}) \right]^{1/\eta} \tag{1.2}$$

$$H_t = \left[\sum_a (\alpha_{ha} \tilde{H}^{\eta}_{k=t-a}) \right]^{1/\eta} \tag{1.3}$$

where a indexes age, t indexes time, and k indexes birth cohort. $\tilde{C}_k$ and $\tilde{H}_k$ give the birth cohort-specific supply of college and high school labor, the effective number of employed college and high school workers in a birth cohort.[19] The skill-specific age group efficiency parameters, α_{ca} and α_{ha}, represent usual human capital accumulation with age and the value of age-specific skills (such as experience or youthful exuberance) assumed to be constant over time and across cohorts. The skill-biased parameters of the technology of production, θ_{ct} and θ_{ht}, are time-varying and permit

[19] A less restrictive model would allow the supplies of college and high school labor to vary within as well as across birth cohorts: $\tilde{C}_{ka} = \tilde{C}_k + u_{ka}$. This would account for the propensity of some workers to obtain a college education later in life, or enter and exit the labor market. For ease of notation I treat the within cohort error component u_{ka} as ignorable. I show that this restriction is empirically justifiable in Appendix C.

secular changes in the relative demand for skilled labor from unspecified sources, such as skill-biased technological change or the changes in the price of skill-specific capital.

The subaggregates imply an elasticity of substitution between ages $\sigma_A = 1/(1-\eta)$ and an elasticity of substitution between the two education groups $\sigma_E = 1/(1-\rho)$. Allowing η to differ for college and high school workers will produce separate elasticities of substitution between ages for these groups.[20]

Taking derivatives gives the marginal products of high school and college labor in cohort k at time t:

$$MP^c_{kt} = \frac{\partial y_t}{\partial \tilde{C}_k} = y_t^{1-\rho}\theta_{ct}C_t^{\rho-\eta}\alpha_{c,t-k}\tilde{C}_k^{\eta-1} \tag{1.4}$$

$$MP^h_{kt} = \frac{\partial y_t}{\partial \tilde{H}_k} = y_t^{1-\rho}\theta_{ht}H_t^{\rho-\eta}\alpha_{h,t-k}\tilde{H}_k^{\eta-1} \tag{1.5}$$

where I've replaced the age subscript on the age efficiency parameters with $a = t - k$.

Efficiency in a spot wage market implies that wages paid will equal the marginal products. Then the log wage gap is given by:

$$gap_{kt} = \ln\left(\frac{\theta_{ct}}{\theta_{ht}}\right) + (\rho-\eta)\ln\left(\frac{C_t}{H_t}\right) + \ln\left(\frac{\alpha_{c,t-k}}{\alpha_{h,t-k}}\right) + (\eta-1)\ln\left(\frac{\tilde{C}_k}{\tilde{H}_k}\right) + \epsilon_{kt} \tag{1.6}$$

where ϵ_{kt} represents determinants of the wage gap outside of the model.

[20] Card and Lemieux (2001) find no empirical support for separate elasticities, and though I do find evidence for small differences, reported in Appendix C, I ignore this here to simplify presentation.

An important restriction imposed by the model is the absence of a cohort-age interaction. As I argued above, the failure of the model in equation (1.6) to predict recent trends in the wage gap (seen in Figure 1.5) is evidence that such an interaction is important. Persistence in the wage gap is one source of a cohort-age interaction.[21]

1.4.2 A wage gap model robust to wage persistence

Wage persistence from implicit contracts, job match quality, and human capital accumulation, will add a cohort-age interaction component to the wage gap. The spot wage market models (match quality and human capital) alter the technology of production described in (1.1), (1.2) and (1.3), while an implicit contract model affects the wage setting mechanism, leaving the technology of production unchanged.[22] Because the general predictions of these models for the wage gap are similar, for the purpose of discussion, I will focus on the match quality mechanism.

The productivity of a worker at a given age depends on the quality of the current match, which in turn depends on initial labor market conditions and past improvements in labor market conditions. The shape of within-cohort age profiles depends

[21]There do exist other theoretical sources of cohort-age interactions in the wage gap. I describe and empirically assess some of these in Appendix C. While I find some support for more flexible parameterizations of equations (1.1), (1.2) and (1.3), these are not sufficient to account for the unexplained portion of the wage gap.

[22]The presence of implicit contracts will affect the wage paid, though neither the underlying structure of the technology of production, nor the marginal products given by equations (1.4) and (1.5). Instead, the worker faces a new utility maximization problem in which she must choose the wage to be paid in each period over the life of the contract subject to the constraint that the firm receives non-negative expected total profits.

on a worker's labor market history. For example, a worker entering the market into poor labor market conditions will obtain a worse job match than a worker entering into good conditions. If labor market conditions do not improve, the worker will be less likely to obtain a better match. So age profiles will be steeper in cohorts with poor initial conditions, which improve over time.

Define γ_{cak} and γ_{hak}, the match quality component of the productivity of a college or high school worker in cohort k at age a. Match quality for college workers, γ_{cak}, is a function of past (i.e. between market entry and age a) labor market conditions faced by cohort k. The augmented education group-specific subaggregates in equations (1.2) and (1.3) are the following:

$$C_t = \left[\sum_a (\gamma_{ca,k=t-a} \alpha_{ca} \tilde{C}^{\eta}_{k=t-a}) \right]^{1/\eta} \tag{1.7}$$

$$H_t = \left[\sum_a (\gamma_{ha,k=t-a} \alpha_{ha} \tilde{H}^{\eta}_{k=t-a}) \right]^{1/\eta} \tag{1.8}$$

The wage gap equation now includes a term representing the dependence of the contemporaneous wage gap on past labor market conditions faced by college and high school workers:

$$gap_{kt} = \ln\left(\frac{\theta_{ct}}{\theta_{ht}}\right)+(\rho-\eta)\ln\left(\frac{C_t}{H_t}\right)+\ln\left(\frac{\alpha_{c,t-k}}{\alpha_{h,t-k}}\right)+\ln\left(\frac{\gamma_{c,t-k,k}}{\gamma_{h,t-k,k}}\right)+(\eta-1)\ln\left(\frac{\tilde{C}_k}{\tilde{H}_k}\right)+\epsilon_{kt} \tag{1.9}$$

Though the discussion above focused on match quality, equation (1.9) is general enough to account for wage persistence from human capital accumulation or implicit contracts. In the case of implicit contracts, the γ parameters represent the dependence of the contracted wage on initial conditions and upward wage adjustments from improved conditions.[23] In the case of human capital accumulation, the γ parameters represent the portion of contemporaneous productivity determined by the human capital growth opportunities in the prior and current job.

1.4.3 Persistence in the wage gap from unemployment

The unemployment rate serves as a proxy for labor market tightness, which can affect job match quality, human capital accumulation opportunities, and the level of a contracted wage. Unemployment rates enter equation (1.9) through the wage persistence parameters $\gamma_{c,t-k,k}$ and $\gamma_{h,t-k,k}$.[24]

[23]Deviations from spot wage markets in wage setting from implicit contracting complicates the interpretation of the structural parameters (e.g. the elasticities of substitution between ages and education), a problem beyond the scope of this paper.

[24]It is not clear *a priori* how unemployment at entry will differentially affect wages of college and high school workers in terms of match quality, human capital accumulation or expected future marginal products.

I consider only the national annualized aggregate unemployment rates.[25] As a result, the unemployment rate faced by a cohort at any given age is the same for college and high school workers. However, because high school workers enter at an earlier age, I observe differences in initial unemployment rates for college and high school workers in a cohort. I assume that college workers enter the labor market at age 23 and high school workers enter at age 19.[26]

I model wage the wage persistence term in the wage gap, $\ln\left(\frac{\gamma_{c,t-k,k}}{\gamma_{h,t-k,k}}\right)$, as a function of the history of unemployment rates faced by college and high school workers between entry and age $t-k$. I include labor market conditions at the age of entry, and the best labor market conditions since entry. Specifically, I include initial unemployment rates for college and high school workers, U_{k+23} and U_{k+19}. I also include contemporaneous unemployment rates, U_t.

In some specifications, I also include and the lowest unemployment rates faced between entry and age $t-k$ for college and high school workers, $U^*_{t-k,k+23}$ and $U^*_{t-k,k+19}$ to distinguish between different mechanisms of wage persistence.[27] The asymmetries

[25]Skill-group specific unemployment rates are available for a portion of the years under consideration, though further sample restrictions will be particularly problematic given the need to observe unemployment rates at young ages for birth cohorts. Given the likely presence of spillover effects across education groups of unemployment levels, the aggregate unemployment rate likely captures the relevant variation.

[26]I show results for alternative choices of entry ages in Appendix D.

[27]In principle, the entire history of unemployment rates faced by a cohort belongs in γ, though such a model is empirically intractable.

of the match and contracting models, predicting that improvements in labor market conditions will lead to higher wages for college and high school workers but worsening conditions will not affect wages, are not present in the human capital accumulation model. In the match and contracting models, including the lowest unemployment rate faced since a cohort's entry into the labor market should reduce dependence of the wage gap on initial conditions. In the human capital accumulation model, the effects of initial conditions dominate.

Finally, I allow the effects of initial and lowest unemployment rates to vary with age. In the match quality model, for example, the interaction between age and initial unemployment accounts for the effects of exogenous job destruction, which will tend to reduce dependence of wages on initial conditions with age, regardless of whether unemployment rates improve below initial levels.

1.5 Data and Empirical Strategy

1.5.1 Current Population Survey March Supplements

Using the 1964-2009 Current Population Survey March Supplements,[28] I construct two samples: a wage sample used to obtain cohort-year specific wage gaps, and a

[28] I exclude the 1963 March CPS as it does not include data on respondents' years of education. The 1962 March CPS does include all necessary variables, but the supplement weights appear to be inconsistent with the weights in the rest of the series. When appropriate, I use corrections to the March CPS series and weights documented by IPUMS CPS (King et al., 2010).

supply sample used to construct the various labor supply measures described below. The wage sample contains full-time, full-year employed men age 26-60 born between 1920 and 1979 with exactly 12 or 16 years of completed education for a total of about 470,000 individuals.[29] The broader supply sample includes all full-time, full-year employed men and women age 20-64 regardless of educational attainment, for a total of about 1.8 million individuals. I follow Lemieux (2006) in constructing the sample and cleaning wages and years of completed education.

To capture variation in the college-high school wage gap across birth cohorts and across time, I calculate wage gaps at the finest level possible, within cohort-year (or, equivalently, age-year) cells. I capture the difference in within-cell mean log wages between male[30] workers with exactly 16 years of completed education from those with exactly 12 years of completed education in a regression of log weekly wages on an indicator for college and an indicator for non-white.[31] I calculate log weekly wages by dividing annual wages and salary by 52.[32] I construct time- and cohort-varying supply

[29]Due to changes in the 1970s in how respondents were asked to report usual hours of work and weeks worked last year, it is necessary to restrict the sample to full-time and full-year workers to create a consistent sample over time.

[30]While changes in education wage gaps for women are no doubt of great interest, models of changes in the demand and supply of female labor are beyond the scope of this paper. Also, as Lemieux (2006) notes, the necessity of limiting the sample to full-time, full-year workers is particularly problematic for women, because of trends in the fraction of women employed part-time. Despite this measurement problem, I do include women in the labor supply measures described below, though the main results are robust to excluding them.

[31]The main results are robust to alternative wage gaps including those which do not condition on non-white and those which ignore the supplement weights.

[32]Lemieux (2006) documents significant and increasing measurement error in wages over time in the March CPS, most likely generated by mis-measured annual earnings of hourly wage earners.

measures from the supply sample, which are then merged with the age-cohort specific wage gaps, covering 1,365 cohort-year cells across 60 birth cohorts. The panel of cohorts is not balanced; only 22 of the 60 birth cohorts are observed from 26 through age 50. I use the supplement weights in constructing all wage and supply measures used in the analysis.[33] See Appendix A for further details about the sample.

1.5.2 Supply measures

Supply measures are in units of effective supply of college or high school workers and include workers of all education levels.[34] To create effective supply measures, I estimate and apply weights by education group (less than high school, high school, some college, more than college), constructed as in Card and Lemieux (2001). Weights for each education group are based upon wages (across all time periods)[35] within the group relative to either high school or college wages. The supply of high school workers is then a weighted sum of workers with fewer than 12 years of completed

While this could tend to lead to upward bias in the college-high school wage gap in later years, this is not expected to lead to bias in the main persistence results below, since these estimates control flexibly for changes in the wage gap over time. The alternative data source, the CPS Merged Outgoing Rotation Groups (MORG), is not ideal for this analysis since it is only available beginning in 1973, limiting the number of birth cohorts observed with labor market earnings, and severely limiting the number of cohorts observed from age 26 through to age 50 or 60.

[33]The CPS supplement weights are designed to sum up to US population estimates in a cross-section. I rely on the weights to aggregate supply counts across survey years for cohort-specific supply measures.

[34]Given the parameterization of the subaggregates defined in equations (1.2) and (1.3), a true effective supply measure would also employ estimates of the efficiency and substitution parameters. Though I do not do this, Card and Lemieux (2001) found this to be unimportant for the wage gap specifications they considered.

[35]The main results in this paper are robust to weights created at the year level, and those which down-weight women relative to men based on relative wages.

education, workers with exactly 12 years of education, and a portion of workers with some college education. Similarly, the supply of college workers is a weighted sum of the remaining portion of workers with some college education, workers with exactly 16 years of completed education, and workers with more than 16 years of education.

Workers with some college (13-15 years of completed education) are apportioned between the effective college and high school supply based upon the mean wage of these workers relative to college and high school workers wages. In practice, because the some college wage is closer to the high school wage than the college wage, approximately 2/3 of workers with some college education are apportioned to the effective high school supply, and 1/3 are apportioned to the effective college supply.

The primary analysis includes two distinct relative supply measures: the aggregate relative supply of college labor and the cohort-specific supply of college labor. In some specifications I also include the aggregate supplies of college and high school labor at a birth cohort's age of labor market entry. I will briefly describe the construction of each of these supply measures from the March CPS. First, I calculate the aggregate relative supply of college labor, $\ln\left(\frac{C_t}{H_t}\right)$, as the log ratio of effective college versus high school full-time, full-year employed male and female workers of age 20-64 observed in year t of the March CPS.

Second, the cohort-specific relative supply, $\ln\left(\frac{\tilde{C}_k}{\tilde{H}_k}\right)$, is the log ratio of effective college versus high school full-time, full-year employed workers observed in cohort k. Because the March CPS samples by year and not by cohort, I do not observe workers of all ages 20-64 for every cohort. If a significant number of workers obtain college education later in life, then the observed cohort-specific relative supply will be biased for cohorts missing the oldest or youngest workers.

I use a regression-based correction to construct cohort-specific relative supplies that are constant within birth cohort. I regress the cohort-age group specific relative supply on birth cohort and age group fixed effects and predict the relative supply for each cohort at age 36-40 (the results are seen in Figure 1.4, discussed earlier). This provides a measure of the cohort-specific relative supply of college labor by age 40 even for birth cohorts for whom individuals aged 36-40 are not observed in the March CPS.[36]

Finally, for use in some robustness checks, I construct aggregate supplies of college and high school labor at the birth cohort's age of labor market entry, $\ln C_{k+23}$ and $\ln H_{k+19}$. For example, college workers born in 1950 face the 1973 aggregate supply

[36]The identifying assumption is that the fraction of workers completing a college education by a given age is constant across birth cohort. Appendix Figure C.1 shows that the cohort-age group supplies are roughly parallel, lending support to this approach. The described regression produces an R-squared of 0.996 implying that very little important variation is lost though this simplifying assumption. In Appendix C, I show that using unadjusted cohort-specific relative supplies does not affect the main results.

of college labor (calculated as described above) when entering the labor market, and high school workers born the same year face the 1969 aggregate supply of high school labor.[37]

1.5.3 Unemployment data

National unemployment rates for individuals age 16 and older from 1948-2009 are obtained from the Bureau of Labor Statistics.[38] As discussed above, a key regressor related to wage persistence is the lowest unemployment rate faced between entry and a given age. For many cohorts this is identical for college and high school workers, but for some cohorts high school workers enter the market in time to face better conditions than college workers. I plot the levels of unemployment faced at ages 19 and 23 in Figure 1.6. High school workers face lower initial unemployment rates in cohorts for which the dashed line (age 19 for high school workers) is below the solid line (age 23 for college workers). There is considerable variation in initial unemployment rates over the sample period ranging from just below 4 percent to nearly 10 percent. Also, though the broad trends in unemployment appear cyclical, the considerable cohort-

[37]For cohorts born before 1945 I cannot construct the aggregate supply at entry for high school workers because in the earliest year of the March CPS series, 1964, workers in these cohorts are not yet 19 years old. As a result, regression models including the aggregate relative supply at entry are limited to cohorts born in 1945 or later.

[38]I do not observe the unemployment rate at entry (age 19) for high school workers born prior to 1929. As a result, most specifications in the empirical work will be limited to cohorts born in 1929 and later.

to-cohort variation in levels of unemployment rates will allow for plausible estimates of small changes in unemployment rates.

1.5.4 Primary estimating equation

Equation (1.9) and the discussion of modeling the wage persistence term, $\ln\left(\frac{\gamma_{c,t-k,k}}{\gamma_{h,t-k,k}}\right)$, as a function of initial and lowest unemployment rates since entry suggests the following regression:

$$
\begin{aligned}
gap_{kt} = {} & \delta_1 year_t + \beta_1 \ln\left(\frac{C_t}{H_t}\right) + \beta_2 \ln\left(\frac{\tilde{C}_k}{\tilde{H}_k}\right) + \delta_2 cohort_k + \beta_3 U_t \\
& + \sum_{a=1}^{N_A} 1(a_{t-k} = a)(\lambda_{1a} + \lambda_{2a} U_{k+19} + \lambda_{3a} U_{k+23} + \lambda_{4a} U^*_{a,k+19} + \lambda_{5a} U^*_{a,k+23}) + \epsilon_{kt}
\end{aligned}
\tag{1.10}
$$

The linear time trend represents contemporaneous changes in relative demand, $\ln\left(\frac{\theta_{ct}}{\theta_{ht}}\right)$. From equation (1.9) and the definition of the elasticity parameters, the elasticity of substitution between age groups is given by $\sigma_A = -1/\beta_1$ and the elasticity of substitution between education groups is given by $\sigma_E = -1/(\beta_1 + \beta_2)$. The linear trend across birth cohorts represents trends in unobserved permanent characteristics of cohorts and unobserved initial labor market conditions.[39] I model the relative age efficiency parameters, $\ln\left(\frac{\alpha_{c,t-k}}{\alpha_{h,t-k}}\right)$, as a set of age group fixed effects, λ_{1a}, grouped into

[39]For example, the cohort trend could represent changes in the relative skills of college and high school workers across cohorts as a result of changes in school quality or other compositional changes. The trend could also represent effects of changes in initial relative demand for college labor.

N_A categories (e.g. $N_A = 7$ categories of five years each and a_{t-k} takes on a value of 1-7). The effects of initial and lowest unemployment rates vary across age groups.

The variation in the wage gap to be explained includes changes in the college-high school wage gap over time, with age, and across birth cohorts. Further, because effects on the wage gap as a result of initial market conditions need not be permanent, the cohort-age interactions in equation (1.10) track the within-birth cohort wage gap as the cohort ages. A non-parametric decomposition of age, time and cohort effects (even ignoring the interaction) is not identified without further restrictions, because knowing any two of age, time, and cohort, one can determine the third (Heckman and Robb, 1985; Beaudry and Green, 2000). However, given sufficiently strong restrictions on the age effects, one can identify not only the main effects, but also the within-cohort age profiles separately from a time trend.

Though I consider more flexible parameterizations, equation (1.10) places linear restrictions on the time and cohort trends, and I will consider age groupings with $N_A = 4$ or sometimes $N_A = 7$ categories. Despite the restrictions imposed on time, cohort and age effects in equation (1.10), the model is more flexible than those in prior work on the wage gap. For example, Card and Lemieux (2001) estimate regressions that are similar to the specification in equation (1.10), but exclude the unemployment rate regressors, age group interactions, and the linear cohort trend.[40] Estimates of the time trend and age efficiency parameters from the specifications in Card and Lemieux (2001) will be biased in the presence of persistent effects of initial conditions on the wage gap, because in the true model, the shape of age profiles changes across cohorts. Estimates of the parameters related to the elasticities of substitution between ages and education, β_1 and β_2, also will exhibit bias from this source of model mis-specification, though the sign of the bias is unpredictable.

1.5.5 Two-stage estimation procedure

One concern in identifying the key parameters of the model is plausible separation of birth cohort and time effects. Though one could estimate equation (1.10) in a single stage OLS regression, this would lead to concerns that unobserved birth cohort-specific contributions to the wage gap would bias parameters on time-indexed regressors, and unobserved time-varying contributions would bias parameters on cohort-indexed regressors. For example, unobserved initial labor market conditions generating persistence in the wage gap, which are correlated with contemporaneous aggregate supply, would lead to bias in the effect of contemporaneous aggregate supply on the

[40]Plausible identification of within-cohort age profiles in the wage gap requires following a reasonably large sample of cohorts from youth to retirement. Without the 12 additional cross-sections of data available since the analysis in Card and Lemieux (2001), identification of the more general model in (1.10) would be challenging.

wage gap. A primary goal of this paper is to separately identify contemporaneous determinants of the wage gap and persistent effects on the wage gap from past labor market conditions. Thus it is imperative that the relationship between cohort-indexed regressors and the wage gap is not confounded by unobserved contemporaneous labor market conditions correlated with past conditions. For these reasons, I use a two-stage estimation procedure.

In the first stage, I regress the cohort-time specific wage gap on a full set of year fixed effects and birth cohort-age group fixed effects.

$$gap_{kt} = \nu_t + \mu_{ak} + \epsilon_{kt} \tag{1.11}$$

Here a indexes workers into four age groups: 26-30, 31-40, 41-50, and 51-60, though I later consider a nested specification with seven age groups. Some grouping or functional form restriction is required since unrestricted single-year age effects would not be identified in a model that also includes single-year birth cohort and year effects.[41] The result of the first stage regression is a semi-parametric decomposition of the wage gap into time and birth cohort-age group components.

[41] The choice of bin width for age groups is arbitrary. Smaller bins could reduce bias at the cost of higher variance. The results are not affected by using the seven versus four age group decomposition, suggesting that the bias concern is unimportant.

Using the decomposition results, in a second stage I estimate each of the following:

$$\hat{\nu}_t = \delta_0 + \delta_1 year_t + \beta_1 \ln\left(\frac{C_t}{H_t}\right) + \beta_3 U_t + \epsilon_t \tag{1.12}$$

$$\begin{aligned}\hat{\mu}_{ak} &= \delta_2 cohort_k + \beta_2 \ln\left(\frac{\tilde{C}_k}{\tilde{H}_k}\right) \\ &+ \sum_{a=1}^{N_A} 1(a_{t-k} = a)(\lambda_{1a} + \lambda_{2a} U_{k+19} + \lambda_{3a} U_{k+23} \\ &+ \lambda_{4a} U^*_{a,k+19} + \lambda_{5a} U^*_{a,k+23}) + \epsilon_{at}\end{aligned} \tag{1.13}$$

Equations (1.12) and (1.13) separate the portion of the main estimating equation (1.10) with time-varying covariates from the portion with cohort and age-varying covariates.

Because the dependent variable in each of regressions (1.12) and (1.13) are estimated, one could use variance weighted least squared (VWLS), weighting by the estimated variance from the first stage. However, this approach will provide inefficient estimates if there is a group component to the residual (Moulton, 1986; Dickens, 1990). In this context, there is reason to expect significant within-year and within-birth cohort correlation in the error terms.

I estimate OLS models with standard errors robust to clustering at the cohort level whenever the model pools multiple age groups, and heteroskedasticity robust standard errors otherwise. Dickens (1990) finds that results from standard heteroskedasticity-robust OLS are very similar to those implementing an alternative efficient GLS procedure. In Appendix B, I implement a GLS procedure similar to one suggested by Dickens (1990) and obtain slightly smaller standard error estimates, suggesting that the simpler approach is conservative.

To address a further concern for estimation, in Appendix B, I show that accounting for the presence of AR(1) serial correlation of errors across birth cohorts does not affect the main results.[42]

1.6 Results

1.6.1 First stage results

Figures 1.7a and 1.7b show the estimates of the fixed effects in equation (1.11), the first stage decomposition of the college-high school wage gap into time and birth cohort-age group components. The decomposition separates the portion of the wage gap explained by time-varying regressors, affecting all birth cohorts and age groups identically, and the portion explained by cohort-age group specific regressors. The adjusted year effects measure the transient effects of contemporaneous labor market conditions on the wage gap, while the adjusted cohort-age group effects measure

[42] I do not address bias or inefficiency in estimation due to the use of estimated regressors in the case of the supply measures. However, given reasonably large samples underlying the aggregate and cohort supply measures, sampling error in the regressors is not likely to play an important role.

persistence in the wage gap and the effects of permanent characteristics of cohorts. It is worth discussing the first stage results in detail, since the fixed effect estimates serve as the dependent variables in the second stage regressions discussed below.

The adjusted year effects in the wage gap show a steeper upward trend than in the unadjusted cross-sections (Figure 1.7a). In a model of the wage gap with no cohort or age effects, the cross-sectional wage gap would be identical to the adjusted gap, because the cohort and age effects would have no explanatory power. The differences between the cross-sectional and adjusted time trends in the wage gap are indicative of large cohort and age effects. The differences are largest in the 1960s and 1970s. For example, in the cross-sectional trend the wage gap increased only by about one-half percent between 1964 and 1980. Adjusted for the characteristics and experiences of birth cohorts, the increase over the same period is about 6 percent.[43]

Figure 1.7b illustrates two facts about the cohort-age group effects in the wage gap. First, as predicted by the presence of imperfect age substitution, the cohort component of the wage gap is inversely related to the cohort-specifc relative supply of labor. The bold solid line shows the cohort effects at age 26-30. The level gives the time-adjusted wage gap for the age group, so the value of -0.15 for the 1950 birth

[43]Interestingly, the adjusted trend much more closely approximates the rate of increase in the 90/10 percentile wage differential (as seen in Autor et al. (2008)). A large portion of the difference between the adjusted and actual cross-sectional wage gap can be accounted for by changes in cohort-specific relative supply, which may not affect the tails of the income distribution.

cohort means that the cohort's actual wage gap in a given year is 15 log points lower than the contemporaneous gap in the adjusted time trend. In other words, the 1950 birth cohort of college graduates at ages 26-30 earns relative wages which are roughly 15 percent lower than the relative wages of other cohorts of college graduates at the same age and holding contemporaneous labor market conditions constant. The cohort effects in the wage gap at age 26-30 are decreasing for cohorts born before 1950, increase for birth cohorts in the 1950s, then return to a decreasing trend. As predicted by the model in Card and Lemieux (2001), the picture inversely mirrors the trends in the cohort-specific relative supply in Figure 1.4.

The second fact apparent from Figure 1.7b is that, in contrast to predictions in prior models of the wage gap, the birth cohort effects in the wage gap change with age, especially between age 26 and 40. The vertical distance between the lines for each age group describes the within-cohort age profile in the wage gap. A larger distance implies a steeper profile.[44] While the downward trend for age 31-40 (bold dashed) prior to the 1945 birth cohorts appears reasonably similar to the trend for age 26-30, this is not the case for newer cohorts. For example, the 1940 cohort faces an approximate 1 percent cohort-specific wage gap penalty, which is essentially constant

[44]The cohort trends by age group in Figure 1.7b do not span all birth cohorts under analysis because a panel of cohorts constructed from cross-sectional surveys will be unbalanced.

through age 41-50. However, the 1950 cohort faces a 15 percent wage gap penalty at age 26-30, but this is reduced to 2 percent by age 31-40.

The changes in within-cohort age profiles represent the key variation in the wage gap to be explained by persistent effects of past labor market conditions. As discussed above, the Card and Lemieux (2001) model provides no predictions for a cohort-age group interaction. In a model with no interaction, the four cohort profiles in Figure 1.7b would be parallel. The steep within-cohort age profiles for cohorts born since the mid 1940s are responsible for the unexplained divergence between the wage gaps for young and old workers since the late 1990s seen in Figure 1.5.

To test whether the changes in within-cohort age profiles evident in Figure 1.7b represent a real feature of the data or are simply sampling noise, I estimate first stage models with alternative restrictions. Table 1.1 gives summary statistics for each of three nested first stage decompositions.[45] The most restrictive model, in column (1), imposes parallel cohort trends (allowing only main cohort and age group effects, excluding their interaction). Because allowing just four age groups is excessively restrictive when excluding the cohort-age group interaction, I allow seven age groups. The restricted specification in column (1) is consistent with the restrictions of the

[45]The model in column (1) is nested within the model in column (3), and the model in column (2) is nested within the model in column (3).

Card and Lemieux (2001) model. The second specification, in column (2), is my primary specification in equation (1.11), shown in Figures 1.7a and 1.7b, which allows an interaction between the cohort effects and four age groups. The final and least restrictive specification, in column (3), allows an interaction between the cohort effects and seven age groups.

The model excluding the cohort-age group interaction, in column (1), has an R-squared of 0.648, while the most flexible model, in column (3), has an R-squared of 0.738, indicating that the cohort-age interaction has explanatory power. The second to last row in Table 1.2 reports p-values from the F-tests of nested specifications. The two restricted models are easily rejected against the model with seven age groups interacted with cohort effects, indicating an important role for changes in within-cohort age profiles in the evolution of the wage gap.[46] Though the model in column (2) is rejected, I will use this decomposition as my baseline decomposition for the second-stage models. All of the second-stage results are robust to the finer decomposition, and the coarser cohort-age group interaction simplifies presentation.

The rejection of the restrictions on age and cohort effects in Card and Lemieux

[46]Because the wage gaps are estimated, a χ^2 test based on the estimated standard errors on the cohort-time specific wage gaps is possible using VWLS. As discussed above, in the presence of group-specific errors these tests are based on an inefficient estimator (whereas the F-tests are robust to this problem). The last row of Table 1.2 presents p-values from the χ^2 tests of the model restrictions, which fail to reject at the 5 percent level. While the baseline model for the two tests is different, the lack of power in the tests which exploit the standard errors is sufficient reason to eschew the VWLS approach in favor of techniques known to improve efficiency.

(2001) indicate that the changes across cohorts in age profiles of the wage gap seen in Figure 1.7a are statistically significant. Prior work on the evolution of the wage gap ignoring the cohort-age group interaction leaves important trends in the wage gap unexplained and is subject to bias if the causes of the interaction effects are related to key regressors. Since in prior work, accounting for wage persistence has reduced the role of contemporaneous labor market conditions in predicting wages, there is reason to suspect that previous work on the evolution of the wage gap has overstated the effects of changes in relative demand and supply. Thus, I next estimate second-stage models of the wage gap robust to cohort-age group interactions. Then I proceed to estimate full versions of the second-stage equations (1.12) and (1.13) including cohorts' unemployment rate histories.

1.6.2 Estimates of models with imperfect substitution

Table 1.2 presents estimates of the second-stage equations (1.12) and (1.13) with imperfect substitution between age and education, but excluding the unemployment rate regressors. I also include comparable estimates from the single-stage regression in equation (1.10) to demonstrate the importance of controlling for unobserved determinants of the wage gap. The estimates of the elasticity of substitution between age groups range between 3 and 4 regardless of the specification, but the elasticity of substitution between education groups and the estimate of the linear time trend in the wage gap (representing changes in relative demand for college labor) are sensitive to specification.[47]

Column (1) of Table 1.2 shows the results of a single-stage regression of the cohort-time specific wage gap, similar to regression models in Card and Lemieux (2001) with imperfect substitution between age and education. I obtain estimates of the elasticities of substitution between education and age of 1.7 and 3.4, respectively. The elasticity of substitution between age groups of 3.4 implies that a 10 percent change in the relative wage of young and old workers leads to a 3.4 percent change in the ratio of young and old workers. In a reduced form sense, the coefficient on the cohort-specific relative supply of -0.296 (the estimate the elasticity of substitution between age groups is based on) says that a 10 percent increase in the relative supply of college labor in the cohort leads to a 3 percent reduction in the wage gap for the cohort. The estimated linear time trend in the wage gap represents wage gap effects from changes in relative demand of 2.2 percent per year.

Columns (2)–(6) show how more flexible specifications affect these key parameters. In these columns, I estimate specifications that account for unobserved characteristics

[47]The elasticity estimates are obtained by taking the inverse of the negative of the coefficient on cohort-specific relative supply (age) and the inverse of the negative of the sum of the coefficients on aggregate relative supply and cohort-specific relative supply.

of cohorts, unobserved contemporaneous labor market conditions, and wage persistence. Even ignoring any effects of wage persistence in the wage gap, unobserved characteristics of cohorts could lead to bias in any or all of the parameters. I include a linear cohort trend in column (2), representing wage gap effects from changes school quality or any other unobserved permanent characteristics of cohorts.[48]

Accounting for a linear cohort trend in column (2), reduces the linear time trend to 1.9 percent per year, but doesn't affect the elasticity estimates. The estimated cohort trend in the wage gap is an increase of four percent for every ten birth cohorts, small relative to the time trend, but statistically significant. However, if effects from unobserved cohort and time characteristics are non-linear, the results in column (2) will also be biased. This would be the case if, for example, school quality (college vs. high school) is not linearly increasing over the sample period, since deviations from the linear trend across cohorts will be correlated with the time trend and age group fixed effects.

The problem of bias from non-linear unobserved time-varying determinants of the wage gap is partially resolved in the single-stage regression in column (3), which replaces the time-indexed regressors with year fixed effects. The estimates of coef-

[48]The cohort trend could also represent persistent effects from relative demand faced by the cohort at labor market entry or other approximately linear trends in initial labor market conditions across cohorts.

ficients on cohort-indexed regressors in column (3) are similar to those in column (2), suggesting that the linear restriction on unobserved time-indexed determinants of the wage gap sufficiently controls for unobserved contemporaneous labor market conditions. The regression in column (3) does not solve the problem of bias from unobserved cohort-specific determinants of the wage gap.

The two-stage procedure allows estimation of both coefficients on cohort-indexed regressors robust to all unobserved time effects, and coefficients on time-indexed regressors robust to all unobserved cohort effects, or cohort-age group interaction effects. I present results from the two-stage procedure in columns (4), (5), and (6). Coefficients above the horizontal line are estimated in a regression of the year fixed effects from the indicated first-stage decomposition on the time-indexed regressors. Coefficients below the line are estimated from a regression of cohort or cohort-age group fixed effects from the indicated first-stage decomposition on cohort- and age-indexed regressors. Columns (4), (5), and (6) of Table 1.2 use the results of the first stage decompositions from columns (1), (2), and (3) in Table 1.1, respectively.

Columns (2) and (4) both restrict cohort and age effects to enter separately, but column (4) implements the two-stage procedure to reduce bias. The results are similar, though column (4) produces a larger elasticity of substitution between age groups, 4.0 versus 3.4.[49] However, accounting for changes in age profiles in the wage gap across cohorts, I find smaller linear time trends and reduced response of the wage gap to aggregate relative supply. In the most flexible model, shown in column (6), the aggregate relative supply no longer enters the model significantly and the time trend is reduced to one percent per year. The reduction in the time trend represents a nearly 50 percent reduction in the contemporaneous effect of changes in relative demand for college educated labor on the wage gap, relative to column (2). The elasticity of substitution between college and high school workers based on the results of column (6) is 3.1, much larger than estimates in prior work.[50]

The small role for contemporaneous labor market conditions after accounting for cohort-age group variation in the wage gap implies that prior work on the evolution of the wage gap has overstated the effects of contemporaneous labor market conditions, but is consistent with prior results in the literature on persistence in wage levels from unemployment rates. Though the estimates of coefficients on cohort-indexed variables are similar across specifications, the fraction of the variance in the cohort-

[49]The two-stage procedure also reduces bias from correlation between unobserved cohort- and time-indexed determinants of the wage gap that are also correlated with the key regressors. This explains the small differences between the results in columns (3) and (4), which both adjust for unobserved year effects.

[50]Despite the smaller role for both contemporaneous relative demand and relative supply in the evolution of the wage gap after accounting for cohort-age group interaction effects, the adjusted time trend shown in Figure 1.7a is steeper than the unadjusted trend. The steeper trend is the result of larger reductions in the negative effects of relative supply than in the positive effects of relative demand.

age effects explained falls from 91 to 70 percent between the specification with no cohort-age group interaction in column (4) and the most flexible specification in column (6). (Recall that it is the estimated dependent variable that changes across the specifications, not the second stage model.) The reduction in explanatory power of the cohort trend and cohort-specific relative supply suggests that there is room for the labor market histories of cohorts in explaining changes in age profiles in the wage gap across cohorts.[51]

1.6.3 Unemployment and persistence in the wage gap

To determine if unemployment rates have a persistent effect on the wage gap, I estimate equation (1.13), the second-stage model of cohort-age group fixed effects. In addition to what was included in Table 1.2, the model includes unemployment rates faced at the approximate age of entry for college and high school workers. I also estimate single-stage versions of the model, but do not report the coefficients on the time-indexed regressors.

Table 1.3 presents the estimated persistent effects of unemployment rates on the wage gap. I show the baseline single- and two-stage specifications with only a linear cohort trend and the cohort-specific relative supply in columns (1) and (4). The

[51] Using the efficient variance estimates from Appendx B, the faction of non-sampling error variance in the cohort-age group effects unexplained by the model in column (5) is 11 percent.

results differ from columns (2) and (4) of Table 1.2 as a result of dropping several of the earliest birth cohorts for which the unemployment rate at age 19 is not available. I also drop the age 51-60 age group from the analysis due to the small sample size (26) for the group, and for ease of presentation.[52]

Columns (2), (3) and (5) introduce the unemployment rate faced by high school educated workers at entry. In the two-stage specification in column (5), a one percentage point decline in the unemployment rate faced by high school educated workers entering the market at age 19 leads to a 1.2 percent smaller wage gap at age 26-30. The wage gap effects of age 19 unemployment do not appear to persist after age 30.

The single stage estimates in column (2) include controls for observable contemporaneous labor market conditions, including the unemployment rate, but show no effects of age 19 unemployment at any age. Replacing the contemporaneous controls with time fixed effects in column (3) produces results similar to column (5), though the point estimate for the cohort trend is smaller and the elasticity of substitution is larger. The differences between columns (2), (3), and (5) suggest that it is imperative to control for unobserved contemporaneous labor market conditions in estimates of persistence from initial unemployment rates.[53]

[52] Though some cohorts and age groups are dropped from the second-stage regression, the first-stage decomposition is estimated using the full sample.

[53] The downward bias of the effect of initial unemployment in column (2) would occur if high initial unemployment rates lead to future unobserved determinants of the wage gap which benefit

A 1.2 percent effect on the wage gap of young workers is generally consistent with the magnitude of initial unemployment effects on wage *levels* in prior literature.[54] That the effects are present 8-12 years after market entry for high school workers implies a longer period of persistence than has been found for wage levels of high school workers in prior work. Though not all of the persistent effects of age 19 unemployment need work through high school educated workers' wages, negative effects on college educated workers' wages from higher age 19 unemployment rates would tend to reduce the total effect on the wage gap.

Column (6) adds to the model the unemployment rate faced by college educated workers entering the market at age 23. The initial unemployment rate for college educated workers does not enter significantly for young workers. Though unemployment at age 19 and 23 for a cohort are positively correlated (0.32), the lack of significant results for the age 23 unemployment rate is not driven by this correlation. Excluding the age 19 unemployment rate from column (6) does not produce significant effects of age 23 unemployment on the wage gap.

---

high school workers relative to college workers. This source of bias is a concern for prior work on persistence in wage levels. Oreopoulos et al. (2008), Kahn (2010), and Genda et al. (2010) use regional variation in unemployment rates at graduation, but include just year fixed effects (instead of region x year), leaving open the concern that past unemployment rates in the region are correlated with future regional labor market conditions. Including year fixed effects at the level of variation in unemployment rates resolves this concern. I discuss controls for unobserved contemporaneous labor market conditions further in Appendix D.

[54]However, the effect on the wage gap is larger than is suggested by the estimates of persistence for college and high school educated workers in Genda et al. (2010). I discuss this further in Appendix D.

The lack of any significant persistent effect of unemployment of the unemployment rate faced by college educated workers entering the market at age 23 on the wage gap is in contrast to prior work on wage levels of college workers. One explanation for the difference is that the age 23 unemployment rate affects both college and high school workers, canceling out the total effect on the wage gap.

Persistent effects of unemployment can go a long way to explaining the residual increase in the wage gap for older workers since the late 1990s. Figure 1.8 shows the predicted trends by age in the wage gap using the regressions in column (5) of Table 1.3. The prediction also includes the unreported predicted values from equation (1.12), the second-stage model of time effects, including the contemporaneous unemployment rate. The model with initial unemployment rates improves the fit for older workers since 1997 by 35% and almost 60% for younger workers, relative to the fit of the Card and Lemieux (2001) model in Figure 1.5.[55]

The success of the fit is a result of lower relative wages due to low initial unemployment rates faced by high school workers, which then fade out with age. Indeed, among the birth cohorts for whom high school workers tend to enter into better economies are those from the late 1940s through the early 1950s (Figure 1.6), approximately

[55]The measure of improvement represents the percent change in the root mean squared differences between the predicted and actual wage gaps in the two models of the wage gap between 1998 and 2009.

the same cohorts for whom the within-cohort age profiles in the wage gap are the steepest (Figure 1.7b). For example, high school workers in the 1950 birth cohort face an initial unemployment rates of about 4 percent, one of the lowest over the sample period. The 1950 birth cohort also experiences one of the steepest age profiles between age 26-30 and age 31-40, approximately a 15 percent increase in the wage gap, indicating that the early relative gains for high school educated workers from low unemployment rates fade with age.

1.6.4 Mechanisms and persistence from unemployment in the wage gap

In Table 1.4, I consider specifications designed to distinguish between mechanisms of wage persistence. In the implicit contracting and match quality models, improvements in labor market conditions after labor market entry should reduce the persistent effects from initial conditions. In the human capital accumulation model, the effects of early human capital investment decisions are not erased by later changes in labor market conditions. Column (1) of Table 1.4 presents the baseline specification from column (5) of Table 1.3, but only shows the effect of the unemployment rate faced by high school educated workers at entry on the wage gap at age 26-30.[56]

Columns (2) and (3) introduce the remaining two regressors from equation (1.13),

[56]The effects for older workers in the specifications in Table 1.4 are not generally statistically significant, though I cannot always rule out large effects.

$U^*_{a,k+19}$ and $U^*_{a,k+23}$, the lowest unemployment rates faced between entry and the end of the age cell (e.g. $U^*_{a,k+19}$ gives the minimum unemployment rate faced by a cohort between age 19 and 30 in the 26-30 age cell, and between age 19 and 40 in the 31-40 age cell). In contrast to the predictions of the match quality and implicit contracts models, controlling for improvements in unemployment rates faced by college and high school workers does not reduce the magnitude of the effects of initial unemployment rates. In column (2) the point estimate for the effect of the unemployment rate faced by high school educated workers at entry on the wage gap at age 26-30 actually increases from 1.2 to 1.6 percent.

The specifications in columns (2) and (3) are similar to those in work attempting to distinguish between mechanisms behind persistence in wage levels (see Beaudry and DiNardo (1991) and Schmieder and von Wachter (2010)). That work finds support for downwardly rigid implicit contracts, in contrast to the lack of predictive power for minimum unemployment rates in explaining variation in the wage gap in columns (2) and (3).[57] Given this contrast, further investigation is warranted. Because the dependent variable is the difference between the log wages of college and high school workers, I consider alternative specifications in columns (4) and (5) in-

[57]Beaudry and DiNardo (1991) and Schmieder and von Wachter (2010) observe the history of unemployment rates for job spells of workers, allowing more precise distinctions between mechanisms than is possible with cross-sectional data. This could explain the lack of precision in the estimates of effects of the lowest rates faced on the wage gap.

cluding *differences* between unemployment rates of college and high school educated workers.

Column (4) includes the difference between $U^*_{a,k+19}$ and $U^*_{a,k+23}$, the measures of lowest unemployment rates faced since entry included in column (3). This difference will be negative for cohorts in which the unemployment rate is lower between age 19 and 22 (after high school educated workers enter and before college workers enter) than the rate at age 23 (college entry). A negative value represents a market that is more favorable for high school workers. The difference will be zero when the unemployment rate at the age of college entry is below or equal to the rates faced between age 19 and 22, *or after the unemployment rate falls below the minimum rate faced between age 19 and 22.* Thus the difference term will only pick up effects on the wage gap from unemployment rates in cohorts in which college educated workers have not faced unemployment rates as low as those of high school workers.

The match and implicit contract models predict that the wage gap should be lower in cohorts with negative values of the difference $U^*_{a,k+19} - U^*_{a,k+23}$, since high school educated workers will have had better opportunities to obtain high quality matches, or receive upward wage adjustments. Also, the initial unemployment rate will have no predictive power. In the human capital accumulation model, the initial unemployment rate will have predictive power in so far as it determines investment decisions. The effects of differences between the initial decisions of college and high school educated workers on the wage gap will not be "erased" by later changes in investment opportunities for college or high school educated workers.

Cohorts in which high school educated workers face better opportunities than college workers by age 30 have lower wage gaps. Column (4) shows that a one percentage point lower difference in the lowest unemployment rates faced by high school and college educated workers, indicating better relative opportunities for high school workers between entry and age 30, leads to a 2.8 percent lower wage gap, though the effect is imprecisely estimated. In column (5), instead of the value of the difference, I include just an indicator for better opportunities for high school educated workers.[58] Cohorts with better opportunities for high school workers have 4.5 percent smaller wage gaps on average. In both columns (4) and (5) the effect of the initial unemployment level is reduced.[59]

The estimates in columns (4) and (5) of Table 1.4 are consistent with the implicit contract and match quality models in which wage levels are related to the best labor

[58] Using the indicator variable ignores variation in the magnitude of differences between conditions faced by college and high school workers, but avoids possible mis-specification given the small number of cohorts available to identify the linear relationship.

[59] Including either or both of the unemployment rate at age 23 and the difference between the lowest unemployment rates faced between entry and age 23 for high school and college educated workers (a main effect for differences in initial differences, constant within cohort) does not affect the results in Table 1.4, nor do these variables enter significantly.

market conditions faced rather than initial conditions. However, given the large standard errors in these results, any claims about the mechanisms driving persistence in the wage gap must be tentative.

1.6.5 Robustness of results and endogeneity concerns

In unreported results, I find that the main conclusions from the empirical work are robust to using the finer, seven category age-cohort interactions from the first-stage regression. I also obtain similar results, reported in Appendix C, from specifications which use cohort-age group specific relative supply of college labor to account for differences across cohorts in the timing of educational attainment. In Appendix D, I discuss the robustness of the results to alternative choices for the age of labor market entry for high school educated workers.

One concern with the main results is the role of endogenous supply of college labor. It is standard practice in the literature interested in estimating elasticities of substitution between education levels and age groups to assume supply is exogenous. Endogenous supply is a potentially critical concern for estimating persistence in wages, however, since it could lead to patterns in the wage gap indistinguishable from wage persistence. If early labor market conditions affect workers' decisions to obtain a college education (or work full time), then the effects on the wage gap of past labor market conditions estimated above may not be true persistence, but instead reflect the compositional effects of selection into schooling.[60]

For example, a high school graduate facing high unemployment rates may choose to delay entry into the labor market by enrolling in college. Initial unemployment rates could also induce selection out of the labor market entirely for both college and high school workers. The direction of bias from endogenous supply in the effect of initial unemployment rates on the wage gap is indeterminate. The reason for indeterminacy is two-fold: 1) the bias depends on an unspecified counterfactual college wage distribution of high school workers, and 2) most plausible counterfactual distributions would imply that both the mean wage level of college and high school workers should fall in response to selection into college education, but not which effect will dominate.

To assess the importance of this source of bias, I estimate the effect of initial labor market conditions on a cohort's relative supply of college labor (the sum of selection into education and labor supply). I find that controlling for other determinants of the decision to obtain a college education and work full time, the effects of initial unemployment rates are small, or disappear entirely. Thus, there is little reason to be suspicious of the persistent effects of unemployment on the wage gap identified in

[60]In one view, the selection effects of initial labor market conditions on education and labor market participation are a source of bias, but in another view, this is considered a form of wage persistence. Indeed, Genda et al. (2010) and Hershbein (2010) are explicitly interested in the effects of initial conditions on labor supply.

this paper.

The dependent variable in Table 1.5 is the cohort-specific relative supply (regression adjusted to age 35-40 levels), and the regressions include one observation per birth cohort. Column (1) includes just the unemployment rate at age 19. In this specification, I find that a one percentage point increase in the unemployment rate is associated with a 7.8 percent increase in relative supply, consistent with a large role for endogenous supply, and troubling for the wage persistence interpretation of the above results. However, with controls for other determinants of a cohort's relative supply, including a linear trend, cohort size, and the aggregate supply of college and high school labor faced at age 19, I can rule out effects larger than 2.8 percent in absolute value with 95 percent confidence.

A 7.8 percent effect on relative supply is equivalent to a decrease in the mean level over the sample period of the percent of the labor force composed of effective high school workers from 63.5 to 61.7 percent, or a 3 percent change. Ignoring effects of compositional changes from heterogeneity in earnings ability, which are of indeterminate sign, the estimate in column (1) could be responsible for a 2.5 percent decrease in the wage gap.[61] This would suggest that the estimated effect of initial

[61] To see this, use the estimate of the effect of cohort-specific relative supply on the wage gap from column (5) of Table 1.3, and the change in supply from column (1) of Table 1.5: $-0.324 * 0.078 = -0.025$.

unemployment rates on the wage gap in Tables 1.3 and 1.4 are biased downward (though, compositional changes complicate this interpretation). However, the results in columns (2)–(5) suggest that the bias is much smaller. Indeed, the vast majority of variation in the cohort-specific relative supply of college labor is accounted for by a linear cohort trend and cohort size – the specification in column (3) accounts for 95 percent of the variation in the cohort-specific relative supply, while the unemployment rate alone only accounts for just 12 percent.

It is difficult to assess the magnitude or even direction of bias from selection because of the indeterminacy of effects from compositional changes on the wage gap. I find evidence that the effects of unemployment on the fraction of workers obtaining a college education, conditional on working full-time, are small, especially once I control for aggregate levels of college and high school workers faced by workers at age 19. Endogenous supply is unlikely to be driving persistence in the wage gap from unemployment rates.

1.6.6 Initial aggregate supply and persistence in the wage gap from unemployment

The aggregate supplies of college and high school educated labor are determinants of the marginal product of college and high school educated workers in equations (1.4) and (1.5). Thus, in addition to unemployment rates, the aggregate supplies faced by workers at the age of labor market entry could lead to persistent effects in the wage gap. However, given the evidence in Table 1.5 of large effects of the aggregate supply faced at entry and the cohort-specific relative supply, there is reason to be skeptical of the persistence interpretation.[62]

However, as key determinants of the relative supply of college labor in a cohort, controlling for initial levels of aggregate supply of college and high school labor in specifications with initial unemployment rates will speak to the concerns of bias from endogenous supply discussed in the previous section. Since aggregate supply and unemployment are directly related (though the relationship between education-group specific supply and unemployment is ambiguous), excluding initial levels of aggregate supply from the model is a particular concern for estimating the effect of initial unemployment rates on the wage gap.

Table 1.6 gives estimates of the persistent effects of unemployment on the wage gap with controls for aggregate supply of college and high school workers at entry.[63] The effect of initial unemployment is robust to controls for the aggregate supply of college and high schools workers at the age of entry. Also, unlike the unemployment

[62]The aggregate supply measures include all workers in the labor market age 20-64 when the birth cohort is 19 years old, so there is no mechanical relationship between the variables.

[63]The sample in Table 1.5 is further limited from that in earlier tables as a result of unobserved aggregate supply at entry for cohorts born prior to 1945.

results, the relationship between initial aggregate supply and the wage gap is constant with age, consistent with an endogenous response of cohort-specific relative supply.

Column (1) shows the baseline specification from column (5) of Table 1.3 with the new sample limitation. The effect of a one percentage point increase in the unemployment rate at age 19 on the wage gap age age 26-30 increases slightly, from 1.2 to 1.5 percent, and the difference is not statistically significant. Column (2) includes the aggregate relative supply of college to high school labor at age 19 interacted with age group. Though the aggregate supply does not enter the model significantly, and has little effect on the persistent effect of initial unemployment rates, the linear cohort trend is no longer significantly different from zero. This suggests that the trend was picking up most of the effects of initial aggregate relative supply on the wage gap and also leaves little role for persistence from initial levels of relative demand.[64]

Column (3) includes the log aggregate supply of high school workers at age 19 separately from the relative supply since only the high school supply enters the high school marginal product. A ten percent increase in the supply of high school labor faced at age 19 is associated with 2-5 percent higher wage gaps through age 50. The unemployment effects are not significantly affected, and if anything, are larger after

[64]However, changes in the permanent characteristics of birth cohorts, including education quality, could work against persistence from relative demand.

controlling for initial aggregate supply. Including the log aggregate supply of college labor at age 23 in column (4) does not affect the results.

The positive effects on the wage gap from initial aggregate supply levels are consistent with endogeneity in the decision of workers to obtain higher education or enter the labor market, especially given that the effect does not fade with age. Persistent effects from initial unemployment are robust to controlling for this source of endogenous supply, which is directly tied to unemployment levels. This is strong evidence that the main results in this paper represent actual persistence in relative wage levels and not compositional effects from the decisions of workers to obtain education or work full time.

1.7 Conclusion

Using a flexible set of restrictions on time, age, and cohort effects, I identify changes across cohorts in wage gap age profiles. Cohorts born in the late 1940s and 1950s have lower cohort-specific wage gaps when young, but see steeper increases in the wage gap with age. Since these cohorts are among the older workers in the 1990s and 2000s, the fade-out of initial negative effects on the wage gap with age is responsible for some of the recent unexplained increases in the wage gap for workers over 45.

An important component of wage inequality is generated by past labor market experiences. I find robust support for a persistent effect of labor market conditions at the age of labor market entry on the college-high school wage gap. The estimated effects are large enough to account for over a third of the unexplained increase in the wage gap for older workers starting in the late 1990s. Indeed, the cohorts with steeper within-cohort age profiles, those born in the 1940s and 1950s, are also the cohorts with lower initial unemployment rates.

A lower unemployment rate at age 19 leads to lower college-high school wage gaps in the birth cohort for young workers. A one percentage point higher unemployment rate causes a one to two percent increase in the wage gap, persistent through age 30. Also, accounting for wage persistence reduces the effects of contemporaneous labor market conditions on the wage gap, including aggregate relative supply and relative demand for college educated labor.

A more tentative result suggests that persistence is driven by downwardly rigid implicit contracts or match quality, and not human capital accumulation. The wage gap is smaller in cohorts in which high school educated workers face better labor market conditions than college educated workers, and initial unemployment has less predictive power. Human capital models predict that effects of initial conditions will persist even after market conditions improve. Additional work is required to identify the mechanisms responsible for persistence in the evolution of the wage gap.

An important role for persistent effects of past labor market conditions in education wage differentials suggests that wage inequality is in part driven by the luck of a cohort to be born so that workers will enter the market during beneficial labor market conditions. Though the effects fade out with age, high school workers entering the labor market to face poor labor market conditions have lower life-time relative earnings. The on-the-job search and match quality explanation for wage persistence implies market inefficiencies are responsible for generating inequality. Persistence from implicit contracts implies workers are not paid their true marginal products. Though the insurance mechanism behind the contracts is efficient, the nature of the insurance is to smooth wages within cohorts, not between them. To alleviate the effects of persistent wage inequality, policy makers should consider targeted cross-generational transfers over transfers designed to alleviate only the inequality generating effects of transitory labor market conditions.

Tables

Table 1.1: Comparison of Alternative First Stage Models

	(1)	(2)	(3)
Number of years	46	46	46
Number of age cells	7	4	7
Number of cohorts	60	60	60
Number of cohort-age group cells	–	169	273
Number of age-time cells	1365	1365	1365
R-squared	0.648	0.688	0.738
Model p-value vs. column (3)	0.0000	0.0000	–
Model p-value vs. unrestricted	0.0530	0.2200	0.6949

Notes: Each column shows summary statistics from a first stage decomposition of the age-time specific log college-high school wage gap into age, year and cohort or cohort-age group fixed effects. Wage gaps are estimated in a series of micro-data regressions of the education-specific wage on an indicator for college and an indicator for non-white, separately by age-time cell. The model p-value vs. column (3) is the result of an F-test of the restrictions relative to those in column (3) (the specifications are nested). The model p-value versus the unrestricted model is the result of a chi-squared test of the column restrictions based on the wage gap standard errors estimated from the micro data. The latter tests are not correctly sized in the presence of group-specific error components (see Dickens (1990)).

Table 1.2: Single- and Two-Stage Estimates of Models of the College-High School Wage Gap

	Single Stage			Two-Stage		
	(1)	(2)	(3)	(4)	(5)	(6)
Year	0.022**	0.019**		0.018**	0.015**	0.010**
	(0.001)	(0.002)		(0.001)	(0.001)	(0.001)
Aggregate relative supply	-0.287**	-0.285**		-0.273**	-0.186**	0.001
	(0.036)	(0.036)		(0.036)	(0.035)	(0.028)
Cohort		0.003*	0.003	0.002**	0.004**	0.004**
		(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Cohort-specific relative supply	-0.296**	-0.296**	-0.284**	-0.251**	-0.293**	-0.323**
	(0.018)	(0.019)	(0.017)	(0.024)	(0.021)	(0.020)
Cohort x age interaction?	–	–	–	No	Yes	Yes
Number of age categories?	7	7	7	7	4	7
Year fixed effects?	No	No	Yes	–	–	–
Elasticity of substitution b/w ed.	1.7	1.7	–	1.9	2.1	3.1
Elasticity of substitution b/w age	3.4	3.4	3.5	4.0	3.4	3.1
	(0.20)	(0.21)	(0.22)	(0.38)	(0.24)	(0.19)
Number of observations in first stage	1365	1365	1365	1365	1365	1365
Observartions in cohort effects model	–	–	–	60	169	273
R-squared	0.576	0.563	0.616	0.910	0.790	0.695

Notes: Columns (1), (2) and (3) show separate OLS regressions with standard errors robust to clustering at the cohort level in parentheses. Columns (4)–(6) each show the result of a pair of second stage OLS regressions with the dependent variable estimated from a first stage regression of the age-cohort specific wage gaps on year fixed effects and the specified cohort and age group fixed effects. Standard errors robust to heteroskedasticity for column (3) and clustering at the cohort level for columns (4) and (5) in parentheses. Standard errors on elasticities calculated using the delta method. The reported R-squareds for columns (4)–(6) are from the model of cohort or cohort x age group fixed effects. All models include age group fixed effects at the level indicated. * significant at 10%; ** significant at 5%.

Table 1.3: Persistent Effects of Initial Unemployment Rates on the College-High School Wage Gap

	Single-Stage			Two-Stage		
	(1)	(2)	(3)	(4)	(5)	(6)
Cohort	0.001	0.001	-0.001	0.005**	0.004**	0.004**
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Cohort-specific relative supply	-0.321**	-0.309**	-0.261**	-0.345**	-0.316**	-0.315**
	(0.026)	(0.030)	(0.025)	(0.027)	(0.031)	(0.030)
Age 26-30 interactions:						
Unemployment rate at age 19		0.006	0.013**		0.012**	0.013**
		(0.004)	(0.004)		(0.004)	(0.004)
Unemployment rate at age 23						-0.000
						(0.003)
Age 31-40 interactions:						
Unemployment rate at age 19		0.002	0.004**		0.003	0.003
		(0.003)	(0.002)		(0.003)	(0.004)
Unemployment rate at age 23						0.001
						(0.004)
Age 41-50 interactions:						
Unemployment rate at age 19		-0.001	-0.002		0.001	-0.002
		(0.004)	(0.003)		(0.004)	(0.004)
Unemployment rate at age 23						0.009**
						(0.004)
Year fixed effects or controls?	Controls	Controls	FE	–	–	–
Elasticity of substitution b/w age	3.1	3.2	3.8	2.9	3.2	3.2
	(0.25)	(0.32)	(0.37)	(0.23)	(0.31)	(0.30)
Number of observations	960	960	960	124	124	124
R-squared	0.701	0.703	0.743	0.785	0.800	0.807

Notes: Each column is an OLS regression. Standard errors robust to clustering at the cohort level in parentheses. Standard errors on elasticities calculated using the delta method. All models include age group fixed effects. Sample is limited to age 26-50. See text. * significant at 10%; ** significant at 5%.

Table 1.4: Mechanisms and Persistent Effects of Initial Unemployment Rates on the Wage Gap

	Second Stage Estimates				
	(1)	(2)	(3)	(4)	(5)
Cohort	0.004**	0.004**	0.005**	0.004**	0.004**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Cohort-specific relative supply	-0.316**	-0.329**	-0.348**	-0.309**	-0.309**
	(0.031)	(0.030)	(0.041)	(0.037)	(0.031)
Age 26-30 interactions:					
Unemployment rate at age 19	0.012**	0.016**	0.011**	0.007*	0.005
	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)
Lowest unemployment rate faced age 19-30 (H)		-0.013	0.009		
		(0.008)	(0.020)		
Lowest unemployment rate faced age 23-30 (C)			-0.018		
			(0.016)		
Difference in lowest unemployment by age 30 (H minus C)				0.028*	
				(0.015)	
Lowest unemployment by age 30 is lower for H than C					-0.045**
					(0.012)
Age 31-40 and 41-50 interactions included but not reported.					
Elasticity of substitution b/w age	3.2	3.0	2.9	3.2	3.2
	(0.31)	(0.28)	(0.34)	(0.39)	(0.33)
Number of observations	124	124	124	124	124
R-squared	0.800	0.816	0.827	0.810	0.822

Notes: Each column is an OLS regression. Standard errors robust to clustering at the cohort level in parentheses. Standard errors on elasticities calculated using the delta method. All models include age group fixed effects. Sample is limited to age 26-50. See text. * significant at 10%; ** significant at 5%.

Table 1.5: Do Entry Conditions Predict Cohort-Specific Supply?

Dependent Variable:	Cohort-Specific Relative Supply				
	(1)	(2)	(3)	(4)	(5)
Cohort		0.023**	0.018**	0.015**	0.023**
		(0.001)	(0.001)	(0.001)	(0.009)
Unemployment rate at age 19	0.078**	-0.023*	-0.026**	-0.031**	0.003
	(0.025)	(0.012)	(0.009)	(0.007)	(0.013)
Aggregate relative supply of labor at age 19					-0.561**
					(0.207)
Log aggregate supply of high school labor at age 19					0.568
					(0.353)
Log cohort size			31.371**	33.335**	24.400**
			(7.694)	(10.524)	(9.672)
Log cohort size squared			-0.943**	-1.007**	-0.735**
			(0.233)	(0.317)	(0.292)
Number of observations	51	51	51	35	35
R-squared	0.124	0.890	0.951	0.896	0.921

Notes: Each column is an OLS regression. Standard errors robust to heteroskedasticity in parentheses. Column (4) presents results of the specification in column (3) with the sample limited to cohorts for which aggregate supply of labor at age 19 is available. * significant at 10%; ** significant at 5%.

Table 1.6: Robustness of Wage Gap Persistence from Unemployment to Endogenous Supply

	Second Stage Estimates			
	(1)	(2)	(3)	(4)
Cohort	0.004**	0.000	-0.003	-0.004
	(0.001)	(0.003)	(0.004)	(0.004)
Cohort-specific relative supply	-0.262**	-0.225**	-0.277**	-0.268**
	(0.054)	(0.061)	(0.065)	(0.066)
Age 26-30 interactions:				
Unemployment at age 19	0.015**	0.013**	0.017**	0.017**
	(0.005)	(0.006)	(0.007)	(0.007)
Aggregate relative supply of labor at age 19		0.086	0.056	0.072
		(0.081)	(0.095)	(0.131)
Log aggregate supply of high school labor at age 19			0.281*	0.297
			(0.162)	(0.209)
Log aggregate supply of college labor at age 23				0.005
				(0.143)
Age 31-40 interactions:				
Unemployment at age 19	0.007*	0.001	0.004	0.001
	(0.004)	(0.005)	(0.006)	(0.007)
Aggregate relative supply of labor at age 19		0.124	0.118	-0.018
		(0.079)	(0.092)	(0.114)
Log aggregate supply of high school labor at age 19			0.222	0.061
			(0.164)	(0.209)
Log aggregate supply of college labor at age 23				0.190
				(0.135)
Age 41-50 interactions:				
Unemployment at age 19	-0.002	-0.005	0.008	0.008
	(0.005)	(0.007)	(0.009)	(0.009)
Aggregate relative supply of labor at age 19		0.092	-0.048	-0.112
		(0.074)	(0.108)	(0.278)
Log aggregate supply of high school labor at age 19			0.496**	0.423
			(0.191)	(0.368)
Log aggregate supply of college labor at age 23				0.094
				(0.313)
Elasticity of substitution b/w age	3.8	4.4	3.6	3.7
	(0.78)	(1.21)	(0.84)	(0.92)
Number of observations	85	85	85	85
R-squared	0.766	0.777	0.796	0.798

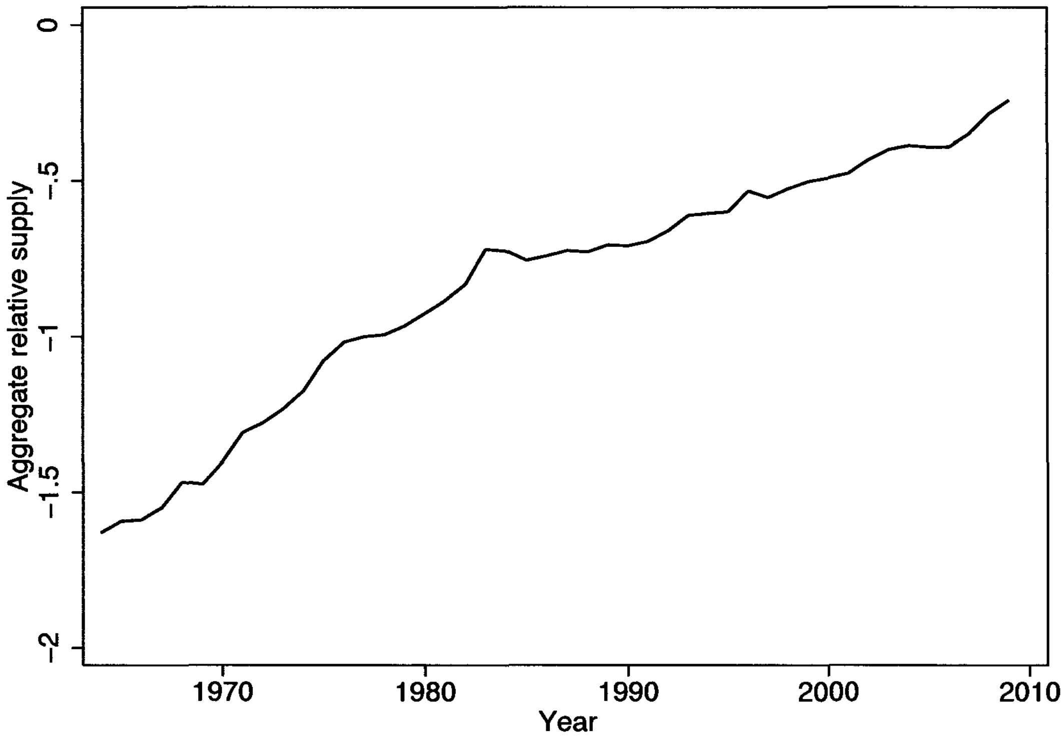
Notes: Each column is an OLS regression. Standard errors robust to clustering at the cohort level in parentheses. Standard errors on elasticities calculated using the delta method. All models include age group fixed effects. * significant at 10%; ** significant at 5%.

Figures

Figure 1.1: Cross-Sectional Trend in the Log College-High School Wage Gap

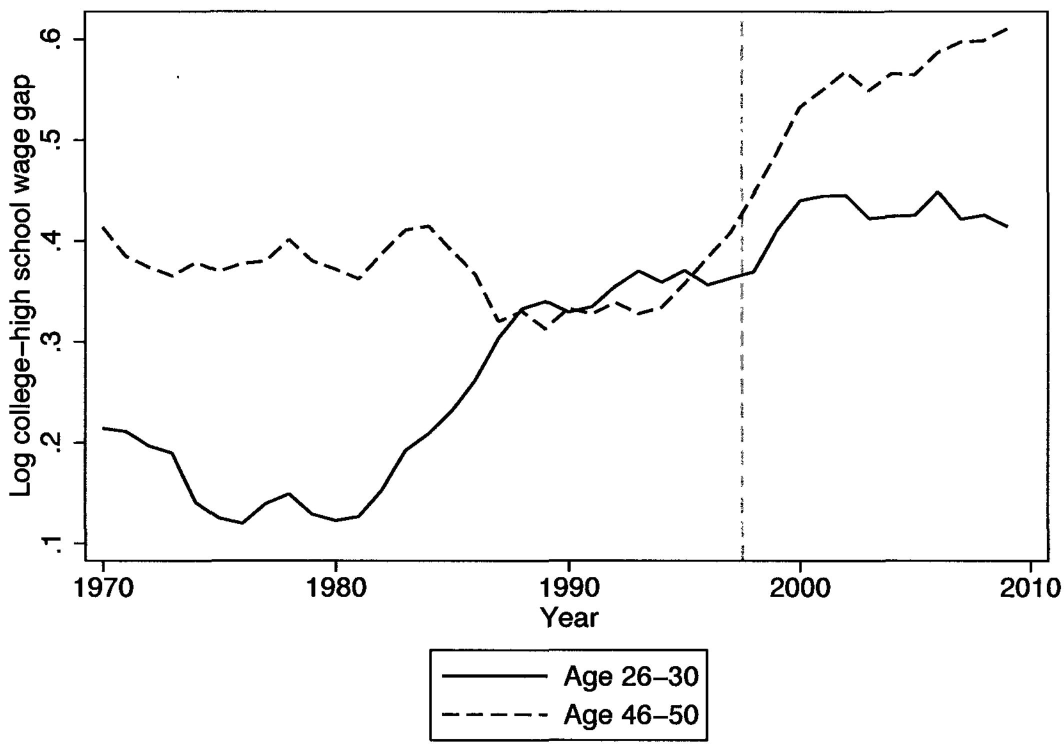
Notes: The sample includes full-time, full-year employed men age 26-60 in the CPS March Supplements between 1964 and 2009. Wage gaps are calculated using workers with exactly 16 years of education versus those with exactly 12 years of education. I discuss the sample and construction of the wage gap in Section 1.5. See text. The figure shows three-year moving averages.

Figure 1.2: Aggregate Relative Supply of College Labor Faced by Cohort



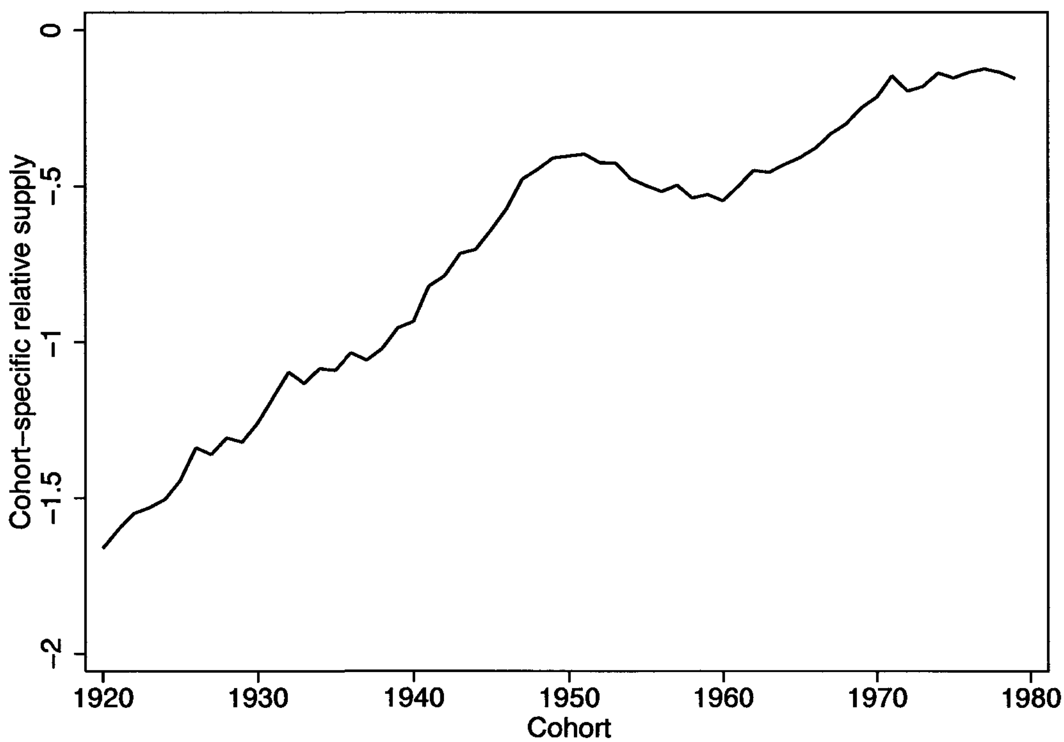
Notes: The sample includes full-time, full-year employed men and women age 20-64 in the CPS March Supplements between 1964 and 2009. The aggregate relative supply of college labor is the log ratio of the supply of effective college labor to the supply of effective high school labor in a year of the March CPS. I construct effective supply measures by estimating weights for education groups, described in Section 1.5. See text.

Figure 1.3: Trends in the Log College-High School Wage Gap by Age



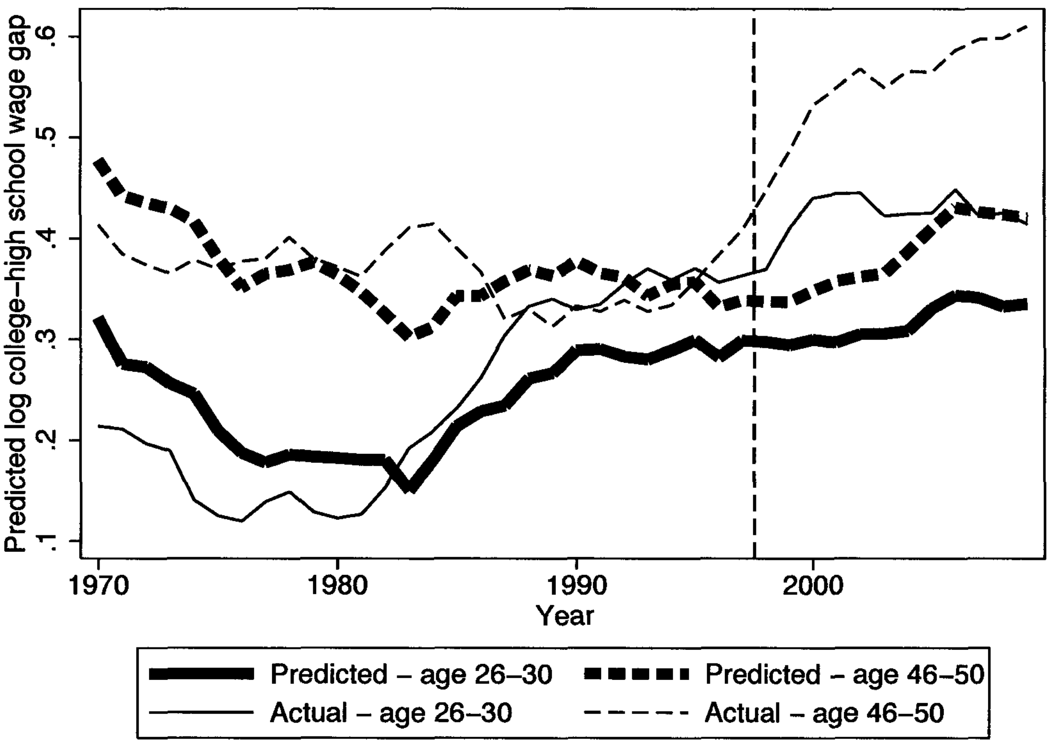
Notes: The sample includes full-time, full-year employed men age 26-60 in the CPS March Supplements between 1970 and 2009. Wage gaps are calculated using workers with exactly 16 years of education versus those with exactly 12 years of education. See text. The vertical line represents 1997, the last year used in the Card and Lemieux (2001) analysis of the wage gap. The figure shows three-year moving averages.

Figure 1.4: The Cohort-Specific Relative Supply of College Labor



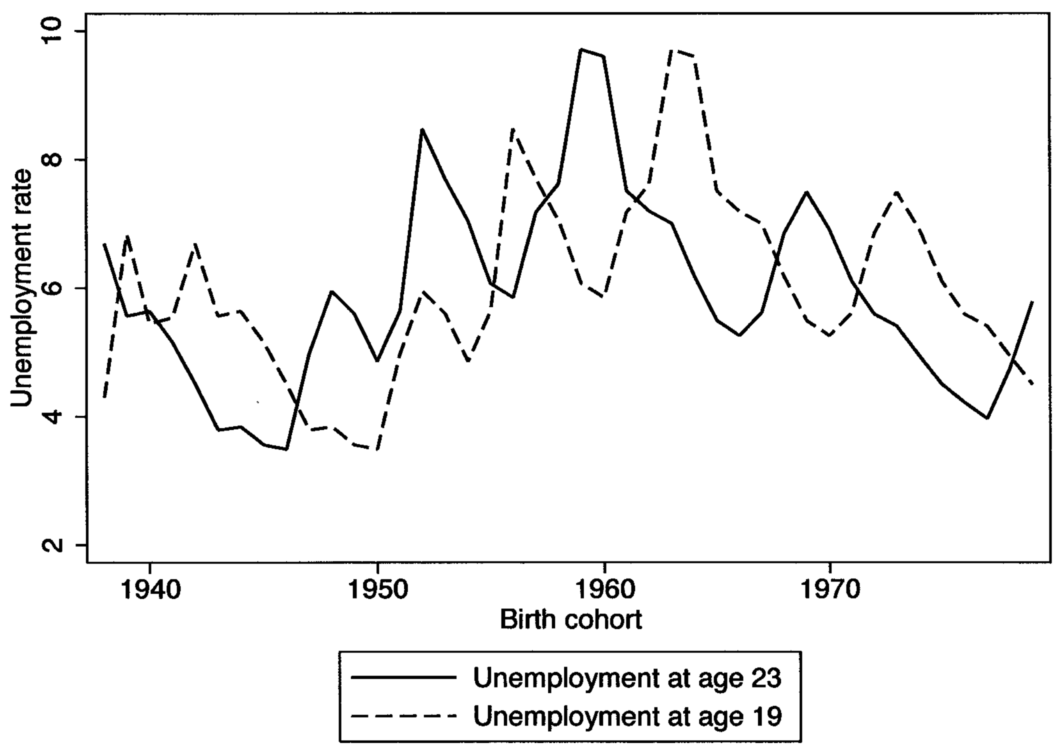
Notes: The sample includes all full-time, full-year employed men and women age 20-64 in the CPS March Supplements between 1964 and 2009. The cohort-specific relative supply of college labor is the log ratio of the supply of effective college labor to the supply of effective high school labor in a birth cohort, normalized to age 35-40 in a regression of the cohort-age group specific relative supply on a full set of cohort and age group fixed effects (see text).

Figure 1.5: Predicted Trends in the College-High School Wage Gap by Age



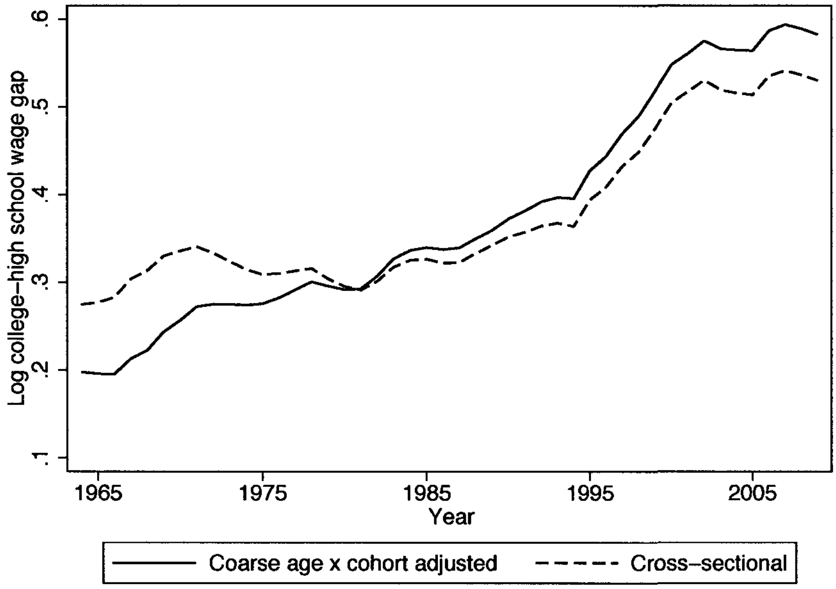
Notes: Predictions are based on estimates in Card and Lemieux (2001) and include a 0.017 percent annual increase in the wage gap, an elasticity of substitution between age groups of 5, and an elasticity of substitution between college and high school educated workers of 1.6. The figure shows three-year moving averages for the actual wage gaps. Predicted wage gaps are normalized by age group to match the 1970-1997 actual mean gap.

Figure 1.6: Unemployment Rates Faced at Age 19 and 23 by Birth Cohort

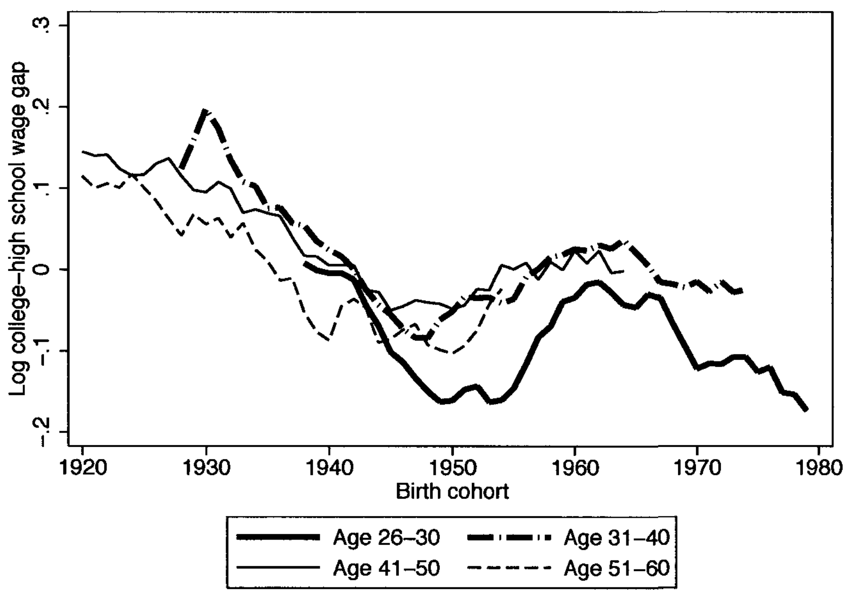


Notes: Source: BLS.

Figure 1.7: Time and Birth Cohort-Age Group Decomposition of the College-High School Wage Gap



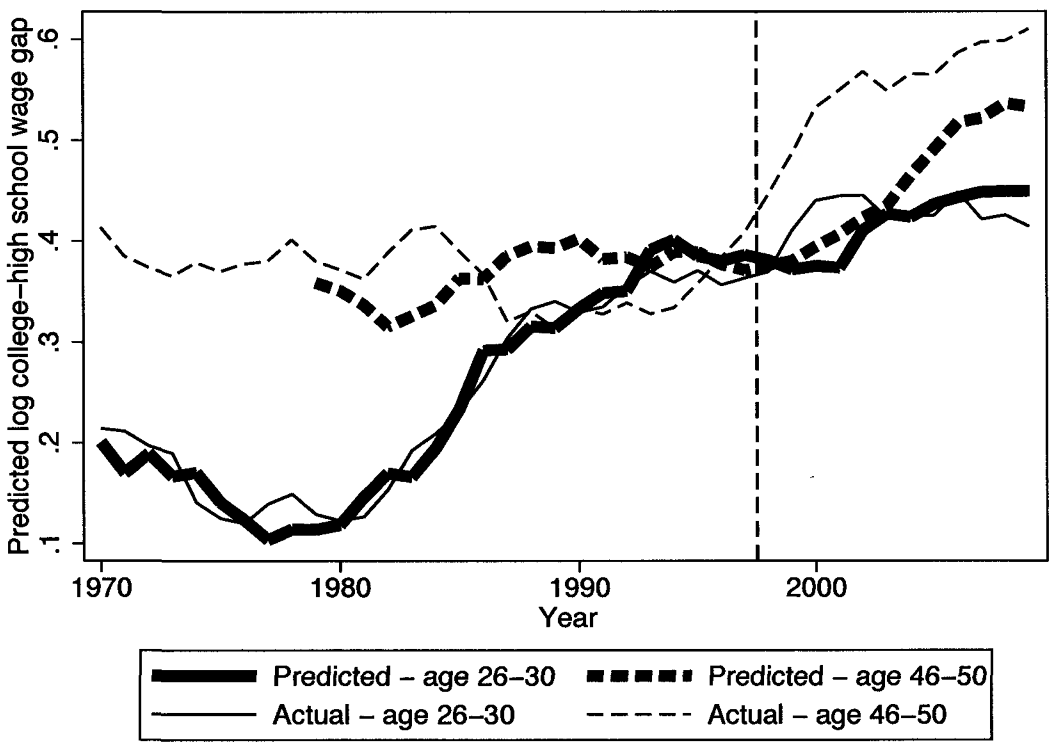
(a) First stage time fixed effects



(b) First stage cohort-age group fixed effects

Notes: The decomposition is obtained from the regression of cohort-time specific log wage gaps on a full set of time and cohort x age group fixed effects in equation (16) in the text. Age groups are 26-30, 31-40,41-50, and 51-60. The figures show three-year moving averages. The time effects are re-meaned to match the mean cross-sectional wage gap. The cohort-age group effects are de-meaned.

Figure 1.8: Actual and Predicted College-High School Wage Gap Including Persistent Effects from Unemployment



Notes: See text. The figure shows three-year moving averages. Predicted wage gaps are normalized by age group to match the 1970-1997 actual mean gap. Predictions include the effect of an estimated 0.004 percent linear increase in the wage gap per birth cohort.

CHAPTER II

What's in a Rating?

2.1 Introduction

Recent education reforms including the federal Race to the Top (RTT) initiative focus on improving teacher quality by instituting evaluation and compensation systems that are based on student achievement. Such evaluation systems often rely in large part on value-added measures of teacher effectiveness. Concerns with the validity and reliability of value-added has led many schools and districts to adopt teacher evaluation systems that rely heavily on supervisor ratings. Indeed, some of the well-known popular reforms in this arena such as the Teacher Advancement Program (TAP) and Denver ProComp place a significant emphasis on supervisor assessment.

The emphasis on subjective ratings, however, has raised a variety of concerns. Some observers worry that principal assessments will dilute inappropriately the focus on student outcomes. Other observers fear that placing greater weight on subjective evaluations will encourage favoritism, nepotism and even discrimination. Another set of concerns centers around the fact that formal performance ratings are almost universally high[1] and thus cannot reasonably be thought to identify teaching quality.

Several recent studies have examined principal attitudes and behaviors with regard to teacher assessment in order to assess the potential importance of such concerns. The existing evidence is limited, but does suggest that principals can differentiate between more and less effective teachers. Moreover, there is evidence that principals use this information systematically in high-stakes hiring and dismissal decisions. Jacob and Lefgren (2008) examined the relationship between subjective (principal holistic assessments) and objective (i.e., value-added) measures of teacher effectiveness, finding that principals could identify the most and least effective teachers but have less ability to differentiate between teachers in the middle of the distribution.[2] Kane et al. (2010) studied the relationship between a teachers effectiveness as measured by value-added and the ratings the teacher received on the basis of classroom observations by trained supervisors. They find that classroom-based observational measures of teacher effectiveness are strongly correlated with student achievement growth. Jacob (2010a) and Jacob (2010b) found that principals in Chicago who were given more

[1]In their report "The Widget Effect" The New Teacher Project (2009) found that in the districts they study over 94 percent of teachers receive positive performance ratings.

[2]Also see Harris et al. (2006).

flexibility to dismiss probationary teachers were more likely to dismiss those teachers with greater absences, weaker educational background and lower value-added scores. Grossman et al. (2010) find classroom-based observational measures of certain instructional practices are highly correlated with teacher effectiveness as measured by value-added.

In this paper, we extend this literature by examining the relationship between the formal ratings that principals give teachers and a variety of observable teacher characteristics, including proxies for productivity such as teacher absences, experience and educational background. To do so, we combine personnel records on all classroom teachers in the Chicago Public Schools (CPS) between 2002-2003 and 2005-2006 with data on the ratings they received from their principal in each applicable year. The benefit of this approach relative to prior work is that it examines the formal evaluations that principals currently give teachers, whereas prior work has relied on surveys in which principals might not evaluate teachers as they would in the institutional context of the formal ratings.[3] By examining the formal ratings we can assess how principals evaluate teachers in a setting where the stakes are reasonably high. Understanding how principals use the performance evaluation tools currently available

[3]The formal ratings are supposed to be based on a principals review of teacher lesson plans, observation of the teacher's classroom and monitoring of student progress.

to them will help policy makers design evaluation and compensation procedures that place appropriate weight on subjective evaluations.

We find that ratings are positively associated with teacher experience for younger teachers, but that there is no relationship between experience and ratings for teachers with more than 10 years of experience. We also find that ratings are negatively associated with teacher absences and positively associated with some measures of educational background. Conditional on observables, men and minority teachers receive lower ratings. Also, conditional on principal, teacher, and student observables, principals in higher performing elementary schools give higher ratings.

These patterns are largely consistent with findings from the teacher effectiveness literature. For example, a large body of prior research has found that early teacher experience is positively associated with student achievement growth (Rockoff, 2004; Rockoff et al., 2008). Similarly, there is evidence that teacher absences and (in some cases) the quality of a teachers college (often viewed as a proxy for the teachers own cognitive ability) are positively associated with student performance (Rockoff et al., 2008). In short, our findings suggest that these formal ratings are strongly related to proxies for productivity, which provides reason to be optimistic about policies that would assign more weight to principal evaluations of teachers.

The remainder of the paper proceeds as follows. Section 2.2 presents background material, and Sections 2.3 and 2.4 describe the data and empirical strategy, respectively. Section 2.5 presents the results, and Section 2.6 concludes.

2.2 Background

The collective bargaining agreement between the Chicago Teachers Union (CTU) and the CPS specifies which teachers are to be evaluated based on probationary or tenured status and prior ratings received. The efficiency ratings procedures relevant to this study went into effect for the 2000-2001 school-year, and were not significantly altered until the 2007-2008 school-year, after the time-frame of our analysis. The agreement specified that tenured teachers previously rated excellent or superior were to be rated every two years, while tenured teachers rated satisfactory were to be rated annually. An exception to these rules is that principals newly assigned to a school were not to rate tenured teachers with prior satisfactory or better ratings until having served in the school for at least 5 months. The only guidance given for the rating of probationary teachers was that newly assigned principals should rate all probationary teachers. The agreement does not specify how frequently returning principals should rate probationary teachers.

While principal ratings of teachers were not a significant factor in teacher dismissals during our period of analysis, the ratings were a factor in some personnel decisions. Administrators make use of the ratings in assigning teachers to summer school positions, and in some school re-assignment decisions. Also, teachers given unsatisfactory ratings are required to participate in drafting a remediation plan with their principal. As we show below, the vast majority of ratings given are positive, and as a result it is teachers at the low end of the quality distribution who are most affected by performance ratings.

The CPS ratings system is similar to those used in many large school districts. Principals in CPS assign teachers a rating in one of four categories, "superior," "excellent," "satisfactory," and "unsatisfactory." The New Teacher Project (2009) profiled rating systems in several districts including the CPS. While districts vary in the number of categories of ratings given from two to five, in all districts, including the CPS, the vast majority of ratings are given in the top one or two categories. The systems are also similar in the frequency of required ratings and the evaluation process.

2.3 Data

The data for this study come from the Chicago Public Schools (CPS) administrative data. Teacher personnel files provide information on teacher background, assignment and efficiency ratings. We supplement this teacher-level data with information on school demographics and principal characteristics from personnel files. We obtain teacher absences from payroll records. Details on how absences are cleaned can be found in Appendix E.

The sample includes all teachers in the Chicago Public School system in the 2002-2003 through 2005-2006 academic years. We exclude individuals who were (1) employed by the central office (including speech pathologists, nurses, counselors) and teachers working exclusively in administrative or professional development capacities; (2) working in a handful of "alternative" schools that serve severely disabled students or other special populations; (3) working less than half-time; or (4) assigned to their school under temporary positions. For teachers in multiple positions (e.g., art or music teachers working in several different schools), we keep only the observation that the school district has designated as the teachers "primary" position.

Our final sample consists of 27,886 unique elementary and high school teachers. For some analysis we limit the sample further to only the 14,318 teachers who are observed in all four years. This four-year sample has two key benefits for estimation relative to the full sample of teachers. First, we will show that some concerns with selection into being rated will be less important for the four-year sample than for the full sample. Second, a balanced panel of teachers over the four years will be important for identifying the relationship between time-varying teacher characteristics and ratings using within teacher variation. However, there are also some concerns with focusing only on the four-year sample. Because teachers must be present in all four years, any teacher whose first year of teaching is after 2002 will be dropped from the sample. As a result, the four-year sample will contain older and more experienced teachers on average. For this reason, we present results for both samples, which are virtually identical in most cases.

2.3.1 Descriptive statistics on principal ratings

While many teachers do receive ratings during our sample period, there seems to be considerable discretion exercised by principals in determining whether to rate a particular teacher in a given year. Roughly 62% of teachers in the main sample received at least one rating from their principal over the four-year period of the analysis. In the four-year sample, 82% of the teachers are rated at least once. About a third of teachers in the full sample, and just over half of teachers in the four-year sample are rated multiple times over the period.

We find considerable variation in the prevalence of ratings across years, as shown in Figure 2.1. Figure 2.1a shows a histogram of the fraction of teachers in a school rated in the odd years of our sample period, 2002-2003 and 2004-2005. Figure 2.1b shows an analogous histogram for the even years, 2003-2004, 2005-2006. A much larger fraction of teachers were rated in the even years. The guidelines in the collective bargaining agreement go some way to explaining this pattern. The vast majority of teachers are given excellent or superior ratings, and so we should expect to see most teachers rated only every other year.[4] Figures 2.1c and 2.1d show histograms of the mean rating in a school separately for the odd and even years respectively. Unlike the histograms in the top row, these figures are qualitatively similar, suggesting that though principals rate fewer teachers in the odd years, the principals are not rating teachers very differently in these years.

The frequency and timing of principals ratings of teachers do not strictly adhere to the requirements in the bargaining agreement. Given the guidelines specified in the collective bargaining agreement, we can predict rating status for teachers in the 2003-2004 through 2006-2007 school years based on their status as probationary or tenured teachers and prior rating (we cannot predict rating status for 2002-2003 because we do not observe the teacher's rating in 2000-2001, two years prior). Roughly a third of tenured teachers fall into each of three categories – those that should be rated, those

[4]Based on the rules in the collective bargaining agreement, we predict that roughly half of tenured teachers in 2004, and only 3% of tenured teachers in 2005 should be rated. Thus it appears that principals tend to default to rating the majority of teachers in the even years when they are required to rate a large fraction of their teachers. Probationary teachers are also much more likely to be rated in the even years with 41 percent rated in 2004, 14% rated in 2005, and 56% in 2006.

that should not, and those for whom the contract does not specify status (such as those who have no prior ratings, and a small number rated unsatisfactory). Among those tenured teachers who should be rated, 74% actually are rated. Among those tenured teachers who should not be rated, 27% actually are rated. 23% of the teachers in the unspecified group are given ratings. This suggests that principals follow the contract in the majority of cases, but that the rules not universally followed or enforced.[5] The apparent discretion of principals in deciding when to rate teachers raises an important concern about sample selection. We discuss this issue in the following section.

Table 2.1 presents summary statistics of teachers, principals and schools for the full and four-year samples. The two samples have roughly the same fraction of teachers by gender and race, but, as expected, the four-year sample is older on average by about 1.5 years. This is a result of the selection of first year teachers after 2003 out of the four-year sample: 5.4% of teacher-years in the full sample are first year teachers compared to 1.3% of teacher-years in the four-year sample. The educational backgrounds of teachers in the two samples are similar. Teachers in both of the samples tend to come from lower selectivity undergraduate institutions on the Barron's scale. 59% of teacher-years in the full sample have higher degrees compared to 61%

[5]Given the uncertainty about the rules governing rating of probationary teachers, we don't perform a parallel comparison for probationary teachers.

in the four-year sample. Teachers are absent an average of 9.3 days per year in the full sample and 8.8 days in the four-year sample.

The full sample represents 602 schools of which 600 are also in the four-year sample. We dont report school characteristics in Table 2.1. 29% of teachers work in mixed race or integrated schools as defined by having over 15% non-Black and non-Hispanic students, the mean enrollment in a school is 1,075 students, and the mean fraction of students testing at or above proficiency in a school is 44%. 66% of teacher-years in the full sample who are given a rating are rated "superior," the highest rating possible, and 28% are rated "excellent". Only 6% of teachers are given ratings in the lower two categories "satisfactory" and "unsatisfactory," and thus for all analysis we combine these two ratings into a single low rating category. As the distribution of ratings is heavily skewed to the top end of the scale, we expect that the ratings are likely to distinguish between those that principals view as mid- and low- quality teachers rather than teachers at the high end of the quality distribution.

Alternative ratings scales or incentive structures could lead principals to rate teachers differently and identify variation in teacher quality at other margins. Given that we only observe ratings in a single institutional setting, we will be limited in what we can say about ratings in alternative environments. In particular we can say nothing about how principals might assign ratings to teachers in the upper tail of the effectiveness distribution. By separately considering both the upper and lower margins that we do observe, we can determine whether principals place different weights on teacher qualities at these margins. However, whether principals are able or willing to distinguish between teachers in finer categories of effectiveness is beyond the scope of this paper.

2.4 Empirical Strategy

The goal of this analysis is to determine the relationship between teacher characteristics and principal ratings of the teacher. We estimate a variety of simple regression models in which the dependent variable is the principal rating of teacher i in school j in year t. The explanatory variables in these models include time-constant teacher characteristics X (such as education credentials and demographics), and time-varying teacher characteristics W (such as years of teaching experience and absences). To examine the influence of teacher characteristics X and W on ratings, we estimate models with school x year fixed effects (α_{jt}) to account for unobserved school-level factors that might be correlated with teacher characteristics and principal ratings.

$$y_{ijt} = X_i\beta + W_{it}\Gamma + \alpha_{jt} + \epsilon_{ijt} \tag{2.1}$$

In some specifications we also include teacher fixed effects μ_i to account for unobserved time-constant teacher factors that might be correlated with W_{it} and principal ratings. While teacher fixed effects account for unobserved time-constant characteristics of teachers, we cannot control for changes in unobserved teacher characteristics across years. When including only school x year fixed effects in equation (2.1), β and Γ are primarily identified off of cross-sectional variation in ratings across teachers within schools. Models including teacher fixed effects are identified from changes in ratings of teachers over time.[6]

To examine the influence of specific school-level characteristics, we estimate specifications that include a set of observable school and/or principal characteristics, Z. One way to do this would be to simply include the school and principal characteristics in equation (2.1) instead of the school fixed effects. This could lead to bias since unobserved school-level factors are no longer absorbed by the school x year fixed effects. Our preferred method is to instead regress the estimated school x year fixed effects from the regression in equation (2.2) on Z.

$$\hat{\alpha}_{jt} = Z_{jt}\Pi + \nu_{jt} \tag{2.2}$$

[6]Very few teachers see their rating change over the four year period: among teachers with multiple ratings, just over 10% are given different ratings.

2.4.1 Sample selection and potential omitted variable bias

As noted above, the apparent discretion principals have in deciding when to rate teachers raises an important concern about sample selection. Specifically, if the principals decision to rate a teacher is correlated with an unobservable characteristic of the teacher and the rating itself, then our estimates between observable teacher characteristics and ratings may be biased. For example, if a principal observes a teachers low effort level (unobserved by the researcher) and chooses not to evaluate the teacher to avoid having to give a low rating, then the lower part of the ratings distribution would be censored. Depending on the relationship between unobserved effort and the regressor of interest (years of teaching experience, for example), the censoring could lead to bias in either direction. If only new teachers shirk, then the magnitude of the estimated relationship between teaching experience and ratings would be too large. If, on the other hand, the more experienced teachers tend to shirk, then the bias would be upwards.

To address selection bias we include a rich set of teacher demographics and background in all regressions. To the extent that these teacher observables are related to the decisions of principals to rate teachers this will mitigate bias. We also include school x year fixed effects to address potential heterogeneity in ratings behavior by principal and school. Finally, in some specifications, we include teacher fixed effects to control for all time-constant teacher characteristics. Even with teacher fixed effects, time-varying teacher unobservables could still be an issue if, for example, a principal believes that a teacher has underperformed due to a temporary low effort level and as a result delays giving the teacher a performance rating. We will present evidence below that this source of selection is not large enough to drive our results.

Of course, it is still possible that the estimates of these characteristics suffer from a more standard omitted variable bias. For example, it may be the case that high rates of absenteeism are associated with a bad attitude or shirking in other dimensions, and it is these factors – rather than the absences per se – that the principal considers when rating the teacher. In this case, one may not be able to say something definitive about principal views regarding teacher absenteeism per se, but rather about behaviors/characteristics associated with absenteeism, all of which presumably speak to productivity in some form or another. One omitted variable known to be related to ratings is a teachers academic value-added. Though our data does not permit its inclusion,[7] Rockoff (2004) finds a positive relationship between teaching experience and value-added, suggesting that the relationship between experience and ratings

[7] Only a modest number of teachers work in grades and subjects in which students take standardized tests, and given the time period of our analysis and the available data, we can only construct value added scores for a small fraction of the already select sample of teachers with performance ratings. For this reason, we do not include value-added measures in our analysis.

estimated from equation (2.1) will be biased upwards (though including teacher fixed effects will account for the time-constant component of value-added). Absences are also known to be negatively related to student achievement (Clotfelter et al., 2007; Miller et al., 2008). There is less concern about bias in the relationship between ratings and education credentials given the finding in Rockoff et al. (2008) that the quality and quantity of a teachers educational background is not strongly related to teacher value-added.

2.5 Results and Discussion

2.5.1 Who is rated?

As we described above, the frequency of principal ratings (as measured by the administrative data we have) do not seem to adhere to the requirements laid out in the teacher contract. Specifically, about a quarter of tenured teachers who should be rated based upon the guidelines are not. Also some teachers are rated only once or not at all during the four-year period of our sample. Only 62% percent of teachers who appear in our data set were rated at least once, and just 33% percent were rated multiple times. For this reason, we first examine whether observable teacher characteristics are associated with being rated. Table 2.2 presents the results from several linear probability models in which the dependent variable is always an indicator for whether the teacher was rated in the given year. Columns 1-3 report results on our full sample i.e., all teacher x year observations from 2003-2006. Columns 4-6 report results on the sample of teachers who appear in our data in all four years. Standard errors clustered by principal are reported in parentheses.

The model shown in column 3 includes school x year fixed effects and a full set of experience indicators along with the teacher characteristics shown in the table. Column 4 replicates this specification on the restricted sample of teachers. The specification shown in column 5 uses the four-year sample but includes teacher fixed effects instead of school x year fixed effects, while column 6 includes both sets of fixed effects. Including teacher fixed effects relies on within-teacher changes in rating status and characteristics over time for identification, rather than within-school x year cross-sectional variation.

Several important findings emerge. First, the school x year fixed effects account for roughly 60 percent of the variation in whether a teacher is rated, indicating that differences across schools and principals account for most of which teachers receive ratings. Second, conditional on the school x year fixed effects, there remains substantial selection in terms of who is rated. Black and Hispanic teachers are somewhat more likely to be rated while teachers with weaker educational backgrounds are less likely to be rated. Teacher absences are negatively related to the likelihood of being rated – i.e., arguably weaker teachers are less likely to be rated.

Third, limiting the sample to teachers in all four years reduces selection issues dramatically, as shown in column 4. The coefficients on teacher race drop substantially and the relationship between educational background and rating disappears. The coefficient on teacher absences declines (in absolute value) by roughly 40 percent. Thus the results using the "four-year" sample of teachers will have fewer selection concerns, though as discussed above, some external validity questions remain.

Figure 2.2a presents the relationship between teacher experience and the likelihood of being rated. The two lines correspond to columns 4 and 5 in Table 2.2. In the model without teacher fixed effects (corresponding to column 4), we see a modest, roughly linear, negative relationship between experience and the likelihood of being rated. First-year teachers are roughly 20 percentage points (50%) more likely to be rated than teachers with ten years of experience. The inclusion of teacher fixed effects makes this relationship even starker. Using within teacher variation, a teacher in her first year of teaching is roughly 50 percentage points (180%) more likely to be rated than when she is in her tenth year.[8] The difference between the teacher and school-

[8] Caution is warranted in interpreting the figure as a full experience profile. Since the data consist of a four-year panel of teachers, the full experience profile cannot be fully identified using only within teacher variation. The difference in ratings between a first- and fourth-year teacher is well-identified, while identification of the difference between a first- and twentieth-year teacher relies on a chain of

year fixed effects relationships indicates that there is some variation in the timing of ratings for observationally equivalent teachers within schools, but that the usual pattern for a teacher is to be rated at least once in the first few years of teaching.

Figure 2.3a presents analogous results for teacher absences. The relationship between absences and the likelihood of being rated is highly linear and negative. The coefficient on annual absences in column 4 of Table 2.2 implies that a teacher with ten additional absences in a school x year is about 2 percentage points less likely to be rated. Consistent with the figure, including teacher fixed effects (columns 5 and 6) doesnt alter the relationship significantly. This negative, but small relationship lessens the concern that principals delay rating teachers who demonstrate a temporarily lower performance level: teachers are only slightly less likely to be given a performance rating in years when they have unusually high absences.

2.5.2 What teacher characteristics are associated with higher ratings?

Table 2.3 shows the relationship between teacher characteristics and principal ratings for the set of teacher-year observations with ratings. The format of the table is analogous to Table 2.2 with the dependent variable now being a continuous measure of the rating the principal received (3=superior, 2=excellent and 1=satisfactory). Note

overlapping teachers at different points in their careers. Note that this is less of a concern for the absence relationships.

that teacher fixed effects pick up about 80% of the variation in ratings in column 5, suggesting an important role for time-constant qualities of teachers in determining rating levels.

To begin, consider the relationship between teacher experience and ratings, shown in Figure 2.2b (which is analogous in structure to Figure 2.2a). In the models without teacher fixed effects, we see that principal ratings increase with teacher experience through year ten, and then level off. These results suggest that a teacher with ten years of experience receives a rating roughly 0.4 points higher than a first-year teacher. Given that the standard deviation of ratings is 0.554, this represents an effect size of roughly 0.7. In the model with teacher fixed effects, we again see a strong positive relationship between ratings and experience through year ten, after which ratings decline with experience such that teachers with 35 years of experience are, on average, rated the same as first-year teachers.

Considering Figures 2.2a and 2.2b together allows one to gauge how much sample selection might influence the results shown in Figure 2.2b. We see that the average rating by experience group is higher when fewer teachers are rated. As we noted above, the general pattern of selection into being rated is that teachers of arguably higher quality are more likely to be rated. This would tend to lead us to overstate the relationship between experience and ratings. However, including teacher fixed effects, which control for all time-constant teacher characteristics, actually increases the magnitude of the experience relationship for newer teachers. The steep experience profile could in part reflect improvements in unobserved (by the researcher) teacher quality with experience, as might be reflected in value-added estimates.

As we discussed above, for selection to be driving the relationship when teacher fixed effects are included, there must be some time-varying, within-teacher unobservable factor determining a principals decision whether to rate a teacher. Principals avoiding rating teachers who performed unusually and temporarily poorly in a given year would generate upwards bias. This is plausible if principals are looking to protect their new teachers and present them in the best light (or avoid any costs associated with giving poor ratings). As discussed, the within-teacher relationship between absences and the probability of having a rating in Figure 2.3a does not suggest a large role for this source of selection. Teachers are only slightly less likely (2 percentage points for 10 additional absences) to be given a performance rating in years when they have unusually high absences. Of course, we cannot rule out the explanation that principals are deciding whether to rate teachers based on some other time-varying factor that we do not observe and that is also uncorrelated with absences.

Figure 2.3b shows the relationship between teacher absences and rating. In the models without teacher fixed effects, absences are negatively related to ratings. For example, a teacher with 10 additional absences during the school year would be rated roughly 0.1 point lower than an otherwise comparable teacher, an effect size of roughly 0.18 standard deviations. The inclusion of teacher fixed effects dramatically reduces (in absolute value) the coefficient on absences, so that ten additional absences are associated with a reduction of 0.03 rating points (0.05 standard deviations). This suggests that principals do not penalize teachers for year-to-year variation in absenteeism, but rather that absences serve as a marker for other unobservable factors associated with teacher quality.

The estimates shown in Table 2.3 indicate that the quality and quantity of a teachers educational background is also associated with ratings, conditional on experience and absences. A teacher with a Masters or Ph.D. degree receives a rating that is 0.05-0.06 points (about 0.1 standard deviations) higher than an observationally equivalent teacher with only a B.A. Effects for college selectivity and having an education major are similar, but smaller in magnitude. Unlike the experience relationship in Figure 2.2b, the relationship between education credentials and ratings are unlikely to be explained by correlations with value-added, given results in other work showing credentials do not explain variation in value-added. Of course, principals could be mistaking credentials for teaching quality in assigning higher ratings to these teachers. On the other hand, education credentials could be related to important dimensions of teaching quality not picked up by value-added scores.

Table 2.3 also indicates that several teacher demographics are associated with ratings, conditional on experience and absences. White female teachers (the omitted group) receive higher ratings than other teachers, with Black and Hispanic male teachers receiving the lowest ratings. Looking at Table 2.2 provides an indication of how important sample selection might be in these results. While there are differences by race x gender in the likelihood of being rated, it is noteworthy that in moving from the full sample to the four-year sample (columns 3 to 4), the magnitude of such selection drops dramatically. Indeed, there is no differential selection into rating for male teachers relative to white female teachers. However, as one moves from columns 3 to 4 in Table 2.3, the coefficients on teacher race-gender remain virtually unchanged. This strongly suggests that the relationship between teacher race and gender and principal rating are not due to sample selection.

While these coefficients are significantly different than zero at conventional levels, the coefficients on Hispanic and Black female teachers are quite small in magnitude (i.e., always less than 0.10 standard deviations). On the other hand, principals give male teachers ratings that are 0.20 to 0.30 standard deviations lower than white female teachers.

While it is tempting to infer from these results much about principal preferences over teacher demographics, several factors make the results difficult to interpret. First, we cannot control for a direct measure of teacher productivity, so that the relationship between demographics and ratings may reflect an unobserved measure of teacher effectiveness that is correlated with a demographic characteristic. Second, the hiring process itself can introduce selection bias with regard to teacher characteristics that are observable at the time of hire, such as race, gender or age. For example, if a principal has a strong preference for female teachers, then a male teacher hired by the principal must have some unobservable asset relative to an observationally equivalent female teacher hired by the principal. Because we do not observe this quality, it may lead us to understate the principals preference for female teachers.[9]

Table 2.4 shows results that use two alternative outcome measures – a binary indicator that takes on a value of one if the teacher received a superior rating and zero

[9]In theory, one could circumvent this concern by focusing on teachers who were not hired by the current principal, although even in this case one might be concerned about correlation of preferences across principals within the same school, particularly if certain views are commonly held in the profession (e.g., male teachers are not as effective as female teachers). In practice, there are not sufficient numbers of cases in which principals switched schools or were newly hired to obtain precise estimates from this approach. Fortunately, this type of selection will not bias coefficients on teacher absences and experience because these factors only become known after hiring.

if she received an excellent or satisfactory rating; and a binary indicator if the teacher received a rating of superior or excellent and zero if she received a satisfactory rating.[10] These results allow us to determine whether the relationships described above were driven by changes at the top or bottom of this (admittedly limited) distribution.

We find that the majority of the relationship between ratings and teacher characteristics is driven by the upper margin: teachers who are rated superior versus those rated excellent. For example, from column 2, a teacher with an additional ten absences relative to an otherwise similar teacher is six percentage points (nine percent) more likely to be rated superior. However, from column 3 the same teacher is only three percentage points (0.3 percent) more likely to be rated either superior or excellent. We see the same general pattern once we include teacher fixed effects in the specifications shown in columns 4 and 5.

Figures 2.4a and 2.4b show results at both margins of ratings for years of teaching experience from the school x year fixed effects and teacher fixed effects models. The solid lines, which show the conditional relationship between the binary superior rating outcome and experience, appear very similar to the experience results in Figure 2.2b. The dashed lines show little relationship between ratings and experience at

[10]An alternative to considering the margins separately, we estimated a multinomial logit model, which produced qualitatively similar results to those presented in Table 2.4.

the lower margin, the superior or excellent rating indicator. In Figure 2.4a teachers with 40 years of experience are no more likely to be rated superior or excellent versus satisfactory or unsatisfactory than teachers with one year of experience (but are more likely to be rated superior versus less than superior). We see in Figure 2.4b that once we include teacher fixed effects there is a small, negative relationship at the lower margin after ten years of experience. The effect at the lower margin is small relative to the much steeper trend in the solid line, the upper margin.

Figures 2.5a and 2.5b show a parallel analysis for teacher absences. Here the dashed and dotted lines appear roughly parallel in both the school x year fixed effects and teacher fixed effects models. Consistent with the linear specification in Table 2.4, however, the lines do diverge slightly as absences increase, so that the upper margin is the stronger relationship. Further, because the baseline mean is higher for the lower margin (0.956) than for the upper margin (0.719) in percent terms the two slopes imply an effect of nine percent at the upper margin versus only 0.3 percent at the lower margin for an additional ten absences.

Given that the majority of teachers are rated superior, we dont believe that the ratings are distinguishing among teachers at the highest level of quality. The results in Table 2.4 and Figures 2.4 and 2.5 suggest that neither do the ratings distinguish between the lowest quality teachers. Rather principals seem to use the ratings system primarily to distinguish between teachers in the middle of the quality distribution. Another possible explanation for these results is that we do not observe the characteristics that lead principals to give teachers low ratings. The R-squared values for columns 2 and 3 of Table 2.4 imply that observables (and school-year fixed effects) account for just 28% of the variation in ratings in the low margin regression, while we can explain 42% of the variation of ratings at the upper margin. These unobserved factors would need to be time-varying as the teacher fixed effect specifications in columns 5 and 6 also show larger relationships at the upper margin.

Table 2.5 presents results on the relationship between principal and school characteristics and ratings. Columns 1 and 2 show results just for the 488 elementary schools in the sample. Observationally equivalent principals in mixed race and integrated schools, schools with over 15% non-black and non-hispanic students, rate an average of 8.5 percentage points (22%) more of the teachers in the school x year than principals in predominantly black schools.[11] Older principals as well as new principals to a school tend to rate a smaller fraction of their teachers. Principals in elementary schools with ten percentage points higher fraction of new teachers rate 1.8 percent-

[11] In unreported results, we find no significant relationships between ratings and interactions between teacher and principal demographics.

age points (5%) more teachers. Figure 2.6a shows the relationship between school achievement (fraction proficient on the ISAT exam) and the conditional fraction rated in school. Principals in high achieving elementary schools do not rate a significantly different fraction of their teachers than those in lower performing schools.

Column 2 of Table 2.5 shows that new principals give lower ratings, but that most other observables about principals are not strongly related to ratings in elementary schools. Student demographics of schools also show no relationship to ratings. However, principals in schools with ten percentage points more teachers failing at least one certification test give ratings which are 0.06 rating points (0.11 standard deviations) lower. A similarly sized effect on ratings is seen in schools with more new teachers.

Figure 2.6b shows that observationally equivalent principals in high achieving elementary schools give higher ratings than those in lower achieving schools. A principal in a school with 50% of students performing at or above proficient levels gives ratings that are about 0.3 points (0.5 standard deviations) higher than in a school with only 20% proficient. The size of the effect fades out for higher performing schools. Mixed race and integrated schools are much more likely to be high achieving school than are schools with high minority compositions and so one might be concerned that the achievement effect on ratings might not be separable from that of mixed race and integrated schools. However, the solid line in Figure 2.6b shows that when one does not control for any school x year level observables, the relationship is only slightly stronger. This suggests that principals give higher ratings when student performance is higher, and not just because they are in schools that have characteristics associated with high performance. Due to a much smaller sample of high schools (358 versus 1,901 school x years), the results in columns 3 and 4 and Figure 2.7 are less precise than those for elementary schools.

2.6 Conclusion

We examine the relationship between the formal performance ratings that principals give teachers and a variety of observable teacher characteristics. Our findings suggest that these formal ratings are strongly related to proxies for productivity. Principals give higher ratings to teachers with more, and better, education credentials, more experience, and fewer absences. Teachers in higher performing schools receive higher ratings. The results for teaching experience are robust to specifications with teacher fixed effects, which rely on within teacher variation, effectively controlling for all time-invariant teacher characteristics. Moreover, we find that the relationship between absences and ratings is smaller when using within teacher variation, indicating that principals do not penalize teachers for year-to-year variation in absences.

These results provide some reason to be optimistic about policies that would assign more weight to principal evaluations of teachers in career paths and compensation. The results tell us that principals use a great deal of information available to them in determining what rating to assign a teacher. This is consistent with principals expending a fair amount of effort to get a teachers rating "right." That principals give lower ratings to teachers who tend to be absent more, but not necessarily who were absent more in the year the rating was given suggests principals are assigning ratings on the basis of underlying quality rather than solely on easily observable characteristics of teachers.

We are limited by the data in what we can say about whether principals are assigning ratings based on factors that are actually related to student performance. The relationship between experience and ratings likely represents improvements in teacher value-added with experience. We also view absences as a proxy for productivity, part of which is measured by academic value-added. In light of evidence in other work that education credentials are unrelated to students academic gains, the relationship between education credentials and ratings could mean that principals are mistaking credentials for teaching quality. On the other hand, education credentials could be related to important dimensions of teaching quality not picked up by value added scores.

Principals consider a great deal of information about their teachers when assigning ratings. Further research is necessary to more fully understand whether principals assign appropriate weight to the available information about teachers when making formal evaluations.

Tables

Table 2 1 Summary Statistics For Elementary and High School Teachers, 2003-2006

	All Teachers	Teachers in All Four Years
Demographics		
Male	0 217	0 215
Black	0 362	0 377
Hispanic	0 125	0 127
Age	45 008	46 562
First year teacher	0 054	0 013
Second year teacher	0 061	0 027
Third year teacher	0 061	0 040
Fourth year teacher	0 054	0 051
Fifth year teacher	0 047	0 048
Six or more years of teaching experience	0 723	0 821
Education		
Barron's rating (1-5)	2 003	2 013
Education major	0 542	0 590
Math major	0 092	0 092
Social Science/Humanities major	0 200	0 206
All other majors	0 185	0 130
Special education or bilingual	0 164	0 170
M A or Ph D degree	0 585	0 608
Teacher Attachment		
Part time	0 033	0 025
New to school	0 134	0 065
Rated last year	0 335	0 406
Efficiency Ratings		
Superior rating	0 227	0 276
Excellent rating	0 097	0 091
Satisfactory/Unsatisfactory rating	0 021	0 017
No rating	0 655	0 616
Fraction rated in 2003	0 161	0 195
Fraction rated in 2004	0 534	0 662
Fraction rated in 2005	0 098	0 089
Fraction rated in 2006	0 577	0 590
Overall fraction of teacher-years rated	0 345	0 616
Absences		
Annual absences	9 270	8 825
Number of observations	84,288	57,272
Number of teachers	27,886	14,318

Notes The table includes all probationary and tenured classroom teachers in Chicago Public Schools in the 2002-2003 through 2005-2006 school years The Barron's rating ranks undergraduate institutions on a selectivity scale from 1 to 5, where 1 is the lowest and 5 is the highest Unranked institutions are coded as 1

Table 2.2: Teacher Level Determinants of Receiving A Rating

	Sample = All Teachers			Sample = Teachers Observed in All Four Years		
Dependent Variable Teacher Rated in Year	Basic Teacher Demographics		All Teacher Controls	All Teacher Controls	Time Varying Teacher Controls	
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher Demographics						
Black female	-0 024**	0 024**	0 023**	0 008*		
	(0 009)	(0 004)	(0 004)	(0 004)		
Hispanic female	0 051**	0 034**	0 036**	0 013**		
	(0 009)	(0 005)	(0 004)	(0 005)		
White male	-0 008	-0 000	-0 001	-0 006		
	(0 006)	(0 004)	(0 004)	(0 004)		
Black male	-0 031**	0 014**	0 018**	0 009		
	(0 012)	(0 006)	(0 006)	(0 007)		
Hispanic male	0 036**	0 026**	0 028**	0 010		
	(0 013)	(0 008)	(0 007)	(0 009)		
Teacher Education						
Lowest undergraduate Barron's rating			-0 042**	-0 003		
			(0 003)	(0 003)		
Education major			0 080**	-0 003		
			(0 004)	(0 004)		
Math major			0 061**	-0 004		
			(0 005)	(0 006)		
Social science or humanities major			0 072**	-0 001		
			(0 005)	(0 005)		
Failed any certification test			0 012**	0 011**		
			(0 003)	(0 004)		
M A or Ph D			0 003	-0 007**	0 043**	0 023*
			(0 003)	(0 003)	(0 017)	(0 012)
Teacher Absences						
Annual absences/10			-0 035**	-0 020**	-0 018**	-0 013**
			(0 002)	(0 003)	(0 005)	(0 003)
Teacher Attachment						
New to school			0 003	0 021**	0 010	0 004
			(0 005)	(0 008)	(0 013)	(0 009)
Rated in prior year			0 035**	-0 021**	-0 362**	-0 331**
			(0 009)	(0 010)	(0 016)	(0 009)
Teaching Experience	All models include indicators for years of teaching experience categories See Figure 2 2a for results					
School x year fixed effects	No	Yes	Yes	Yes	No	Yes
Teacher fixed effects	No	No	No	No	Yes	Yes
Number of observations	84288	84288	84288	57272	57272	57272
Number of teachers	27886	27886	27886	14318	14318	14318
Number of schools	602	602	602	600	600	600
Mean of dependent variable	0 345	0 345	0 345	0 384	0 384	0 384
R-squared	0 016	0 608	0 619	0 710	0 580	0 843

Notes Each column is an OLS regression with the indicated fixed effects Standard errors robust to clustering at the principal level are in parentheses All models include controls for teacher fund and certification area indicators In addition to what is shown in the table the set of teacher demographics includes a cubic in teacher's age, and dummies for years of teaching experience categories The set of teacher attachment variables also includes a dummy for specially designated classroom teacher, a dummy for teacher in multiple schools, and a dummy for part time teachers The teacher education variables also include a special education or bilingual certification indicator The lowest undergraduate Barron's rating category includes schools rated in the least competitive category as well as unrated schools The specifications with teacher absences include a linear trend in absences for teachers having 40 or fewer absences (shown), an indicator for having more than 40 absences, a linear trend for teachers with more than 40 absences, and an indicator for missing absences The teacher fixed effect regressions do not include the cubic in age, but when the school-year fixed effects are not included a set of year fixed effects is included See Figures 2 2a and 2 3a for results for non-linear teaching experience and absences

Table 2.3: Relationship Between Teacher Ratings and Teacher Characteristics

Dependent Variable 3=Superior, 2=Excellent, 1=Satisfactory/Unsatisfactory	Sample = All Teachers with a Rating in Year			Sample = Teachers in All Four Years with a Rating in Year		
	Basic Teacher Demographics		All Teacher Controls	All Teacher Controls	Time Varying Teacher Controls	
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher Demographics						
Black female	-0 216**	-0 044**	-0 037**	-0 039**		
	(0 015)	(0 010)	(0 010)	(0 011)		
Hispanic female	-0 035*	-0 053**	-0 033**	-0 029*		
	(0 019)	(0 014)	(0 014)	(0 016)		
White male	-0 137**	-0 100**	-0 106**	-0 106**		
	(0 018)	(0 013)	(0 013)	(0 014)		
Black male	-0 393**	-0 196**	-0 180**	-0 179**		
	(0 032)	(0 023)	(0 024)	(0 029)		
Hispanic male	-0 220**	-0 200**	-0 180**	-0 194**		
	(0 036)	(0 022)	(0 022)	(0 025)		
Teacher Education						
Lowest undergraduate Barron's rating			-0 025**	-0 021**		
			(0 008)	(0 009)		
Education major			0 038**	0 024**		
			(0 009)	(0 011)		
Math major			0 002	0 002		
			(0 015)	(0 016)		
Social science or humanities major			0 004	0 004		
			(0 010)	(0 012)		
Failed any certification test			-0 034**	-0 022		
			(0 011)	(0 013)		
M A or Ph D			0 061**	0 052**	-0 034	-0 042
			(0 008)	(0 009)	(0 036)	(0 036)
Teacher Absences						
Annual absences/10			-0 106**	-0 093**	-0 033**	-0 035**
			(0 007)	(0 008)	(0 010)	(0 009)
Teacher Attachment						
New to school			-0 257**	-0 197**	-0 133**	-0 101**
			(0 017)	(0 023)	(0 026)	(0 030)
Rated in prior year			-0 098**	-0 134**	-0 022	-0 005
			(0 015)	(0 018)	(0 023)	(0 021)
Teaching Experience	All models include indicators for years of teaching experience categories See Figure 2 2b for results					
School x year fixed effects	No	Yes	Yes	Yes	No	Yes
Teacher fixed effects	No	No	No	No	Yes	Yes
Number of observations	29045	29045	29045	21988	21988	21988
Number of teachers	17335	17335	17335	11858	11858	11858
Number of schools	541	541	541	533	533	533
Fraction with superior rating (3)	0 658	0 658	0 658	0 719	0 719	0 719
Fraction with excellent rating (2)	0 281	0 281	0 281	0 237	0 237	0 237
Fraction with less than excellent rating (1)	0 060	0 060	0 060	0 044	0 044	0 044
Mean of dependent variable	2 598	2 598	2 598	2 676	2 676	2 676
Standard deviation of dependent variable	0 601	0 601	0 601	0 554	0 554	0 554
R-squared	0 124	0 414	0 441	0 427	0 841	0 899

Notes Each column is an OLS regression with the indicated fixed effects Standard errors robust to clustering at the principal level are in parentheses See notes on Table 2 2 for details on controls included in the models but not shown See Figures 2 2b and 2 3b for results for non-linear teaching experience and absences

Table 2.4: Alternative Teacher Ratings Outcomes and Teacher Characteristics

	Outcome = Rank (Baseline) (1)	Outcome = Superior Rating (2)	Outcome = Superior or Excellent Rating (3)	Outcome = Rank (Baseline) (4)	Outcome = Superior Rating (5)	Outcome = Superior or Excellent Rating (6)
Teacher Demographics						
Black female	-0.039**	-0.037**	-0.001			
	(0.011)	(0.009)	(0.004)			
Hispanic female	-0.029*	-0.025*	-0.003			
	(0.016)	(0.013)	(0.005)			
White male	-0.106**	-0.088**	-0.018**			
	(0.014)	(0.011)	(0.005)			
Black male	-0.179**	-0.137**	-0.042**			
	(0.029)	(0.022)	(0.011)			
Hispanic male	-0.194**	-0.152**	-0.042**			
	(0.025)	(0.020)	(0.011)			
Teacher Education						
Lowest undergraduate Barron's rating	-0.021**	-0.018**	-0.003			
	(0.009)	(0.007)	(0.004)			
Education major	0.024**	0.022**	0.002			
	(0.011)	(0.009)	(0.004)			
Math major	0.002	0.006	-0.004			
	(0.016)	(0.013)	(0.007)			
Social science or humanities major	0.004	0.005	-0.001			
	(0.012)	(0.010)	(0.005)			
Failed any certification test	-0.022	-0.022*	0.000			
	(0.013)	(0.011)	(0.005)			
M.A. or Ph.D.	0.052**	0.045**	0.007**	-0.034	-0.029	-0.004
	(0.009)	(0.007)	(0.003)	(0.036)	(0.030)	(0.017)
Teacher Absences						
Annual absences/10	-0.093**	-0.063**	-0.030**	-0.033**	-0.022**	-0.011**
	(0.008)	(0.006)	(0.003)	(0.010)	(0.008)	(0.005)
Teacher Attachment						
New to school	-0.197**	-0.156**	-0.041**	-0.133**	-0.106**	-0.027*
	(0.023)	(0.018)	(0.011)	(0.026)	(0.020)	(0.015)
Rated in prior year	-0.134**	-0.087**	-0.047**	-0.022	-0.028	0.005
	(0.018)	(0.015)	(0.008)	(0.023)	(0.018)	(0.009)
Teaching Experience	All models include indicators for years of teaching experience categories. See Figure 2.4 for results					
School x year fixed effects	Yes	Yes	Yes	No	No	No
Teacher fixed effects	No	No	No	Yes	Yes	Yes
Number of observations	21988	21988	21988	21988	21988	21988
Number of teachers	11858	11858	11858	11858	11858	11858
Number of schools	533	533	533	533	533	533
Mean of dependent variable	2.676	0.719	0.956	2.676	0.719	0.956
R-squared	0.427	0.416	0.276	0.841	0.841	0.742

Notes: Each column is an OLS regression with the indicated fixed effects. Standard errors robust to clustering at the principal level are in parentheses. The sample is limited to teachers observed in all four years with a rating in the year. See notes on Table 2.2 for details on controls included in the models but not shown. See Figures 2.4 and 2.5 for results for non-linear teaching experience and absences.

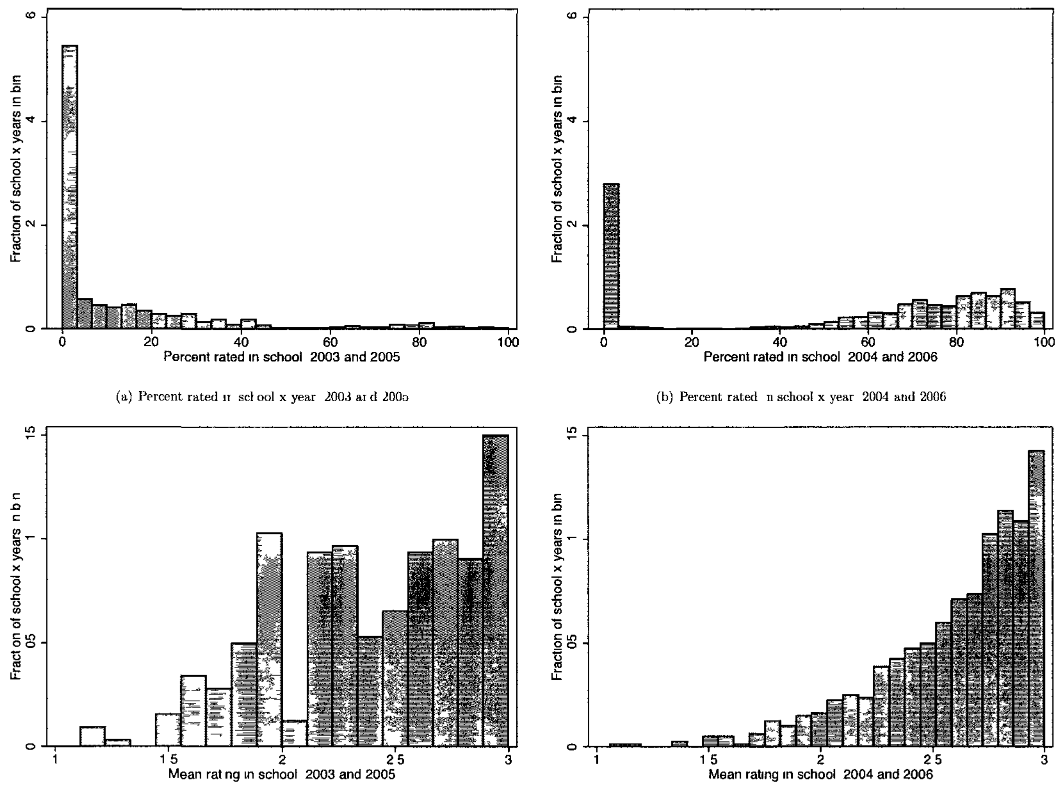
Table 2.5: School Level Determinants of Receiving A Rating and Ratings

Dependent Variable =	Elementary Teachers in All Four Years		High School Teachers in All Four Years	
	Teacher Rated in Year	Ranked Rating	Teacher Rated in Year	Ranked Rating
	(1)	(2)	(3)	(4)
School Demographics				
Predominantly minority	0.046	0.052	-0.061	-0.183
	(0.038)	(0.059)	(0.071)	(0.156)
Predominantly hispanic	0.023	-0.063	-0.219	0.278
	(0.049)	(0.074)	(0.143)	(0.275)
Mixed race or intergrated	0.085**	-0.044	-0.178*	0.198
	(0.043)	(0.065)	(0.099)	(0.192)
Enrollment/100	0.003	-0.001	0.006*	0.005
	(0.003)	(0.004)	(0.003)	(0.006)
Fraction proficient in high school	All models include cubics in fraction proficient			
Fraction proficent in elementary school	See Figures 2.6 and 2.7 for results			
Principal Demographics				
Principal new to school	-0.112**	-0.169**	0.129**	0.016
	(0.025)	(0.041)	(0.056)	(0.123)
Male principal	0.023	0.066*	0.141	0.118
	(0.027)	(0.040)	(0.087)	(0.182)
Principal black	-0.029	-0.044	0.129	0.066
	(0.030)	(0.046)	(0.092)	(0.185)
Principal hispanic	-0.017	-0.085	0.002	0.507**
	(0.036)	(0.054)	(0.113)	(0.228)
Principal black male	-0.042	0.023	0.179*	0.010
	(0.038)	(0.061)	(0.107)	(0.217)
Principal hispanic male	-0.037	-0.033	-0.168	0.549*
	(0.052)	(0.078)	(0.142)	(0.287)
Principal 50-55 years old	-0.019	0.022	-0.025	0.068
	(0.022)	(0.033)	(0.064)	(0.130)
Principal 55-60 years old	-0.041*	0.057*	0.027	0.012
	(0.022)	(0.034)	(0.064)	(0.123)
Principal over 60 years old	0.082**	0.065	0.002	0.337**
	(0.027)	(0.043)	(0.074)	(0.155)
Undergraduate Barron's rating (1-5)	0.003	0.004	0.011	0.089*
	(0.009)	(0.014)	(0.026)	(0.053)
School Year Level Teacher Demographics				
Fraction w/ lowest undergraduate Barron's rating	-0.074	-0.132	-0.013	-0.724*
	(0.069)	(0.107)	(0.227)	(0.432)
Fraction w/ M.A. or Ph.D.	-0.036	0.105	-0.066	0.478
	(0.070)	(0.111)	(0.220)	(0.488)
Fraction failed any certification test	0.143	-0.611**	-0.248	0.562
	(0.116)	(0.180)	(0.355)	(0.799)
Mean annual absences/10	-0.077	-0.074	-0.011	0.528*
	(0.047)	(0.075)	(0.138)	(0.285)
Fraction new to school	0.179**	0.458**	-0.045	-0.496*
	(0.088)	(0.142)	(0.141)	(0.281)
Fraction rated in prior year	0.035	-0.132**	0.127	-0.014
	(0.032)	(0.058)	(0.089)	(0.204)
Number of observations	1901	751	358	126
Number of schools	488	440	112	93
Mean of dependent variable	0.393	2.701	0.362	2.607
R squared	0.384	0.367	0.378	0.584

Notes: All columns show the results of a second stage regression of the school-year fixed effect estimates from column (4) of Table 2.2 or Table 2.3 on school and principal demographics and a set of year fixed effects. The standard errors for the second stage models are estimated assuming an identity weighting matrix, which provides consistent, but not efficient results. In addition to what is shown, the set of controls for school demographics also includes an indicator for missing student achievement, an indicator for multiple schools in the same building, and indicators for whether a school is a magnet school or achievement academy. The set of controls for principal demographics also includes an indicator for whether the principal has a B.A. in education and missing data indicators. The set of controls for school-year level teacher characteristics also includes means of all of the teacher level vairables from Table 2.2. For elementary schools school academic achievement is the enrollment weighted percentage of third, fifth, and eighth grade students in a school x year testing at or above proficiency on the ISAT averaged across math and reading. For high schools school achievement is the percentage of students in a school x year testing at or above national norms on the PSAE. The proficiency scores are demeaned separately for elementary and high schools so the coefficient on the elementary school indicator is the conditional difference in the fraction rated in elementary and high schools at the mean proficiency levels. Mixed race or integrated schools are schools with over 15% non-black and non-hispanic students.

Figures

Figure 2 1 Density of Percent Rated and Mean Rating in School x Years



(a) Percent rated in school x year 2003 and 2005

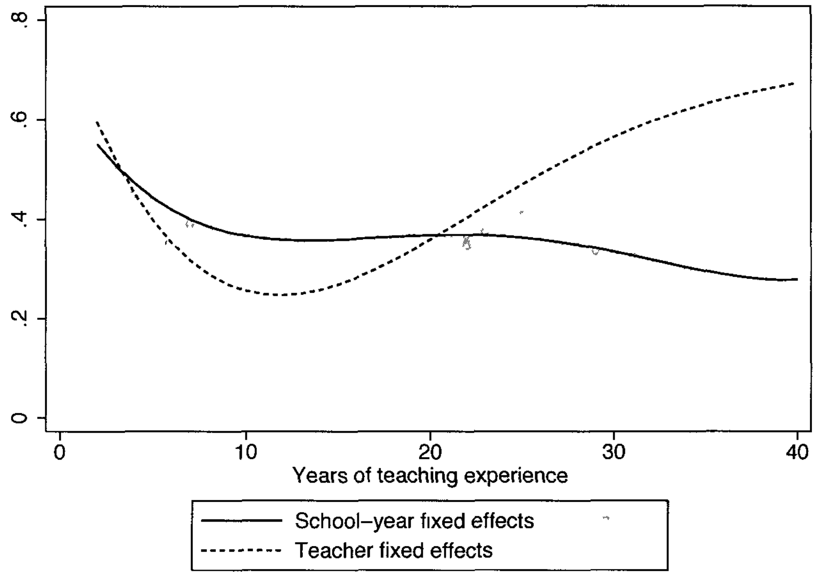
(b) Percent rated in school x year 2004 and 2006

(c) Mean rating in school x year 2003 and 2005

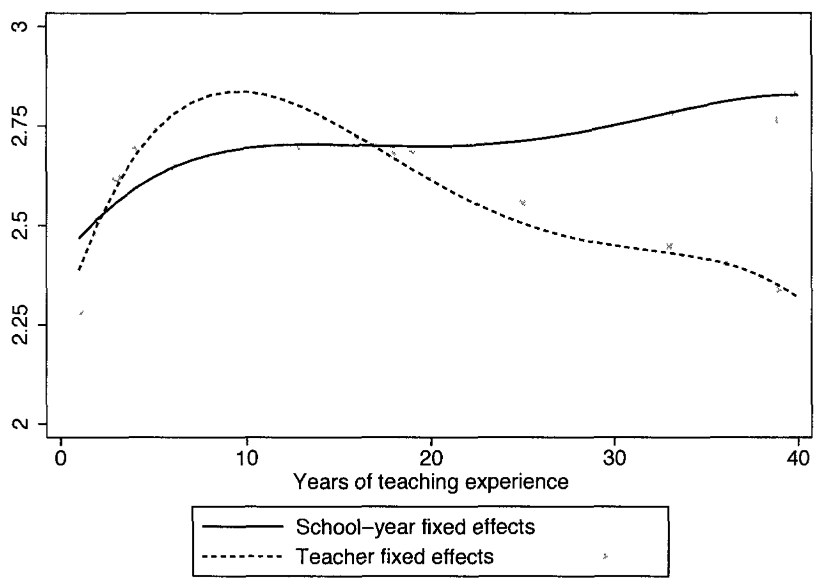
(d) Mean rating in school x year 2004 and 2006

Notes The sample used to generate the figure is the set of all probationary and tenured teachers in 2002-2003 through 2005-2006 A superior rating is coded as a 3, excellent as 2, and satisfactory and unsatisfactory ratings as 1 Panels (c) and (d) only contain school x years that rate at least five teachers

Figure 2.2: Probability of Receiving a Rating and Mean Rating by Teaching Experience



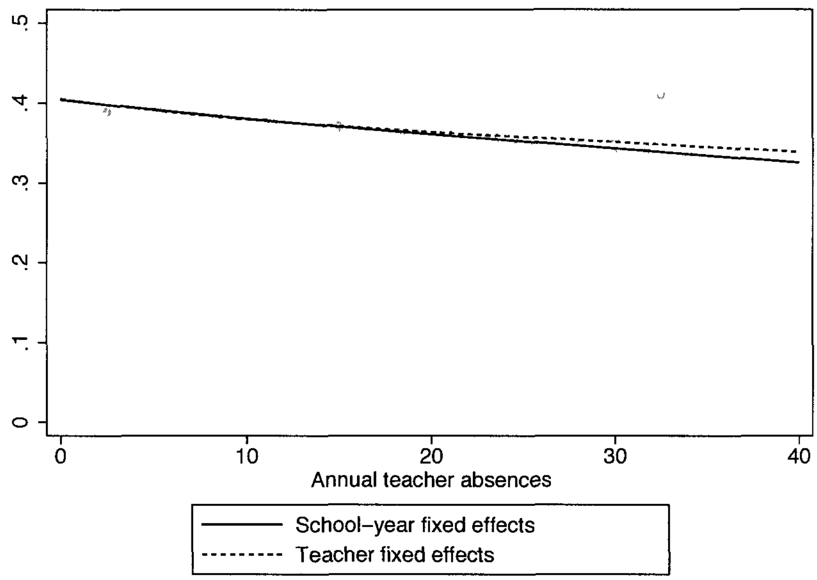
(a) Probability of receiving a rating



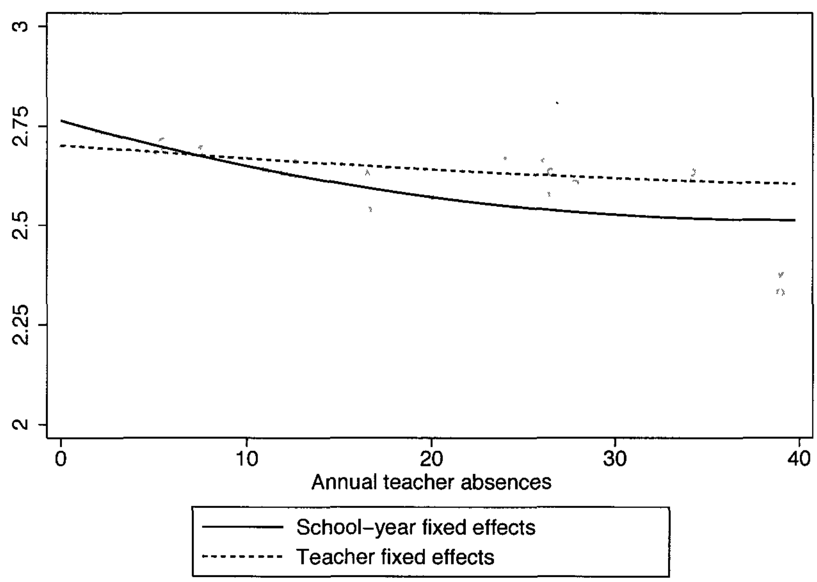
(b) Mean rating

Notes: The sample used to generate the figure is the set of probationary and tenured teachers observed in all four years between 2003 and 2006. The y-intercept of each line is calculated at $\bar{X}$, the mean of the teacher covariates included in the model. The trend lines show estimates of a quartic in experience conditional on the full set of teacher covariates described in the notes of Table 2.2 and the indicated set of fixed effects. The scatterplots show the conditional mean for each year of experience. A superior rating is coded as a 3, excellent as 2, and satisfactory and unsatisfactory ratings as 1.

Figure 2.3: Probability of Receiving a Rating and Mean Rating by Teacher Absences



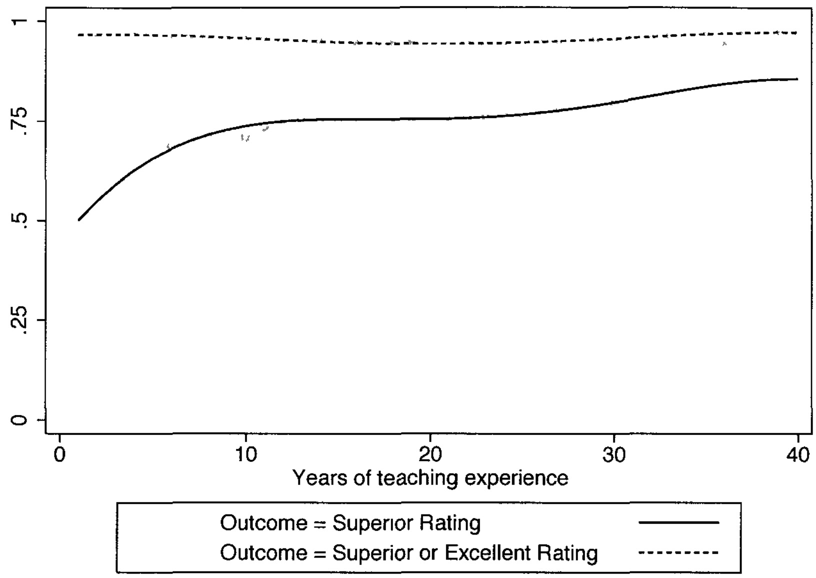
(a) Probability of receiving a rating



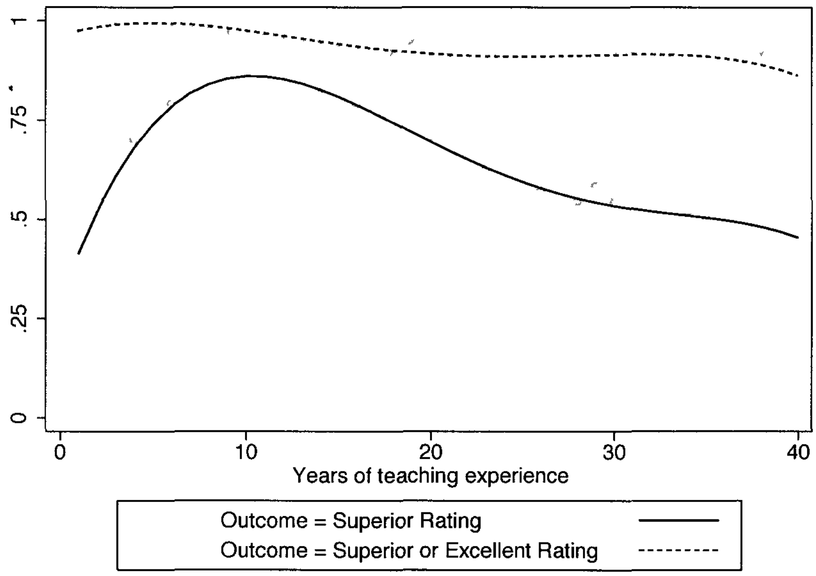
(b) Mean rating

Notes: The sample used to generate the figure is the set of probationary and tenured teachers observed in all four years between 2003 and 2006. The y-intercept of each line is calculated at $\bar{X}$, the mean of the teacher covariates included in the model. The trend lines show estimates of a quartic in absences conditional on the full set of teacher covariates described in the notes of Table 2.2 and the indicated set of fixed effects. The scatterplots show the conditional mean for each value of teacher absences rounded to the nearest integer. A superior rating is coded as a 3, excellent as 2, and satisfactory and unsatisfactory ratings as 1.

Figure 2.4: Alternative Ratings Outcomes by Teaching Experience



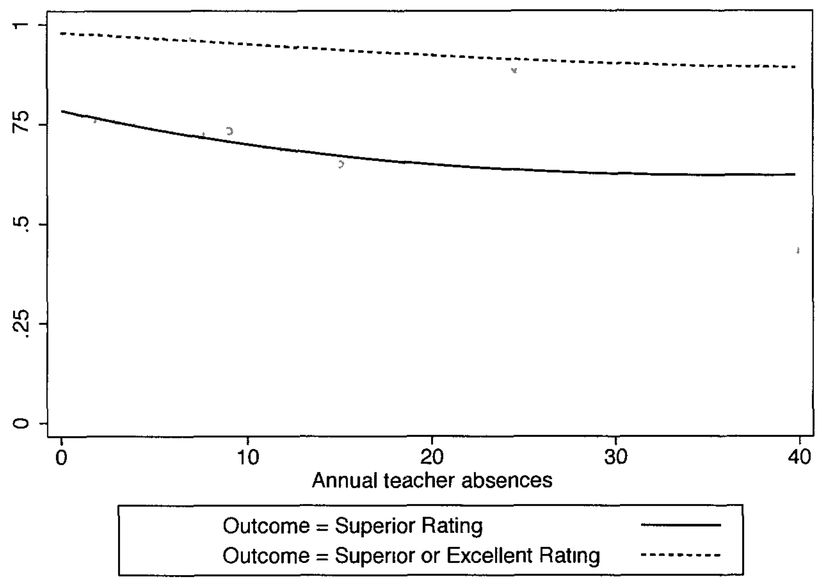
(a) School-year fixed effects



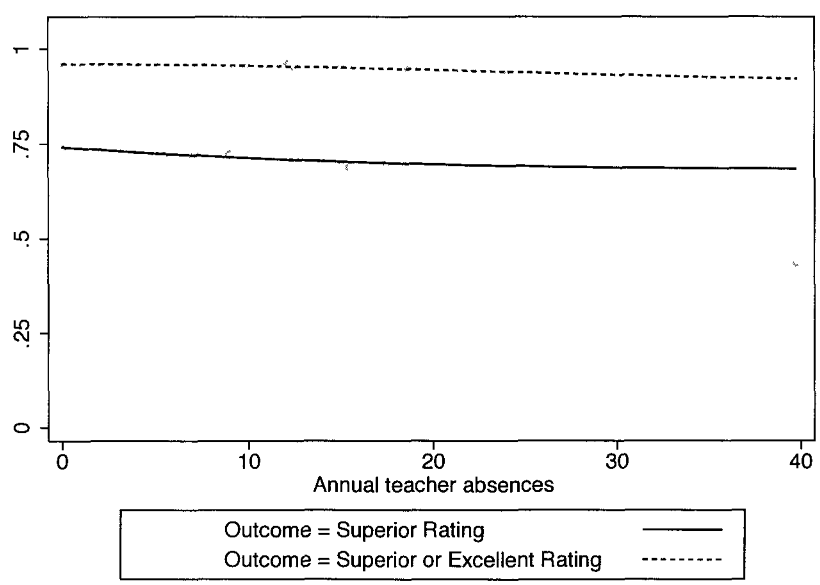
(b) Teacher fixed effects

Notes: The sample used to generate the figure is the set of probationary and tenured teachers observed in all four years between 2003 and 2006. The y-intercept of each line is calculated at $\bar{X}$, the mean of the teacher covariates included in the model. The trend lines show estimates of a quartic in experience conditional on the full set of teacher covariates described in the notes of Table 2.2 and the indicated set of fixed effects. The scatterplots show the conditional mean for each year of experience.

Figure 2.5: Alternative Ratings Outcomes by Teacher Absences



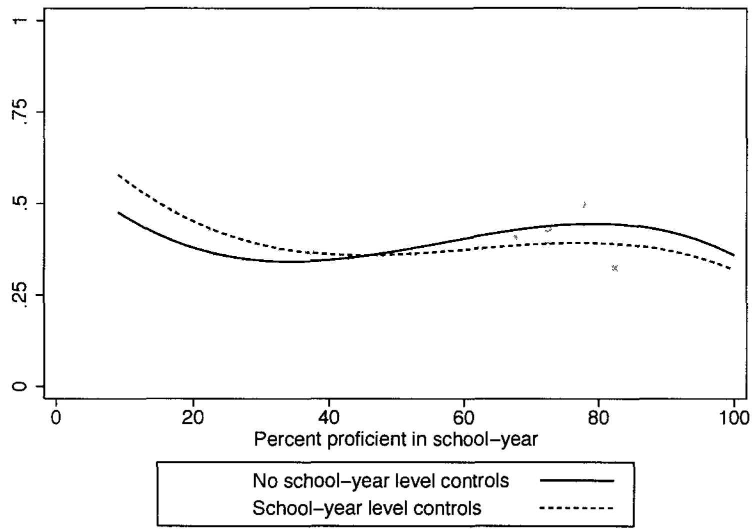
(a) School-year fixed effects



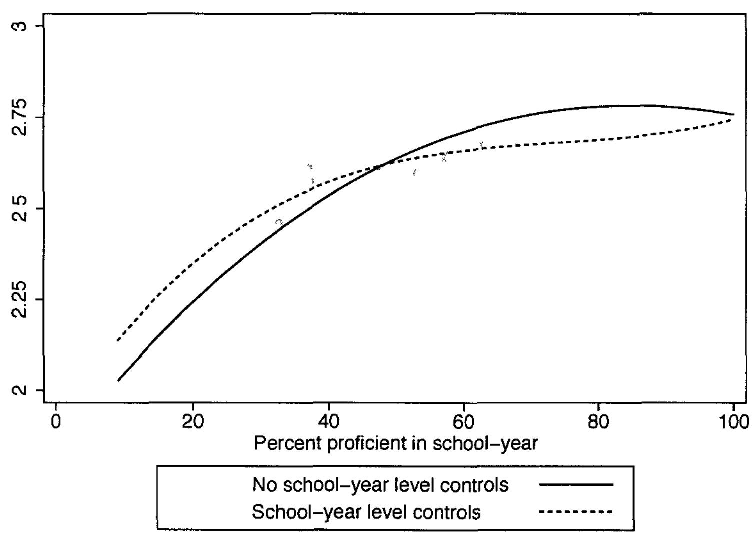
(b) Teacher fixed effects

Notes: The sample used to generate the figure is the set of probationary and tenured teachers observed in all four years between 2003 and 2006. The y-intercept of each line is calculated at $\bar{X}$, the mean of the teacher covariates included in the model. The trend lines show estimates of a cubic in absences conditional on the full set of teacher covariates described in the notes of Table 2.2 and the indicated set of fixed effects. The scatterplots show the conditional mean for each absence category.

Figure 2.6: Probability of Receiving a Rating and Mean Rating by School Achievement in Elementary Schools



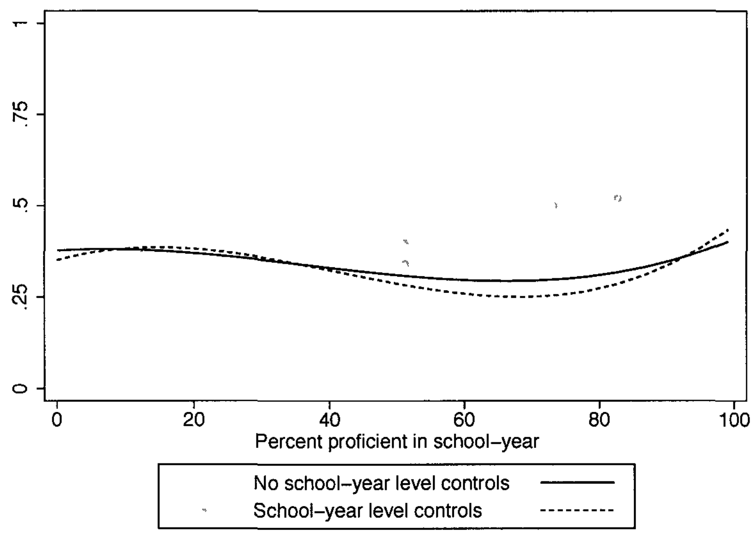
(a) Probability of receiving a rating



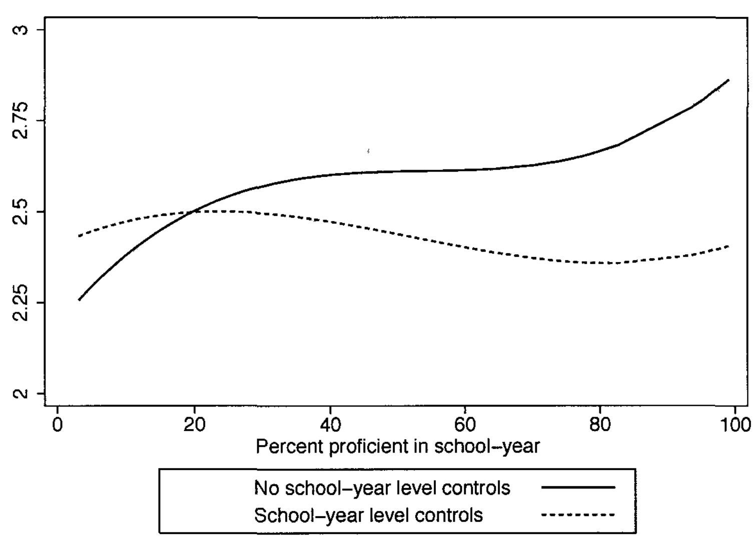
(b) Mean rating

Notes: The sample used to generate the figure is the set of all probationary and tenured teachers observed in all four years between between 2003 and 2006. The y-intercept of each line is calculated at $\bar{Z}$, the mean of the principal covariates included in the model. The trend lines show estimates of a cubic in percent proficient from a regression of school-year fixed effects from a model including the full set of teacher covariates on year dummies and school-year covariates as indicated. The scatterplots show the conditional mean rating or fraction rated for bins of fraction proficient. The points are shown at the mean fraction proficient for each bin. See the notes on Table 2.5 for more details on estimation and the school achievement measure..

Figure 2.7: Probability of Receiving a Rating and Mean Rating by School Achievement in High Schools



(a) Probability of receiving a rating



(b) Mean rating

Notes: The sample used to generate the figure is the set of all probationary and tenured teachers observed in all four years between between 2003 and 2006. The y-intercept of each line is calculated at $\bar{Z}$, the mean of the principal covariates included in the model. The trend lines show estimates of a cubic in percent proficient from a regression of school-year fixed effects from a model including the full set of teacher covariates on year dummies and school-year covariates as indicated. The scatterplots show the conditional mean rating or fraction rated for bins of fraction proficient. The points are shown at the mean fraction proficient for each bin. See the notes on Table 2.5 for more details on estimation and the school achievement measure.

CHAPTER III

School Entry Policies and Learning: Test Score Effects of Pre-Kindergarten Enrollment, Maturity, and Time in School

3.1 Introduction

Children in the US generally begin kindergarten late in their fourth year or during their fifth year of age. The timing of entry into school is in part determined by school entry policies. For example, if a state's entry cutoff date is September 15th, then a child who turns five prior to September 15th will begin school a full year earlier than children with birth dates in late September through December. Thus, under compliance with the policy, a child with a birth date in early September will be a grade level ahead of students born in late September of the same birth year. In addition, the student born in early September will be relatively younger than peers in her same entering cohort.

Compliance with entry policies is not universal, however, as parents may delay entry an additional year if they wish, or circumvent the entry policy to begin formal schooling early. Parents, especially those with high socioeconomic status, increasingly are giving their children an extra year out of school – a practice often called "redshirting" – to take advantage of perceived benefits to children of being relatively older than their classmates (Deming and Dynarski, 2008).

In addition to parents' decision about when to enroll their child in school, the state (or in some cases the school district) must decide on a cutoff date. An earlier cutoff date will make entering classes older, since more children will miss the cutoff date. Perhaps as a result of increasing emphasis on end-of-year test scores, many states have moved up their cutoff dates, which tends to increase the age at which children take these tests (Elder and Lubotsky, 2009).[1]

Age is not the only relevant mechanism that states and parents need consider; entry policies generate variation not only in the age of students when they enter school, but also in their activities prior to entering kindergarten. Figure 3.1 plots the reported pre-kindergarten enrollment of children by month of birth relative to

[1]Higher test scores as a result of children being older will not necessarily translate into long- or even medium-run gains. Students who enter school when older are more likely to drop out of high school when they exceed the compulsory schooling age at an earlier grade level (Angrist and Krueger, 1991, 1992; Deming and Dynarski, 2008). Deming and Dynarski (2008) argue that as a result of redshirting and the effects of compulsory schooling laws, the trend across states toward the "graying" of kindergarten also is likely to increase inequality in educational outcomes.

the state school entry cutoff using data from the National Assessment of Educational Progress Long-Term Trend (LTT).[2] Students assigned to enter school early, those on the left of the cutoff line, report participating in pre-kindergarten at a lower rate (about 66 percent) than those assigned an extra year out of school before entry (about 71 percent).[3] Thus, relatively older students in a given grade level might perform better on cognitive assessments than their younger peers as a result of their age (both relative to their entering cohort and absolute age effects from maturity and development), or as a result of richer pre-school experiences.

If older-entering students have higher test scores than younger peers in their cohort only because they participate in richer pre-school activities, then parents might be better served by enrolling students in pre-school activities at a younger age rather than redshirting.[4] Also, if differential pre-school activities are responsible, at least in part, for "age-at-test" effects, then by moving up entry cutoff dates, states raise student outcomes in part by passing off some of the costs of schooling to parents (or those who subsidize pre-school programs).

The primary research question motivating this paper is whether investments in

[2] I discuss the data and formal estimation of the effects of the entry policies below.

[3] The measure of pre-kindergarten enrollment is based on student responses to a survey question, so there could be significant measurement error in the levels.

[4] The benefits of enrolling in pre-school at an early age are likely much smaller, or even negative, for very young children, e.g. a two year old is unlikely to derive much benefit from pre-school programs targeted to four year olds.

children's pre-school activities can account for the effects of entry policies on student outcomes. I develop a dynamic model of human capital accumulation while in school and demonstrate that a simple model with no relative age effects, but with differences in early childhood experiences, could account for patterns in the effects of entry policies found in prior work. The key contribution of the paper is to utilize information on pre-school experiences of students and student and parent investments in schooling to explain the evolution of the effects of entry policies with age.

For the empirical analysis, I employ the National Assessment of Educational Progress Long-Term Trend (LTT) surveys of 9 and 13 year-old students in the US over time. Using the LTT series, I estimate the "reduced form" effects of entry policies on math scores measured when students are the same age, but in different grades, and also measured at the same assigned grade, but at different ages. To estimate these parameters, I use a combination of fuzzy regression discontinuity design and instrumental variables approaches. In both cases, I exploit the school entry policies as a source of exogenous variation in timing of school entry affecting student outcomes through age 13. I also exploit variation in the response to and effects of entry policies across subgroups to separately identify the effect of pre-kindergarten participation from the effects of age and time in school.

I find that the implied effect of pre-kindergarten attendance on math scores needed to account for the entire effect of entry policies often attributed to age is implausibly large. However, I do find evidence of very large effects of pre-kindergarten attendance on math scores, though the estimates are imprecise. I find that the effects of entry policies on math scores (holding either age or assigned grade constant) fade out between age 9 and 13. The decline with age in the effects of entry policies is consistent with a model of human capital accumulation with depreciation of early learning gains.

The rest of the paper proceeds as follows: Section 3.2 reviews the prior literature using state school entry policies as a source of exogenous variation in measuring maturity and schooling effects on academic outcomes; Section 3.4 presents the data used in the analysis; Sections 3.5 and 3.6 present the empirical strategy and results. The paper concludes with a discussion of the implications of the key findings.

3.2 Literature Review

Students enter school at different ages for a variety of reasons. First, students in a birth cohort are different ages (measured in months or days) on the first day of school simply as a result of variation in day of birth. Next, parents have some ability to choose the age of enrollment by "red-shirting" their child, the increasingly common practice of delaying a child's entry into school (Deming and Dynarski, 2008). Finally, state school entry policies determine the earliest age at which a student is allowed to begin school.

Prior work has used entry policies as a source of plausibly exogenous variation in age of entry into school to estimate parameters related to the production of student outcomes, including the value of time in school, maturity or development, and age relative to peers. However, this work has focused almost exclusively on reduced form effects with ambiguous relevance, and it has ignored differences in out-of-school experiences between early- and late-entering children.

Students who enter school when relatively older than their same-grade peers tend to perform better on cognitive assessments. When comparing students who enter school when relatively older to younger entering students *when they are the same age*, however, the "young" students (who have completed an additional year of formal schooling) outperform the "old" students. I refer to these two types of comparisons as the "same grade, different age" and "same age, different grade" comparisons. The former comparison is often discussed as an estimate of the effect of age (encompassing "relative age" differences[5] as well as maturity or development). The latter comparison is discussed as an estimate of the value of time in school relative to the typical out-

[5]One can think of relative age effects as arising from the degree of success of the match between curriculum and age, or as a result of peer effects.

of-school experiences of a child. (In Section 3.3, I develop a model to formalize these concepts and describe these comparisons in a structural context.)

3.2.1 Same grade, different age

The motivation for the same grade, different age comparison is to identify parameters related to the effect of age[6] on student outcomes: most generally, the reduced form effect of delaying school entry on student outcomes measured when in the same grade as early-entering students. Cross-sectional comparisons of students of different ages within a grade level will produce biased estimates of the effects of age on student outcomes to the extent that non-compliance with the entry policies leads to selection on student characteristics related to the outcome of interest.[7] To address the problem of selection bias, prior work estimating the reduced form effect of relative age has used an instrumental variables (IV) approach, exploiting the variation in entry age generated by compliance with the entry policies by instrumenting for actual age of entry with assigned age of entry.

[6]Though conceptually distinct, here and elsewhere in the text, I refer to effects of relative age and absolute maturity as the effects of age, generally. Empirically, plausibly separating the effects of relative age and absolute age is not generally possible because of the collinearity between the age of the student at the time of an assessment and the relative age of the student to her peers. See Black et al. (2008) for one effort.

[7]Additional bias would result from any direct effects of day of birth on outcomes. Dickert-Conlin and Elder (2010) find no evidence of the manipulation of the timing of births to game the entry cutoffs, nor of differences in infant or maternal health differences by birth date.

Datar (2006) and Elder and Lubotsky (2009) find that students induced by entry policies to delay kindergarten entry by a full year perform about 0.5 to 0.8 standard deviations better on standardized tests than their earlier entering (younger) peers at the time of entry. Datar (2006) also finds evidence of small additional gains by the end of kindergarten for the older students, though Elder and Lubotsky (2009) show that these small gains quickly dissipate and that the effect of the year of relative age difference is reduced to about 0.2 standard deviations by eighth grade. Bedard and Dhuey (2006) and McEwan and Shapiro (2008) also examine the test scores of students in later grades and find that some of the early differences between young and old students persist through at least eighth grade.[8]

The large performance difference between relatively older and younger students in the fall of kindergarten, before participation in formal schooling can have an effect on student outcomes, suggests that absolute maturity or differential out-of-school experiences of children (and not relative age) is an important part of the reduced form same grade, different age comparison. The decreasing same grade, different

[8]Complicating the interpretation of effects of entrance age in later grades is the finding (for students in Chile) that relatively older students (by one year) are 2.1 percentage points less likely to be retained in first grade (McEwan and Shapiro, 2008), and the finding that relatively younger students are more likely to be diagnosed with a learning disability such as ADD/ADHD (Elder and Lubotsky, 2009; Dhuey and Lipscomb, 2010). Also, Bedard and Dhuey (2006) show that relatively younger students are less likely to attend college, though given the interaction of entrance age with compulsory schooling laws (Angrist and Krueger, 1992), this effect cannot be solely attributed to the persistent effects of the front-end entrance age policy.

age comparison as students advance through school presents a puzzle for the relative age explanation for the entry age premium after kindergarten. While relative age differences within an entering class might become less significant as students age, given the small magnitude of the relative gain in test scores for older students between the start and end of kindergarten, fading relative age effects cannot explain the large drop in the entry age premium between the end of kindergarten and the middle grades.[9] However, as I show in Section 3.3, the decline in the same grade, different age comparison is consistent with depreciation of the effects of early childhood investment.

3.2.2 Same age, different grade

The motivation behind the "same age, different grade" comparison is to obtain estimates of parameters related to the effect of instructional time. Papers investigating the effects of time in school tend to use regression discontinuity (RD) approaches to compare students very close in age, but born on different sides of the school entry cutoff date. While the RD approach nets out the effects of absolute age (maturity), the age of students on either side of the cutoff fall at the opposite, extreme ends of the age distribution of their same grade peers, so relative age effects remain a confounding

[9]Also, "multiplier" effects of accumulated knowledge on the rate of new learning (a complementarity between the stock of human capital and learning (Cunha et al., 2006; Bedard and Dhuey, 2006; Elder and Lubotsky, 2009)) would tend to produce increasing same grade, different age effects with grade level, as the direct benefits of being older compound with time.

factor. Thus, as with the "same grade, different age" comparison, even given ideal data and no selection concerns, the parameter obtained in this sort of analysis is not a clean estimate of a single structural parameter.

Early attempts at estimating the effect of time in school using variation in age of entry generated by school entry policies such as Cahan and Davis (1987) and Cahan and Cohen (1989) suffer from severe selection concerns as a result of excluding non-compliers with entry policies from analysis. Cascio and Lewis (2006) address these selection concerns by sampling on the age of students rather than grade, thus comparing students with birth dates on either side of the school entry cutoff date regardless of actual enrolled grade level. High school students induced to be in a higher grade level than same-age students as a result of entry policies perform better on assessments of cognitive skills.

However, because students with more schooling are also relatively young for their grade level, the RD parameter is a lower bound for effects of a year of school when relative age effects (benefitting relatively older students) are present. Students with more schooling had less time out of school and are less likely to participate in formal pre-school activities, providing additional reasons to treat the RD parameter as a lower bound for the effect of a year in school.[10]

[10]Complicating interpretation further, because both school entry policies and compulsory school-

3.2.3 Effects of early childhood programs

Because entry policies affect the pre-school experiences of children, some portion of the effects of entry policies discussed above are likely due to the benefits of early childhood programs. Recent work has found that the quality of schooling received at an early age (pre-kindergarten and kindergarten) has long-run effects on educational attainment and labor market outcomes (Cunha et al., 2006; Heckman et al., 2010; Chetty et al., 2010; Dynarski et al., 2011).[11] The benefits of participation in early childhood programs appear immediately; a wide body of work has found large effects of early childhood programs on student outcomes measured within a few years of participation, some using experimental or quasi-experimental evidence.[12]

Anderson (2008) re-examines evidence from three influential randomized trials of early childhood interventions: The Abecedarian Project, The Perry Preschool Program, and the Early Training Project. In all three programs, the treated students

ing laws bind for students in the Cascio and Lewis (2006) sample, any direct effects of time in school are the result of both front-end and back-end attendance policies. An extra year of schooling at a young age could have different (persistent) effects from an extra year directly leading to degree completion. Also, front- and back-end interventions likely affect different student populations, associated with different local average treatment effects (LATE). This last point is related to the violations of monotonicity for LATE estimates of the effects of compulsory schooling laws discussed in Barua and Lang (2009) and Aliprantis (2010).

[11] Long-run effects of entry policies such as those found in Angrist and Krueger (1991) and Angrist and Krueger (1992) could be partly determined by early childhood experiences, and not only the effects of compulsory schooling laws on educational attainment.

[12] Magnuson et al. (2007) use non-experimental methods to estimate the effect of pre-kindergarten attendance on test scores in kindergarten and first grade. Gormley, Jr. and Gayer (2005) describe earlier non-experimental studies of pre-kindergarten enrollment and student outcomes.

have larger IQ scores at age 5, by about 0.6 to 1 standard deviation, than the control students. The IQ effect fades out with age, though the programs have long-run effects on educational attainment and labor market outcomes.[13] The Head Start program, has also been shown to have positive effects on student outcomes. Ludwig and Phillips (2007) use data from the National Head Start Impact Study, a randomized control trial, to show that participation in Head Start leads to higher scores on a variety of assessments at age 3 and 4 by 0.1 to 0.4 standard deviations.[14]

One concern with results from the Head Start experiment and the other preschool program experiments is that it is not clear what the control group is experiencing.[15] The effects of a preschool program relative to no program participation might be expected to be larger than if some members of the control group participate in another program. Also, Head Start and the programs studied by Anderson (2008) tend to be targeted and highly structured, making it difficult to bound the effects of participation in the types of pre-school programs typical for students in the US. Indeed, we know much less about the effects of typical pre-kindergarten programs on typical children.

Gormley, Jr. and Gayer (2005) evaluate a mandatory pre-kindergarten program in

[13]See Heckman et al. (2010) for additional evidence of the long-run effects of the Perry Preschool Program.

[14]Also see Puma et al. (2005) for results from the National Head Start Impact Study.

[15]One exception is Currie and Thomas (1995), who find a similar effect of Head Start (based on family fixed effect models) on test scores relative to both students who attend another type of preschool and those who attend no preschool at all.

Oklahoma using a "same age, different grade" regression discontinuity approach and find that students born just before the entry cutoff perform about 1/3 of a standard deviation better on cognitive and some non-cognitive assessments than those just missing the cutoff at the time of pre-kindergarten exit for the early-entering students (so that the late-entering students had yet to enroll in pre-kindergarten).[16] Though, as with grade-level effects in later grades, the possibility of unobserved differences in pre-school (here pre-pre-kindergarten) activities complicates interpretation. For example, students who miss the pre-kindergarten entrance cutoff may be more likely to be enrolled in day care programs in the year prior to pre-kindergarten enrollment.

3.3 Theory

The parameters obtained from empirical strategies that exploit variation from school entry policy are best understood in the framework of a model of human capital accumulation. A model capturing the contributions of relative age, maturity, time in school, and early childhood experiences to human capital accumulation informs the interpretation of the "same age, different grade" and "same grade, different age" parameters.

I model the human capital accumulated by a student born on day c measured at

[16] Barnett et al. (2005) find similar results for non-mandatory pre-kindergarten participation across five states using a similar RD design.

a given age t as a function of the depreciated stock of accumulated human capital, and the contributions of family investments I_t and time spent in school θ_t

$$h_{ct} = \beta h_{c\,t-1} + I_t(Y, S_{ct}) + \theta_t(S_{ct}, h_{c\,t-1}) \tag{3 1}$$

$1 - \beta$ gives the rate of human capital depreciation Investment is a function of family resources Y and can vary with the student's current grade in school S_{ct} The effect of time in school on human capital accumulation is a function of S_{ct} and the stock of human capital the student brings into the school year [17] (Equation (3 1) suppresses an individual subscript for economy of notation) Time t indexes academic years and gives the age of a student in years on the first day of school

Maturity or development enters the model as a component of I, and can vary in importance with age t Whether a child participates in an early childhood program also enters through I and varies with Y and where the child's birthdate c falls relative to the school entry cutoff date As with investment, the t subscript on the schooling function θ allows the contribution of a given grade level S to vary with age, introducing relative age effects into the model

[17] Equation (3 1) is similar to the model put forward in Elder and Lubotsky (2009) One difference is that in their model, it is not the stock of human capital h that enters the schooling contribution function θ, but the age of entry into formal schooling In both cases the idea is to capture the concept of "grade readiness" (complementarity between a child's knowledge base or maturity and new learning in school) The formulation in equation (3 1) is more general since a student's stock of human capital is only in part determined by the age of entry

Because not all students enter school at the same age, the human capital accumulated by age t by observationally equivalent students born in the same calendar year will vary. State school entry policies generate variation in a student's grade at age t so that students can be enrolled in formal schooling ($S \geq 0$), pre-kindergarten or other activities in the year prior to formal schooling ($S = -1$), or no schooling activities ($S = -2$):

$$S_{ct} = \max\{t - 5 + Z_c, -2\} \tag{3.2}$$

where Z_c is an indicator for whether the student's day of birth occurs before the school entry cutoff.

In this framework, at age $t = 5$ a student with $Z_c = 1$ will be enrolled in first grade ($S_{ct} = 1$), while a student born just after the cutoff will be enrolled in kindergarten ($S_{ct} = 0$). Under perfect compliance with entry policies, *all* students will be 5 years old (and some number of months) on the cutoff date in the year of kindergarten enrollment, but some will have a birthday between the start of school and the cutoff date. The latter are considered age $t = 4$, while the former, who turned 5 before starting kindergarten, are $t = 5$.

For now, ignoring the human capital stock term in θ_t (no stock-learning complementarity), equation (3.1) can be rewritten by iteration as:

$$\begin{aligned} h_{ct} &= \beta^{t-1} I_1(Y,-2) + \beta^{t-2} I_2(Y,-2) + \beta^{t-3} I_3(Y, S_{c3}) \\ &+ \sum_{k=0}^{t-4} \beta^k [I_{t-k}(Y, S_{c,t-k}) + \theta_{t-k}(S_{c,t-k})] \end{aligned} \tag{3.3}$$

Here I assume no initial human capital and no non-investment contribution of out-of-school activities: $h_{c0} = \theta_t(Y,-2) = \theta_t(Y,-1) = 0$. The first three investment terms represent the contribution of early investment and possible pre-school enrollment prior to entry into kindergarten. The summation gives the contribution of investment and learning obtained during and after a child's fourth year of age on human capital.

As discussed above, the prior literature on the effects of entry policies has tended to focus on two parameters: the difference in human capital between differently aged students enrolled in the same grade, and the difference in human capital between students in different grades measured when they are the same age. In the context of the model in equation (3.1), these two parameters, which I will refer to as the same grade (Δ_t^{SG}) and same age (Δ_t^{SA}) comparisons, can be expressed as follows:

$$\Delta_t^{SG} \equiv \lim_{a \to c} h_{a,t+1}(Y, Z_a = 0) - \lim_{b \to c} h_{bt}(Y, Z_b = 1) \tag{3.4}$$

$$\Delta_t^{SA} \equiv \lim_{b \to c} h_{bt}(Y, Z_b = 1) - \lim_{a \to c} h_{at}(Y, Z_a = 0) \tag{3.5}$$

The limits in equations (3.4) and (3.5) ensure that the comparison is between students born on either side of, but arbitrarily close to the school entry cutoff c. The same grade comparison, Δ_t^{SG}, is the difference in the stock of human capital at a given year in school between two students identical in every way except on which side of the entry cutoff their date of birth lies.[18] The same age comparison, Δ_t^{SA}, is the difference in the stock of human capital at a given age between two students identical in every way except on which side of the entry cutoff their date of birth lies. Conceptually, the key difference between the two parameters is the timing of the measurement of the stock of accumulated human capital. Both differences can be expressed as a function of the key parameters of the human capital model. For $t \geq 4$:

$$\Delta_t^{SG} = \sum_{k=0}^{t-4} \beta^k \{I_{t+1-k}(t-4-k) + \theta_{t+1-k}(t-4-k) - I_{t-k}(t-4-k) - \theta_{t-k}(t-4-k)\}$$
$$+ \beta^{t-3}(I_4(-1) - I_3(-1)) + \beta^{t-2} I_3(-2) + \beta^{t-2}(\beta - 1)(I_2(Y, -2) + \beta I_1(-2)) \tag{3.6}$$

[18]Equation (3.4) describes a comparison between students in different entering cohorts (i.e. they are not classroom peers) – the relatively older late-entering students from one entry cohort and the relatively younger early-entering students from the subsequent cohort – measured when they are in the same grade. One could estimate a slightly different parameter, as Elder and Lubotsky (2009) do, in which the comparison is between students in the same entering class: $\Delta_t^{SG'} \equiv E[h_{c-1,t+1}|Z_{c-1} = 0] - E[h_{ct}|Z_c = 1]$. This alternative comparison is attractive as a result of easier access to cross-sectional data which samples students by grade level than the panel data required to estimate (3.4). Under the assumption of no calendar year of birth effects, which would lead to bias in $\Delta_t^{SG'}$, and no school-year effects (such as year-to-year changes in school quality or curriculum), which would lead to bias in Δ_t^{SG}, the two are equivalent.

and,

$$\Delta_t^{SA} = \sum_{k=0}^{t-5} \beta^k \{I_{t-k}(t-4-k) + \theta_{t-k}(t-4-k) - I_{t-k}(t-5-k) - \theta_{t-k}(t-5-k)\}$$
$$+ \beta^{t-4}(I_4(0) + \theta_4(0) + \beta I_3(-1) - I_4(-1)) - \beta^{t-3} I_3(-2) \tag{3.7}$$

The summation term in both expressions captures differences in investment and learning obtained by students while enrolled in formal schooling. Both expressions also involve the $I_3(Y,-1)$ and $I_4(Y,-1)$ terms (the Y argument is dropped for economy), which represent the contribution (and likelihood) of pre-school enrollment at ages 3 and 4. The same grade comparison has additional terms representing the long-run effect of early childhood investment on human capital, while the age 1 and 2 investment terms drop out of the same age comparison. The early investment terms remain in the same grade comparison as a result of an extra year of discounting for the older students.

Prior work has obtained positive estimates of both Δ_t^{SG} and Δ_t^{SA} for elementary school students. In the same grade comparison, this result is often attributed to the benefit to a child of being relatively older than one's peers (i.e. a net positive summation term). Elder and Lubotsky (2009) show that $\Delta_t^{SG} > 0$ even at kindergarten entry ($t = 4$), suggesting the positive estimates are due to differences in the effect of early childhood investments (including normal childhood development) between the early and late-entering students, rather than benefitting more from formal schooling. Indeed, if late-entering students are more likely to participate in pre-school ($I_4(Y,-1) > I_3(Y,-1)$), this fact alone could account for the positive same grade comparison.[19]

In contrast, assuming $I_4(Y,-1) > I_3(Y,-1)$ works *against* a positive same age comparison. If Δ_t^{SG} is positive as a result of differences in pre-school investment, then in the context of the model, $\Delta_t^{SA} > 0$ implies positive effects of time spent engaged in formal schooling on human capital accumulation.

One way to make this discussion clearer is to assume that investment and learning while enrolled in school are identical across grade level and independent of age, so that $I_t(Y,g) = I(Y)$ and $\theta_t(g) = \theta$ for $t \geq 4$ and $g \geq 0$. This rules out relative age effects and treats all grade levels identically. I also assume no investment for children until the year prior to kindergarten. Then (3.6) and (3.7) reduce to:

$$\Delta_t^{SG} = \beta^{t-3}[I_4(Y,-1) - I_3(Y,-1)] \tag{3.8}$$

$$\Delta_t^{SA} = \beta^{t-4}[I(Y) + \theta] - \beta^{t-4}[I_4(Y,-1) - \beta I_3(Y,-1)] \tag{3.9}$$

[19]Here I implicitly assume that 3 and 4 year olds benefit identically from pre-school enrollment conditional on being enrolled.

Larger early childhood investments for late-entering children lead to positive Δ_t^{SG}, but ambiguously signed Δ_t^{SA}, since pre-kindergarten investments and the gains from an extra year in school work in opposite directions.

The simple model with no relative age effects can account not only for the signs of Δ_t^{SG} and Δ_t^{SA} but also for changes in the magnitude of the parameters with t. As students in a given entry cohort advance, the relative benefit to older students of additional pre-kindergarten investment will depreciate, and the magnitude of the same grade comparison will fall. It can be shown that:

$$\Delta_{t+j}^{SG} - \Delta_t^{SG} = (\beta^j - 1)\Delta_t^{SG} < 0 \tag{3.10}$$

$$\Delta_{t+j}^{SA} - \Delta_t^{SA} = (\beta^j - 1)\Delta_t^{SA} \tag{3.11}$$

The direction of changes in Δ_t^{SA} with age are ambiguous. The depreciation of the value of the additional schooling obtained by early-entering students relative to late-entering students will lead to a smaller same-age comparison for older students. However, as the lingering effects of differential pre-kindergarten investment depreciate, the same-age comparison will grow larger.

While relative age effects might be important for human capital accumulation of students while in school, they are not strictly necessary to explain commonly estimated signs of levels and changes in the same-age and same-grade comparisons. These patterns could be a result of differential early childhood investments and normal depreciation of human capital with age.

These results are not changed importantly by allowing for an explicit interaction between human capital acquired in a year of school θ and the student's prior stock of human capital, representing the "multiplier effect" on past learning discussed above (available upon request). However, if the complementarity is large enough, Δ_t^{SG} will grow with time in school and not fall as in equation 3.10.[20]

3.4 Data

I utilize the restricted-use National Assessment of Educational Progress Long-Term Trend (LTT), which includes nationally representative cross-sectional samples of students enrolled in public or private schools in the US at ages 9 and 13[21] between 1990 and 2004.[22] The LTT was designed to chart the academic achievement of children in the United States over time, and so includes assessments in mathematics that are

[20]Specifically, if γ is the marginal benefit of bringing additional human capital into a school-year, then Δ_t^{SG} will grow with time if $\gamma\theta > 1 - \beta$, the rate of human capital depreciation.

[21]The samples of 17 year olds are unsuitable for use in this analysis, since the LTT sampling frame for student date of birth coincides with the entry cutoff date for the majority of states; in the approximately 20 states (the actual number varies over time) with cutoff dates in September, I observe few, if any, 17 year-old students born on one side of the entry cutoff. Because the LTT is based on school-sampling, it does not include students who have dropped out of school. This potentially large selection problem is another reason to limit the analysis to the samples of 9 and 13 year olds, for which dropout is unlikely to be a problem.

[22]The LTT surveys extend back to 1971, though surveys prior to 1990 are not suitable for use in the analysis since they do not record student year and month of birth and/or state of residence.

comparable across time.[23] The math scores are reported as item response theory (IRT) scaled scores.[24] The surveys include students' month of birth and current state of residence.[25]

Unlike most school administrative data and many other surveys used to identify the effects of entry policies (such as the ECLS-K), the LTT samples students based on their age regardless of current grade of enrollment. Age-based sampling reduces selection concerns important for an analysis of the effects of school entry policies, which "treat" students based on their age and not on the basis of enrolled grade. The age-based sampling frame ensures that I observe a random sample of students born on a given day regardless of the grade of enrollment and also that an identical assessment is given to students regardless of grade of enrollment.[26]

Across the six survey years in the available sample (1990, 1992, 1994, 1996, 1999, and 2004), I observe 71,062 students, though I impose several sample restrictions, described in Appendix F. Most of the sample restrictions are a result of dropping state-years for which insufficient numbers of students are observed with birth dates

[23]The LTT includes a separate set of cross-sectional samples with reading scores, but these do not include the question about pre-kindergarten participation in the student survey.

[24]The IRT scores are reported as a set of four "plausible value" estimates of each student's math or reading ability. I average the four values to obtain a single score for each student.

[25]While the LTT does include current state of residence in most survey cross-sections, I do not observe the state of residence at the time of school entry. Using state of residence as a proxy for entry date will lead to some measurement error in assigning entry cutoff dates to students.

[26]The survey uses a matrix sampling design so that not all children take identical questions, but the assignment of test booklets is random. For more details on the sampling and survey design, see Rogers and Stoeckel (2007).

on either side of the entry cutoff.

The final sample includes 116 (108) state-years for 9 (13) year olds, representing 18,961 (17,940) students. In each sample, 30 states remain, though only four of these are observed in every year. The 30 states are distributed across the midwestern, southern, and western regions of the US. In the northeastern region, only Deleware in 1999 for 9 year olds remains. This limitation of the data should be considered in assessing the external validity of the results in this paper. Appendix Table F.1 tabulates the school entry cutoffs for states in the final sample. Just under 3/4 of students in the analysis sample live in state-years with an early September or late August cutoff date. Appendix Table F.2 describes the state-years that remain in the final samples of 9 and 13 year olds.

Table 3.1 gives summary statistics of the final analysis samples for 9 and 13 year olds. All student characteristics are based on student self-reports. Just under half of students are boys, 15-18 percent of students are black, 9 percent are Hispanic and 70-73 percent are white. School characteristics are based on administrator surveys. Students on average attend schools with a racial makeup similar to the sample population. More than 80 percent of students attend schools with no bilingual students, and more than 60 percent of students attend schools with no English as a Second Language (ESL) students. About half of students attend schools with 11-50 percent free lunch eligible students, but significant numbers of students attend schools composed of both very few and nearly all free lunch eligible students. About 3/4 of students report attending some kind of pre-kindergarten program, one of the key outcomes for the analysis. This is in contrast to kindergarten attendance, which is near universal.

The student and school demographics descriptives reported in the table constitute the control variables I include in the analysis described below. The sample size for some questions is smaller as a result of non-response. When variables with missing values are included in the analysis, missing values are coded as 0 and a separate indicator for missing is included.

Panel A of Table 3.2 gives the distribution of grade level of enrollment for 9 and 13 year olds by assigned entry cohort along with the mean math score for each cell. 81 percent of students assigned early entry are in the 4th grade at age 9, and 86 percent of students assigned late entry are in the 3rd grade at age 9, indicating a high degree of compliance with school entry policies. Panel B of Table 3.2 restricts the sample to only students born within one month of the entry cutoff. Non-compliance for these students is significantly higher.

The distribution of 13 year olds across grades is similar, with slightly smaller fractions of students in their assigned grade level (8th grade for assigned early-entering students and 7th grade for assigned late entry), possibly as a result of grade retention between age 9 and age 13.[27]

The non-compliers for students assigned early entry tend to be behind grade, consistent with redshirting, while 12 percent of 9 year olds assigned late entry are actually above grade, indicating early entry. Thus, many students are actually enrolled in school a year ahead of the earliest entry date set by the state's entry policy.[28] An additional 3 percent of students assigned late entry delay entry an additional year or are retained in grade prior to age 9. As one might expect, students in higher grade levels tend to perform better on the mathematics assessment, though one cannot separate the role of selection into grade level and the effect of grade level itself in the differences in test scores between grades (to say nothing of possible age effects, differences in out of school experiences, etc.).

[27]Given trends over time in the incidence of redshirting (Deming and Dynarski, 2008), some of the difference between 9 and 13 year old compliance rates could be due to year effects.

[28]Some of the apparent non-compliance could be a result of measurement error in state of residence.

3.5 Empirical Strategy

3.5.1 Primary estimating equation – same age, different grade

To estimate Δ_t^{SA}, the reduced form effect of entering school a year early, measured at a given age, I implement a regression discontinuity (RD) design. I estimate the following reduced form RD equation separately for 9 and 13 year olds:

$$Y_{ijstb} = \beta_0 + \beta_1 T_{ist} + f(m_{ist}) + g(m_{ist})T_{ist} + X_i\theta_1 + W_j\theta_2 + \mu_{st} + \nu_{tb} + \epsilon_{ijstb} \quad (3.12)$$

where i, j, s, t, and b index student, school, state, year, and test booklet. T_{ist} is an indicator for assigned early entry – whether student i's birthdate falls ahead of the school entry cutoff in state s in the year of the student's first opportunity to enter school given age in year t. The running variable in the RD is m_{ist}, student i's month of birth relative to the relevant entry cutoff. For example, if the entry cutoff is at the start of September, then a student born in July is assigned $m = -2$ (September birthdays are assigned $m = 0$). To flexibly account for the relationship between the running variable and the outcome of interest, the preferred specifications implement the functions $f()$ and $g()$ as quadratics in relative month of birth.[29]

The vectors X_i and W_j represent student demographics and school-level observ-

[29] Because in most states the entry cutoff is at the start of September, leaving only four months on the assigned late entry ($T = 0$) side of the cutoff, and because I only observe students' month of birth, not exact day, I am limited in how flexibly I can implement these functions.

ables included as control variables. If the key identifying assumption for the RD design – that students born near the school entry cutoff date are identical in all respects except in their assignment to early or late entry into school – is satisfied, then control variables are not strictly necessary. I include X_i and W_j in some specifications to illustrate the insensitivity of the RD estimate to their inclusion and also to reduce the variance of the model residual, producing smaller standard error estimates. For the same reasons, in some models I also include state-year fixed effects μ_{st} and test booklet-year fixed effects ν_{tb}. Even in specification with no control variables, I include survey year fixed effects.

The key discontinuity parameter is the estimate of β_1, the coefficient on the assigned early entry indicator. $\hat{\beta}_1$ gives the estimated reduced form effect of assigned early entry relative to assigned late entry for students born near the cutoff, when they are the same age. Under perfect compliance with the entry laws, this provides an estimate of Δ_t^{SA}, but as discussed above, significant non-compliance is present in the form of redshirting, school promotion policy, and students who enter school ahead of their assigned entry. Thus, for some purposes, the reduced form discontinuity estimate must be scaled to account for the frequency of non-compliance. To estimate the effect of being induced by entry laws to be enrolled a full grade level G ahead at age 9 or 13 on a student outcome Y, I implement a fuzzy RD using two-stage least squares (2SLS) estimation with first stage (FS) and structural equations (SM) as follows:

$$\begin{aligned}(FS): G_{ijstb} &= \gamma_0 + \gamma_1 T_{ist} + f^{FS}(m_{ist}) + g^{FS}(m_{ist})T_{ist} \\ &+ X_i\Gamma_1 + W_j\Gamma_2 + \mu^{FS}_{st} + \nu^{FS}_{tb} + \phi_{ijstb}\end{aligned} \tag{3.13}$$

$$\begin{aligned}(SM): Y_{ijstb} &= \pi_0 + \pi_1 G_{ist} + f^{SM}(m_{ist}) + g^{SM}(m_{ist})T_{ist} \\ &+ X_i\Pi_1 + W_j\Pi_2 + \mu^{SM}_{st} + \nu^{SM}_{tb} + \eta_{ijstb}\end{aligned} \tag{3.14}$$

The estimate of γ_1 gives the effect of assigned early entry on grade attainment of students, which will be unity under perfect compliance. The estimate of π_1 gives the effect of entering formal school early scaled up to a full grade level.[30]

It is not entirely clear if the discontinuity estimate given by the estimation of equations (3.13) and (3.14) is the appropriate scaling. For example, one might consider promotion and retention rates, unlike redshirting and early entry, to be part of the effect of assigned early entry. If so, these factors should not be included in the first stage correction for non-compliance, as they implicitly are in equation (3.13). Put another way, one parameter of interest is the effect of entering school a full year early, rather than the effect of being enrolled one grade level higher. If retention and promotion rates do differ across the cutoff (shown to be the case for Chilean children

[30]The superscripts SM and FS indicate that the fixed effect and functional form estimates in these equations are allowed to differ between the structural and first stage equations, and from those in the reduced form regression in equation (3.12).

in McEwan and Shapiro (2008)), then the first stage discontinuity estimate of $\hat{\gamma}_1$ obtained using the grade-level outcome is likely smaller than what one would obtain using years since school entry (as a result of higher retention rates in the assigned early entry cohort). Using grade level instead of years since entry as the first stage outcome would tend to bias the 2SLS estimate of π_1 upward.[31]

Because the entry policies affect pre-kindergarten experiences as well as the time of entry into kindergarten, I consider an alternate scaling of the reduced form effect of early entry on student outcomes at a given age, based on years of completed schooling, rather than grade level. Replacing G in equations (3.13) and (3.14) with S, the sum of enrolled grade level, an indicator for kindergarten attendance, and an indicator for pre-kindergarten attendance, produces an estimate of π_1 that accounts for differential early childhood participation in schooling.[32] However, scaling by completed schooling is also subject to criticism regarding treatment of retention and promotion rates.

3.5.2 Primary estimating equation – same grade, different age

Using the students on only one side of the cutoff, I implement instrumental variables (IV) techniques similar to those used in Datar (2006), Bedard and Dhuey (2006),

[31] Because I observe grade level, but not years in school, I cannot test for differences in retention and promotion rates using the LTT data.

[32] This specification implicitly treats all grade levels equivalently, so that a student who completes pre-kindergarten through grade 2 and a student who completes kindergarten through grade 3 are considered to have an equivalent schooling histories.

and others to estimate Δ_t^{SG}, the effect of entering school a year late when measured at a given number of years since the assigned year of school entry.

Non-compliers with entry policies are treated differently in a grade-based sample than in an age-based sample. Previous work, using data with grade-based sampling, has included all students enrolled in a given grade regardless of their assigned grade given by the entry policy. The oldest students in a grade-based sample are those who were redshirted (or retained). Redshirted students in the grade-based samples will be a full year older than compliers with entry policies born on the same day. In contrast, in an age-based sample, redshirted students are not relatively older than others in the sample, but are instead observed when they are in an earlier grade level. Only the compliers with entry policies are treated identically in the two sampling frames.

In the grade-based sample, non-compliers (ignoring students in non-compliance from entering school ahead of their assigned entry) will outperform the non-compliers from an age-based sample, who are a full year younger. Because students assigned to enter school when younger are more likely to be redshirted, the differences in the age of measurement for non-compliers will lead to a smaller relationship between test scores and assigned age of entry in a grade-based sample than in an age-based sample. Conceptually, the age-based sample holds assigned grade fixed for all students, whereas the grade-based sample fixes enrolled grade. Using data with an age-based sampling frame, I estimate the same assigned grade, different age comparison, and also, by de-meaning the math score outcome by grade level, the same enrolled grade, different age comparison.

Prior work has used an instrumental variables approach to estimating Δ_t^{SG}, with assigned entry age based on state school entry policies instrumenting for actual age of entry:[33]

$$(FS): E_{ijstb} = \gamma_0 + \gamma_1 A_{ist} + X_i\Gamma_1 + W_j\Gamma_2 + \mu_{st}^{FS} + \nu_{tb}^{FS} + \phi_{ijstb} \tag{3.15}$$

$$(SM): Y_{ijstb} = \pi_0 + \pi_1 E_{ist} + X_i\Pi_1 + W_j\Pi_2 + \mu_{st}^{SM} + \nu_{tb}^{SM} + \eta_{ijstb} \tag{3.16}$$

where E gives the age of entry into kindergarten in months, and A gives the assigned age of entry in months based on state entry cutoffs. A is not directly analogous to the relative month of birth m used in the RD estimation of same age, different grade effects, since students start school at different ages not only because of entry policies, but also due to natural variation in month of birth. For example, a student born in July in a September entry cutoff state will take on values of $m = -2$ and $A = 62$, or five years and two months (assuming the school year begins in September). However,

[33]Here I recycle greek letters from the previous subsection for ease of interpretation, thus the reader should infer no equivalencies.

if the state cutoff were instead in October, then $m = -3$, while the value of A for the student is unchanged (again assuming the school year begins in September).

I do not observe the exact age of entry E into kindergarten for students, which complicates the estimation of equations (3.15) and (3.16). I address this two ways. First, I can impute age of entry based on current grade of enrollment and ignoring the possibility that a students' current grade level may be affected by past retention decisions. Ignoring retention carries similar concerns to those discussed in relation to the scaling of the RD parameter above. A second solution is to limit focus to the reduced form equation, which is identified even without knowledge of the actual age of entry for students:

$$Y_{ijstb} = \beta_0 + \beta_1 A_{ist} + X_i\theta_1 + W_j\theta_2 + \mu_{st} + \nu_{tb} + \eta_{ijstb} \tag{3.17}$$

The reduced form results give the intent-to-treat (ITT) effect of entry policies. While the ITT effect is a well-defined policy parameter, it is not informative of the effects of entering a full year earlier on student outcomes. Though the ITT effect is not directly related to the structural parameters discussed in the theory section, the ITT and IV effects are identical in the case of perfect compliance; otherwise, the ITT parameter given by the reduced form estimate from equation (3.17) is a lower bound for π_1, the effect of entering school one full year early on Y.

Because the effects of age are potentially non-linear, my preferred specifications allow quadratic functional forms in assigned age A and actual age of entry E. As in the RD estimation, I am limited in the flexibility I can allow in the estimation the regressions in this section. In data sampled on age, the entry cutoff date divides students into two assigned entry cohorts. To hold assigned grade fixed, I focus on students on just one side of the cutoff. Because the LTT samples only 9 year olds and not 8 or 10 year olds, I do not observe students born a full year apart on one side of the cutoff. For example, if the cutoff is September 1st, the LTT only includes the students born in January through August in the school entry cohort, providing just 8 months of support.[34]

3.6 Results

3.6.1 First stage and reduced form same age, different grade effects of entry policies

First, I examine the first stage effects of state school entry policies on grade level and time in school at ages 9 and 13. Figures 3.2a and 3.2c plot the mean grade level of 9 and 13 year-old students by month of birth relative to the state school entry cutoff

[34]Alternatively, given that I observe students who make the cutoff in school year y, I would need to observe students who miss the cutoff in school year $y + 1$ (and not y at the time of observation in the LTT), so that students on both sides of the cutoff are observed in the same assigned grade.

faced by each student when entering kindergarten.[35] The grade level of students born before the cutoff, those assigned early entry into school, is higher than those born after the cutoff. The magnitude of the difference between students on either side born within a month of the cutoff date is well below a full grade level, indicating significant non-compliance with the entry laws (and/or differences in retention rates).

In addition to the plot of sample means, the solid lines in Figure 3.2 show the results of estimation of the first stage regression equation (3.13) with $f^{FS}()$ and $g^{FS}()$ implemented as quadratics and including only survey year fixed effects as control variables. The dashed lines show 95 percent confidence intervals (around the relative month of birth means) based on standard errors robust to clustering at the state-year-relative month of birth level. The discontinuity estimates based on this specification for 9 and 13 year olds are reported in columns (1) and (3) of Table 3.3. Students born just before the entry cutoff are induced by state policies to be enrolled 0.40 grades higher at age 9 and 0.36 grades higher at age 13 relative to those born just after the cutoff.[36]

Because the first stage results in Table 3.3 rely on the quadratic functional form assumption, I also include the p-value from an F-test of the quadratic restriction

[35]The means are adjusted for survey year fixed effects but include no other controls.

[36]The smaller effect for 13 year olds, while not statistically significant, is possibly a result of additional grade retention between age 9 and 13.

against the saturated model (group means by relative month of birth) in brackets. Given the apparent success of the fit seen in the figure, it is not surprising that the p-values indicate a failure to reject at conventional confidence levels. The estimates in columns (2) and (4) include the full set of additional control variables, which improves precision, but does not importantly affect the discontinuity point estimates.

Because entry policies also affect the experiences of children prior to formal school enrollment (see Figure 3.1), I examine effects of entry policies on an alternative measure of educational attainment that accounts for differences in pre-kindergarten enrollment rates across the cutoff. Figures 3.2b and 3.2d present results analogous to 3.2a and 3.2c, but with an outcome that counts the total years of schooling completed by the students, including reported pre-kindergarten enrollment.

Students assigned late entry are more likely to attend pre-kindergarten, which works against the magnitude of the grade-level first stage equation. While the results in Figures 3.2b and 3.2d do not appear much different than those in 3.2a and 3.2c (apart from a shift up in the y-intercept from counting kindergarten and pre-kindergarten as years of schooling), the second row of Table 3.3 reports discontinuity point estimates which are between 9 and 18 percent smaller than the estimated effect on enrolled grade level.[37] The smaller, but not zero, first stage effects that account

[37]The sample size is reduced in the years of schooling specifications relative to the grade-level

for pre-kindergarten enrollment indicate that only a fraction of students induced to enter school late by the policy compensate by attending pre-kindergarten.

In Figure 3.3 and in the last row of Table 3.3, I present estimates of the reduced form effects of entry policies on standardized test scores. Figure 3.3a shows a generally downward sloping trend in math scores for 9 year olds with a discontinuity at the entry cutoff. The downward slope is consistent with positive age effects on math scores, though compliance rates are also changing with relative month of birth m. I explore this relationship in more detail when considering the same grade, different age comparison results, below. The downward trend for 13 year olds is less pronounced (the IRT scaling is designed to make the 9 and 13 year-old scores comparable on the same scale), and a discontinuity is not apparent.

Results in Table 3.3 show a 7.3 point (1/4 standard deviation) effect of the entry policies on math scores at age 9, but a less than one point (0.03 standard deviation) effect at age 13, and not statistically significantly different from 0. Indeed, with 95 percent confidence, I can rule out effects on math scores at age 13 larger than 3.1 points (0.11) standard deviations. Again, the p-values from tests of the quadratic restriction indicate no cause for concern about the functional form assumptions.

specifications as a result of non-response to the questions of pre-kindergarten and kindergarten enrollment.

3.6.2 Scaled effects of schooling on test scores

As discussed above, to adjust for non-compliance rates, the reduced form effects of assigned early entry into school on math scores can be scaled in a number of ways. Because entry policies affect student experiences both before and after entering kindergarten, I scale the effects of entry policies by current grade level of enrollment, which ignores early childhood experiences, and years of completed schooling, including pre-kindergarten participation.

Table 3.4 shows the results of 2SLS estimation of equations (3.13) and (3.14). For comparison, in the first row, I also present the results of an OLS regression of test scores on grade level and the same set of controls as in the 2SLS regression, but without including regressors specific to the fuzzy RD design. Age 9 students who are enrolled in 4th grade perform, on average, 30 points (a full standard deviation) higher than those enrolled in 3rd grade. The analogous comparison for 13 year olds in 8th versus 7th grade is 19 points (0.7 standard deviations). These cross-sectional grade-level comparisons are likely to overstate the causal effect of participating in an additional year of schooling on test scores because the students in higher grades are also relatively older, and some may have been enrolled in school earlier as a result of having higher ability or parents who are inclined to invest heavily in schooling.

The 2SLS estimate of the effect of a grade level in school is 18 points (0.6 standard deviations) for 9 year olds and indistinguishable from no effect (I can reject effects larger than 0.3 standard deviations) for 13 year olds. The interpretation of the 2SLS effect on math scores scaled by grade level as the effect of a year in school ignores differential retention rates, pre-kindergarten experiences, and differences in the age of students relative to peers in their entry cohort across the cutoff. Under the assumption that more early-entering (young-for-grade) students are retained than late-entering students, the parameter scaled by the grade-level first stage is an upward-biased estimate of the effect of time in school. However, effects of entry policies on pre-kindergarten attendance and relative age effects work in the opposite direction and suggest that the 2SLS estimates in the second row are actually lower bounds for the effect of time in school. Even so, the benefits of time in school can account for at least half of the cross-sectional difference in math scores by grade level for age 9 students.

The alternative scaling, shown in the third row of Table 3.4, adjusts for the extra year of pre-school obtained by more late-entering students than early-entering students. In this framework, a 9 year old who attended pre-kindergarten through 3rd grade is considered to have been exposed to the same amount of schooling as a student who never attended pre-kindergarten and is enrolled in 4th grade. This scaling gives larger estimates of the effect of a year in school: about 0.8 standard deviations for 9 year olds and an upper bound of 0.4 standard deviations for 13 year olds (but still not significantly different from zero).

The decline with age in the effects of time in school is consistent with the model presented above, in which the early human capital gains from the first additional year in school (whether pre-kindergarten or kindergarten) depreciate over time. One alternative explanation, however, is that students simply learn less in 8th grade relative to 7th grade than in 4th relative to 3rd grade. Later, I attempt to separate the effects of grade level from pre-kindergaten participation. Regardless of which story is responsible for the decline in the effect of a year in school with age, that the estimates accounting for early childhood experiences are larger imply that the effects of early schooling investments are important for understanding the human capital stock of students of all ages. Also, though the effects of additional early schooling on test scores seem to depreciate quickly, this does not necessarily mean that its effects on other outcomes do not persist. For example, Chetty et al. (2010) and Dynarski et al. (2011) find persistent effects of kindergarten classroom quality on long-run educational and earnings outcomes, even though effects on math scores fade quickly.

The RD estimates of the effects of time in school on test scores eliminate bias from selection into early entry due to differences across the cutoff in the underlying propensity of families to invest in schooling. However, to the extent that early entry directly affects investment, the results in Table 3.4 may reflect differential investment and effort, and not only an extra year in school. I explore this further in Appendix G. The results in Appendix Table G.1, which suggest that early entry negatively affects investment while in school, give further reason to consider the 2SLS RD estimates as a lower bound for the effect of time in school on math scores.

Finally, as a test of the RD research design, I test for discontinuities in student and school demographics across the cutoff. Table 3.5 shows no evidence of discontinuities in these outcomes for 9 year olds. However, the tests for 13 year olds reveal statistically significant discontinuities in student and school characteristics, though the magnitude of the differences are small.

3.6.3 Reduced form same grade, different age effects of entry policies

While time in school accounts for much of the growth in test scores between grade levels, students' age might account for some of the growth. In Table 3.6, I explore the effects of being assigned to enter school at relatively younger age on math scores measured at a given number of years since the year of assigned entry. Columns (1) and (2) give estimates of equation (3.17) with and without control variables. The effect of being assigned entry into school when one year older on math scores five years after assigned entry (so most students are in 4th grade) is about 17 points (0.6 standard deviations). This effect falls to about 9 points (0.3 standard deviations) by the end of the ninth year after assigned entry.

Figure 3.4 shows the relationship between assigned entry age and math scores graphically. As discussed above, I only include students born before the relevant state school entry cutoff date in the analysis, limiting the support of assigned entry age to students of between 5 years and 5 years and 8 months (5.67 years) of age. Figures 3.4a and 3.4b suggest that the linear restrictions in columns (1) and (2) in Table 3.6 are mis-specified. Column (3) of Table 3.6 presents the results of the quadratic specification also shown by the solid line in the figures. Calculating the marginal effect at two points near the extremes of the support of entry age, I find much larger marginal effects of assigned entry age on math scores for younger students than for older students.

The marginal effect on math scores five years after assigned entry of an additional *month* of assigned entry age for a student assigned to enter at 5.1 years of age is about 2 points (0.07 standard deviations), while an additional month for a student assigned to enter at 5.6 years of age only provides effects about half as large. The estimates presented in the table represent the effects of a full year of age, but given the interpretation as marginal effects, scaling back to a month is appropriate.[38]

Because of non-compliance with entry policies (and because of grade retention) not all students included in the regressions in columns (1)–(3) are actually in the same grade, despite being assigned to the same grade (4th grade or 8th grade) by entry policies. If the effects of assigned age of entry represent actual effects of maturity or relative age, and not differences by grade level, then the effects should be present even when the comparison is between students actually enrolled in the same grade. In column (4), I demean the math score outcome by grade level (using all students in the full analysis sample, on both sides of the cutoff) prior to running the regression, so that behind-grade students are not penalized. The results are similar to those in column (3) at the older 5.6 year margin, but show no effects of increasing the age of assigned entry at the 5.1 year margin. Figure 3.5 shows results using the demeaned outcome graphically.

The smaller effects at the young end of the distribution in the demeaned outcome specification are not surprising given the higher proportion of these students who are below grade. An effect of assigned entry age on the demeaned outcome close to zero

[38]The fact that the identifying variation in the IV procedure approximates the experimental manipulation of age of entry locally also demands caution in the interpretation of the effects as a full year of age. The IV manipulation is in contrast to a RD design, which I cannot perform due to data limitations, which would effectively manipulate age of entry by a full year of age.

for the youngest students could reflect a very low rate of compliance with the law (i.e. frequent redshirting), no relative age or maturity effects, or some combination of the two. (I adjust for non-compliance below.) At least for the students at the upper age margin, additional time before starting school leads to math score gains, which cannot be explained away by grade-level effects. Even for students at the higher age margin, after nine years in school, the effects of an older assigned entry age are no longer statistically significantly different from zero, though I cannot reject effects as large as 0.9 points (0.03 standard deviations) for one additional month of time prior to beginning school. The magnitudes of these estimates are similar to those reported in Elder and Lubotsky (2009).

To sum up the results in this section, assigning students to entry at an older age produces large test score gains, especially for the youngest of potential entering students. However, much or all of this effect is due simply to the fact that marginally older students are less likely to be redshirted (and thus have completed more school t years after assigned entry), and not due to any direct benefits of age. However, the test score gains of older students due to additional time out of school are consistent with maturity or relative age effects. As with the effects of additional time in school, the benefits of "the gift of time" fade quickly as the student progresses through school.

3.6.4 Scaled effects of entry age on math scores

Though I do not directly observe the age at which students enter kindergarten, I can impute entry age based on current grade of enrollment. To do so, I assume no retention and that all students attend kindergarten. Then, I can instrument for students' imputed actual entry ages with assigned entry age based on month of birth and the relevant school entry cutoff to estimate the effect of entering school a full year (or month) later on math scores at five and nine years after assigned entry. This IV procedure does not address differences in the early childhood experiences of children, and to the extent that older-entering children are more likely to participate in pre-kindergarten, it will produce upward-biased estimates of the effects of age. I ignore this issue, temporarily, to present results comparable to prior work on the same grade, different age comparison.

Table 3.7 shows the results of the first stage and 2SLS estimation using assigned age of entry as an instrument for actual entry age. The first stage estimates in column (1) indicate that on average, a student assigned to enter school one year early tends to enter on average just 0.6 years early. This estimate is misleading, however, since as with the reduced form, the effects of assigned age on age of entry are non-linear. The quadratic specification in column (2) indicates that entry policies have a near one-for-one effect on actual entry age for students assigned to enter at 5.6 years of age, while students assigned entry at an earlier age tend to enter school fewer than three months early in response to early assigned entry by a full year. Figure 3.6 plots the first stage relationship, showing the pattern of redshirting for students assigned early entry (a flatter relationship), who are much more likely to enter school after their assigned age.[39]

Column (3) of Table 3.7 shows the 2SLS estimate of the effect of entry into school when one year older on math scores. Five years after assigned entry into school, students who entered a full year older, perform 29 points better (one standard deviation). After nine years, the effect is reduced by about half. Given the non-linearities in both the first stage and reduced form, I estimate a more flexible specification with quadratics in both the instrument and the imputed entry age measure. The results of this specification in column (4) show very large entry-age effects for the students assigned entry at the youngest margin and smaller effects at the older margin. As with the reduced form estimates in Table 3.6, the large estimate at the lower margin has less to do with age than with the grade level of enrollment for these children: the 52-point effect at the 5.1 year margin reflects a smaller probability of being redshirted, and thus a higher grade of enrollment after five years, given a marginal increase in

[39]Some of what I attribute to redshirting here could be grade retention.

the age of assigned entry.

As in the reduced form results, I estimate a final specification using a math score outcome demeaned by grade level prior to estimating the model. This specification essentially eliminates any effects of grade level on math scores, giving instead the effects of age of entry in a hypothetical world in which all students are enrolled in the same grade. The linear specification in column (5) indicate benefits of a month of age of 0.7 points (0.02 standard deviations) five years after assigned entry, falling to about half of that by the ninth year. The quadratic specification in column (6) does not have the power to distinguish differences in the magnitude of effects between the two assigned age margins.

Given the evidence of non-linearities in the effects of age, one should be cautious about extrapolating these results beyond marginal changes in entry age to the effects of a full year of age. Even if one assumes the effects of age do scale linearly (almost certainly leading to an overestimate), then five years after assigned entry a year of delayed entry provides about 8 points on the math assessment, about half the 18-24 point effect of a year of formal instruction reported in Table 3.4. Together, these results suggest that parents who redshirt obtain a double windfall for their children's test scores, as the child is more likely to attend pre-kindergarten (additional schooling), and reaping the (smaller) benefits of age. However, such arithmetic double-counts any benefits of pre-school enrollment: it is possible that the *only* reason age appears to have a direct effect on math scores is precisely because students who enter when older have richer out of school experiences.

3.6.5 Can pre-kindergarten enrollment account for the effects of entry policies?

While the prior literature exploiting variation from entry laws has often worked under the assumption that the policy instrument works only through grade or age, this need not be the case, which importantly affects the interpretation of estimates of causal effects.[40] Taking an approach at the opposite extreme from the prior literature, I ask, if inducing pre-kindergarten enrollment were the only pathway through which school entry policies affected student test scores, how large is the effect of pre-kindergarten enrollment on test scores implied by the reduced form estimates of entry policies?

Large effects of pre-kindergarten enrollment on test scores would mean that the estimates of the effects of time in school in the second row of Table 3.4 are too small (the "years of completed schooling" scaling adjusts for pre-kindergarten attendance,

[40] Bound et al. (1995) make this point in discussion of Angrist and Krueger (1991) and Angrist and Krueger (1992).

but assumes pre-kindergarten and all other grade levels have identical effects on math scores), and that the estimates of the effects of age in Table 3.7 are too large. In Table 3.8, I estimate the effects of entry policies on math scores under the assumption that the entire effect works through pre-kindergarten enrollment.

Columns (1) and (2) give the first stage effects of entry policies on pre-kindergarten enrollment. The regression discontinuity results indicate that assignment to late entry increases pre-kindergarten enrollment rates by 6.8 percentage points. Using only students in the assigned early-entry cohort I estimate an equation analogous to the same grade, different age specification in equation (3.17), and find a nearly identical effect of 6.2 percentage points for a year of assigned entry age. The similarity of these results lends credibility to the two main identification strategies used in the analysis above and also simplifies the interpretation of the remaining results in Table 3.8.

I present 2SLS estimates of versions of equations (3.15) and (3.16) with the first stage outcome replaced with pre-kindergarten attendance in columns (3) and (4). The implied effect of pre-kindergarten attendance is an implausible 265 points (nearly 9 standard deviations) on the math assessment at age 9, five years after assigned entry. However, this estimate does not adjust for differences in grade level due to redshirting, which is correlated with the assigned age of entry instrument. I do not present similar results for the RD specification because the implied effects of pre-kindergarten attendance on test scores measured at a given age are negative; the effects of entry policies on pre-kindergarten attendance and grade level work in opposite directions, thus pre-kindergarten participation cannot account for the effects of grade level on math scores.

To net out the effects of grade level, I replace the equation outcome with the grade-level demeaned math score, as in column (4) of Table 3.6. Thus, in the assigned entry age IV, and assigned grade level cutoff RD results shown in columns (5) and (6), students are not penalized for being a grade level behind their assigned grade. For the RD comparison, netting out grade level leaves differences in relative age and early investment between students on either side of the cutoff. To account for variation in the demeaned math score outcome across the entry cutoff and with changes in assigned entry age, the effect of pre-kindergarten would need to be about 67 points (about two and a half standard deviations). The implied effect is nearly identical across the two research designs, though the estimates are imprecise.[41]

The 2.6 standard deviation point estimate is much larger than prior estimates of the effects of pre-school participation on cognitive outcomes, including Gormley, Jr.

[41] I do not present results for 13 year-old math scores due to very weak first stage effects of entry policies on reported pre-kindergarten attendance.

and Gayer (2005), who found effects on the order of 1/3 of a standard deviation prior to kindergarten entry. One explanation for the large effects in Table 3.8 could be downward bias in the first stage due to mis-reporting of pre-kindergarten attendance by students. Taken at face value, however, the implausibility of the magnitude of the estimates of the effects of pre-kindergarten enrollment on test scores implies that a single channel cannot account for all of the test score effects generated by the entry policies. However, one would be equally in error to ignore what could still be a large role for pre-school experiences in the evolution of human capital while in school.

3.6.6 Heterogeneity

Table 3.9 presents the results of the age 9 RD first stage and reduced form estimation by student subgroups. Column (1) gives the baseline for all students, and columns (2) and (3) split the sample by public and private students. All of the effects of entry policies are stronger for public students than for private students, which is expected given that private schools need not generally impose the same entry cutoff as the state's public schools. Columns (4) and (5) present results by gender. Boys tend to be redshirted more often than girls, and are also more likely than girls to use the extra assigned year out of school to attend pre-kindergarten. The smaller reduced form estimate for boys could be a result of a smaller difference in enrolled grade level across the cutoff than for girls, or the larger difference in pre-kindergarten participation. The remaining columns present results by student race. White students are less likely to comply and more likely to attend pre-kindergarten when assigned an extra year out of school than black students. As with the gender comparison, the white students also have a smaller reduced form estimate. Given the smaller sample sizes when looking at subgroups, the standard errors tend to be much larger than in the full sample, so care should be taken in comparing point estimates across the columns of Table 3.9.

Table 3.10 gives the analogous results for the same grade, different age estimation. Compliance with the entry laws follows a similar pattern in these results as in the RD, but the quadratic specification of the IV allows a look at difference in the compliance rate for students assigned to enter at 5.1 versus 5.6 years of age. Public school students, girls, and black students have the highest rates of compliance at the younger margin, though public school students and girls are more likely to enter school early when assigned to enter at 5.6 years of age. Compliance rates are much higher at the older margin (except for students in the "other race" category). The point estimates in the first stage effects of entry policies on pre-kindergarten participation show less variance across groups than in the RD specification, but black students are still the last likely to use extra time out of school for pre-kindergarten attendance, and boys and white students are among the most likely.

I report the same grade, different age reduced form estimates by subgroup at the bottom of Table 3.10. The reduced form point estimates vary across subgroup, but high variance in the estimates make interpretation with the naked eye difficult.

3.6.7 Separate identification of the effects of age, pre-kindergarten, and time in school on math scores

Because the entry policies affect student outcomes through several potential mechanisms yet provide just one valid instrument,[42] plausibly separating the effects of age, time in school, and early childhood experiences on test scores is challenging. Such an effort requires further assumptions (or additional instruments). In this section, I exploit variation across states and over time in the first stage responses to entry policies – age, grade in school, and pre-kindergarten participation – to explain variation in the reduced form effects of entry policies.

Using cross-state and cross-time variation to separately identify channels through which entry policies produce effects on math scores requires strong assumptions about unobserved determinants of student response to entry policies. For identification, I

[42] Some work has attempted to exploit differences across states in the date of the entry cutoff or in the average age of students' peers in an entry cohort as an additional instrument.

assume that unobserved characteristics of states and years that lead to variation across subgroups in compliance and response (including pre-kindergarten enrollment) are uncorrelated with the error term in the structural model. Thus, any reasons for differences in compliance across states and years are not also determining outcomes. For example, if students in states with excellent pre-school programs are less likely to comply with entry policies (e.g. more redshirting), and if participation in these programs is positively related to math scores, then estimates of the effect of entry age on test scores based on cross-state variation will be biased downwards (and effects of pre-kindergarten participation on math scores will be biased upwards).

I estimate the primary RD and IV specifications for the effects of entry policies on math scores for student subgroups. Specifically, for each state-year cell, I obtain estimates of the first stage RD and IV parameters $\hat{\gamma}_{st}^{G}$, $\hat{\gamma}_{st}^{P(SA)}$, $\hat{\gamma}_{st}^{E}$, $\hat{\gamma}_{st}^{P(SG)}$, grade level and pre-kindergarten for RD, and entry age and pre-kindergarten for IV, respectively. I also estimate the math score reduced forms for RD and IV, $\hat{RF}_{st}^{SA}$ and $\hat{RF}_{st}^{SG}$. Then in a second stage I estimate:

$$\hat{RF}_{st}^{SA} = \lambda_0 + \lambda_1 \hat{\gamma}_{st}^{G} + \lambda_2 \hat{\gamma}_{st}^{P(SA)} + e_{st} \tag{3.18}$$

$$\hat{RF}_{st}^{SG} = \psi_0 + \psi_1 \hat{\gamma}_{st}^{E} + \psi_2 \hat{\gamma}_{st}^{P(SG)} + v_{st} \tag{3.19}$$

using variance weighted least squares (VWLS) weighting by the estimated variance of the reduced form from the first stage subsample regressions.[43] Given no correlation between the structural error term in equation (3.14) and the error term in equation (3.18), λ_1 recovers the effect of a grade level on math scores, and λ_2 recovers the effect of pre-kindergarten participation on math scores. Similarly, given no correlation between the structural error term in equation (3.16) and the error term in equation (3.19), ψ_1 recovers the effect of entering school a year older on math scores, and ψ_2 also recovers the effect of pre-kindergarten participation on math scores.

In my preferred specifications, I also include state and year fixed effects to control for time-invariant characteristics of states and differences across years related to the costs and benefits to compliance with entry laws. With state and year fixed effects in the model, bias from unobserved determinants of compliance and response to entry laws will only result if these determinants change within a state over time.

Table 3.11 shows results of the estimation of equation (3.18). Each column is an OLS or VWLS regression 116 (108) observations for age 9 (13), representing the state-year cells available in the analysis sample.[44] Focusing on the preferred specifications in columns (4) and (9), estimates at age 9 suggest about a 1/3 standard deviation

[43] I do not correct for the estimated first stage regressors on the right-hand side of the regression.
[44] Appendix Table F.2 describes the states and years in the analysis sample.

effect of pre-kindergarten on math scores, and about 1-1.5 of a standard deviation effect of a grade level on math scores. By age 13, the effect of pre-kindergarten remains about 1/3 of a standard deviation, and the effect of grade level has declined to about 1/2 of a standard deviation. OLS results are similar. Omitting the fixed effects does not affect the estimate of the effect of pre-kindergaten participation on math scores, suggesting that the result is not driven by unobserved characteristics of states, and reducing concerns of large bias in the results. The effect of grade level for 9 year olds is reduced slightly when the state and year fixed effects are removed.

Table 3.12 shows results of the estimation of equation (3.19) by OLS for 9 year-old students (five years after assigned entry). These results are much noisier than the analogous same age, different age results in Table 3.11, and also rely on linearity assumptions in the first stage and reduced form specifications (equations (3.15) and (3.17)). The point estimates in the preferred specifications in columns (4) and (7) suggest very small effects of pre-kindergarten on math scores and small or negative effects of entering school when older on math scores.

3.7 Conclusion

The effects of entry policies on student outcomes work through several channels. Age effects and instructional time have received much attention in the prior literature, but the effects of entry policies on pre-school experiences and parent and child investments in human capital accumulation while in school have largely been ignored.

I find that an extra year in school increases math scores by about 3/4 of a standard deviation at age 9, with no discernible effect by age 13, though the estimates are likely lower bounds. One explanation for the fade out with age is rapid depreciation of early learning. I find that a year of age provides gains of about 1/3 of a standard deviation of academic achievement five years after school entry (likely an upper bound), and no distinguishable effect after the ninth year. As with the effects of time in school, maturity and relative age effects could diminish with age as a result of depreciation, especially if they are most relevant for very young students. Regardless of the reason for the fadeout, the results are inconsistent with models of age effects in which a child's stock of human capital (related to maturity) is complementary to instructional time, which would predict widening effects of relative age with time in school.

An alternate explanation for the effects of entry policies often attributed to age is differences in early childhood investment in learning. To account for the the full effect of school entry laws on student achievement usually attributed to age, pre-kindergarten enrollment would need to have implausibly large effects on test scores. Using variation across states and years in the propensity of students to comply with and respond to entry laws, I find evidence of effects of pre-kindergarten participation of at about 1/3 standard deviation at age 9 and 13.

The results in this paper imply that redshirting has limited potential to benefit students academically since the effects of age (relative or absolute) on test scores are small. However, if redshirting allows more parents to enroll students in pre-kindergarten, students will benefit at least in the early years of schooling. I find evidence that the effects of early school interventions persist through age 13, but through pre-kindergarten participation rather than any benefits of being older in itself. However, some of the long-run benefits of redshirting may not be captured by test scores; early school experiences have been shown to have long-run effects on educational attainment and labor market outcomes, even after test score effects fade (Cunha et al., 2006; Heckman et al., 2010; Chetty et al., 2010; Dynarski et al., 2011).

Tables

Table 3.1: Summary Statistics of Sample

	Age 9			Age 13		
	Mean	SD	N	Mean	SD	N
Student Demographics						
Male	0.492		18,961	0.490		17,940
Black	0.176		18,961	0.152		17,940
Hispanic	0.089		18,961	0.087		17,940
White	0.706		18,961	0.734		17,940
Other race	0.030		18,961	0.027		17,940
School Characteristics						
Percent black	17.4	25.4	17,820	15.7	24.4	16,952
Percent hispanic	9.0	18.0	17,820	9.1	17.4	16,952
Percent white	70.0	29.7	17,820	72.3	28.4	16,952
0-5 percent free lunch	0.126		15,169	0.154		13,808
6-10 percent free lunch	0.101		15,169	0.111		13,808
11-25 percent free lunch	0.225		15,169	0.266		13,808
26-50 percent free lunch	0.247		15,169	0.258		13,808
51-75 percent free lunch	0.153		15,169	0.143		13,808
76-100 percent free lunch	0.148		15,169	0.069		13,808
No bilingual students	0.840		13,000	0.808		11,450
1-5 percent bilingual	0.060		13,000	0.103		11,450
6-100 percent bilingual	0.101		13,000	0.088		11,450
No ESL students	0.620		13,056	0.606		11,469
1-5 percent ESL	0.234		13,056	0.301		11,469
6-100 percent ESL	0.146		13,056	0.093		11,469
Investment in Schooling						
Attend pre-kindergarten	0.731		17,284	0.718		16,531
Attend kindergarten	0.978		18,686	0.980		17,549

Notes: The math and reading summary statistics are based on separate samples of students. Student characteristics are based on survey responses of the student. School characteristics are based on administrator surveys. Surveys differ across between ages, and within and across survey years. Sample size varies as a result of non-response, and also because some questions are not included in all test booklets.

Table 3.2: Grade of Enrollment and Math Scores by Assigned Entry Cohort

Panel A – All Students				
	Assigned Early Entry		Assigned Late Entry	
	Fraction	Mean Math Score	Fraction	Mean Math Score
Age 9				
Grade 2	0.004	176.8	0.032	178.3
Grade 3	0.188	205.7	0.855	215.2
Grade 4	0.808	242.4	0.112	241.1
N	13,071		5,870	
Age 13				
Grade 6	0.010	233.2	0.051	236.0
Grade 7	0.211	255.2	0.821	270.0
Grade 8	0.779	283.3	0.127	281.5
N	12,640		5,274	

Panel B – Students Born within One Month of Cutoff				
	Assigned Early Entry		Assigned Late Entry	
	Fraction	Mean Math Score	Fraction	Mean Math Score
Age 9				
Grade 2	0.006	175.9	0.024	182.2
Grade 3	0.349	211.5	0.775	215.1
Grade 4	0.645	240.1	0.201	241.0
N	1,765		1,697	
Age 13				
Grade 6	0.015	240.5	0.036	234.5
Grade 7	0.341	259.7	0.744	270.3
Grade 8	0.645	281.7	0.219	282.9
N	1,643		1,463	

Notes: Students assigned early entry are those with birthdates ahead of their state's entry cutoff date, and those assigned late entry have birthdates after the cutoff date.

Table 3.3: First Stage and Reduced Form Effects of School Entry Policies

	Age 9		Age 13	
	(1)	(2)	(3)	(4)
First stage – grade level	0.400**	0.406**	0.363**	0.373**
	(0.029)	(0.023)	(0.030)	(0.023)
	{0.6247}	{0.1691}	{0.8408}	{0.5549}
Control mean	3.08		7.08	
Standard deviation	0.37		0.42	
N	18,961		17,940	
First stage – years of schooling	0.329**	0.341**	0.321**	0.341**
	(0.034)	(0.029)	(0.038)	(0.030)
	{0.9930}	{0.8499}	{0.4358}	{0.2069}
Control mean	4.81		8.81	
Standard deviation	0.06		0.06	
N	17,158		16,442	
Reduced form – scaled score	6.7**	7.3**	-1.3	0.9
	(1.6)	(1.1)	(1.7)	(1.1)
	{0.7244}	{0.1352}	{0.5190}	{0.3591}
	[0.24]	[0.25]	[-0.05]	[0.03]
Control mean	217.0		269.7	
Standard deviation	28.7		28.2	
N	18,961		17,940	
Controls?	No	Yes	No	Yes

Notes: Each cell is a first stage or reduced form discontinuity estimate from a separate RD regression (Equations (3.13) and (3.12) in the text) with standard errors clustered by state-year-relative month of birth in parentheses, and the p-value from an F-test of the quadradratic parametric model versus the saturated model in brackets. The reduced form test score discontinuity is reported in standard deviation units in square brackets. All models include year fixed effects. Control variables include test booklet-year fixed effects, state-year fixed effects, and the student and school demographic controls listed in Table 3.1. Years of completed schooling is reported grade level plus indicators for attended pre-k and kindergarten. The indicators for attended pre-k and kindergarten are based on student self-reports. The sample sizes in the years of schooling regressions are smaller as a result of dropping students with no response to the pre-k or kindergarten attendence survey question. Treating non-response as non-attendance of pre-k and kindergarten results in slightly smaller point estimates on the years of schooling first stage, but the differences are not statistically significant. See text for discussion of sample restrictions and further details. * significant at 10%; ** significant at 5%.

Table 3.4: Estimated Effect of Starting School a Year Early on Math Scores

	Age 9		Age 13	
	(1)	(2)	(3)	(4)
Non-experimental "effect" of grade level	30.4**	29.9**	18.9**	18.5**
	(0.5)	(0.4)	(0.6)	(0.5)
	[1.09]	[1.07]	[0.69]	[0.68]
Scaled by grade level first stage	16.9**	18.0**	-3.6	2.5
	(3.6)	(2.3)	(4.7)	(2.8)
	[0.59]	[0.63]	[-0.13]	[0.09]
Scaled by years of schooling first stage	22.9**	23.6**	-1.7	4.8
	(4.5)	(2.9)	(5.3)	(3.1)
	[0.81]	[0.83]	[-0.06]	[0.17]
Controls?	No	Yes	No	Yes

Notes Each cell is a scaled discontinuity estimate from a separate 2SLS regression (Equations (3.13) and (3 14) in the text) with standard errors clustered by state-year-relative month of birth in parentheses and the discontinuity estimate scaled as a fraction of a standard deviation in student test scores in brackets. The non-experimental estimates for age 9 (13) students are based on an OLS regression of test score on an indicator for enrolled 4th (8th) grade relative to 3rd (7th) grade students in other grades are dropped from these comparisons with standard errors clustered by state-year-relative month of birth in parentheses. The control variables are described in Table 3.3 notes Dependent variable means, standard deviations, and sample sizes are reported in Table 3.3. * significant at 10%; ** significant at 5%.

Table 3.5: Tests of Smoothness of Student and School Demographics

Dependent Variable	Male	Black	Percent Black in School	Student Demo Proj	School Demo Proj	Student and School Demo Proj
	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity, age 9	0 006	-0 009	-1 174	-0 031	-0 001	-0 031
	(0 019)	(0 014)	(0 809)	(0 035)	(0 034)	(0 035)
Control mean	0 475	0 179	16 828	-0 007	-0 026	-0 021
Standard deviation	0 499	0 384	25 220	1 001	0 999	0 995
N = 18,961						
Discontinuity, age 13	-0 014	0 027**	1 923**	-0 092**	-0 088**	-0 103**
	(0 023)	(0 013)	(0 838)	(0 035)	(0 036)	(0 034)
Control mean	0 490	0 143	14 480	0 014	-0 002	0 006
Standard deviation	0 500	0 351	23 568	0 992	0 994	0 990
N = 17,940						

Notes· Each cell is a scaled discontinuity estimate from a separate RD regression (Equation (3.13) with the outcome replaced with the column heading) with standard errors clustered by state-year-relative month of birth in parentheses. Only the state-year and test booklet-year fixed effects are included as control variables. Projections are the predicted values from a regression of math scores on the demographic controls in Table 3.1 * significant at 10%; ** significant at 5%.

Table 3.6: Reduced Form Effect of Assigned Entry Age on Math Scores

	No controls (1)	With controls (2)	Quadratic specification (3)	Outcome demeaned by grade level (4)
Panel A – 5 Years After Assigned Entry (Age 9)				
Assigned entry age	16.8**	16.9**	142.1**	-90.2**
	(1.6)	(1.1)	(63.8)	(53.9)
	[0.57]	[0.57]		
Assigned entry age squared			-11.7**	8.9**
			(5.9)	(5.0)
Marginal effect at 5.1 years			23.1**	0.0
			(3.4)	(2.8)
			[0.78]	[0.00]
Marginal effect at 5.6 years			11.4**	8.9**
			(2.9)	(2.5)
			[0.38]	[0.34]
Mean	235.5	235.5	235.5	-1.1
Standard deviation	29.8	29.8	29.8	26.2
N	13,023	13,023	13,023	13,023
Panel B – 9 Years After Assigned Entry (Age 13)				
Assigned entry age	9.7**	9.1**	97.8	-58.8
	(2.0)	(1.2)	(69.0)	(62.3)
	[0.33]	[0.31]		
Assigned entry age squared			-8.3	5.7
			(6.4)	(5.8)
Marginal effect at 5.1 years			13.4**	-0.8
			(3.5)	(3.1)
			[0.45]	[-0.03]
Marginal effect at 5.6 years			5.1	4.9
			(3.3)	(3.1)
			[0.17]	[0.18]
Mean	277.3	277.3	277.3	-1.9
Standard deviation	29.6	29.6	29.6	27.6
N	12,517	12,517	12,517	12,517

Notes: Each column in a panel is a separate OLS regression (Equation (3.17) in the text) with standard errors clustered by state-year-assigned age of entry in parentheses and the estimates scaled as a fraction of a standard deviation in student test scores in brackets. The control variables, included in all specifications, are described in Table 3.3 notes. See text for discussion of sample restrictions and further details. * significant at 10%; ** significant at 5%.

Table 3.7: Effect of Entering School When One Year Older on Math Scores

	First Stage – Imputed Age of Entry		2SLS – Math Score		2SLS – Demeaned Math Score	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A – 5 Years After Assigned Entry (Age 9)						
Entry age	0.596**	-6.67**	28.5**	487.9**	7.9**	-44.2
	(0.019)	(1.11)	(2.4)	(98.7)	(1.6)	(79.2)
			[0.96]		[0.27]	
Entry age squared		0.678**		-42.7**		4.8
		(0.103)		(9.1)		(7.3)
Marginal effect at 5.1 years		0.239**		52.2**		5.2
		(0.061)		(6.0)		(4.7)
				[1.75]		[0.20]
Marginal effect at 5.6 years		0.917**		9.5**		10.0**
		(0.049)		(3.8)		(3.3)
				[0.32]		[0.38]
Panel B – 9 Years After Assigned Entry (Age 13)						
Entry age	0.635**	-7.32**	14.3**	242.2**	3.5**	-34.3
	(0.020)	(1.17)	(2.1)	(82.5)	(1.7)	(72.2)
			[0.48]		[0.12]	
Entry age squared		0.742**		-20.9**		3.5
		(0.109)		(7.5)		(6.6)
Marginal effect at 5.1 years		0.248**		28.8**		1.1
		(0.064)		(6.0)		(5.1)
				[0.97]		[0.04]
Marginal effect at 5.6 years		0.990**		7.9**		4.5*
		(0.050)		(2.5)		(2.5)
				[0.27]		[0.16]

Notes: Each column in a panel is a separate OLS or 2SLS regression (Equations (3.15) and (3.16) in the text) with standard errors clustered by state-year-assigned age of entry in parentheses and the estimates scaled as a fraction of a standard deviation in student test scores in brackets. The independent age variables in columns (1) and (2) is assigned entry age. The control variables are described in Table 3.3 notes. Dependent variable means, standard deviations, and sample sizes are the same as reported in Table 3.6. See text for discussion of sample restrictions and further details. * significant at 10%; ** significant at 5%.

Table 3.8: Implied Effect of Pre-Kindergarten Attendence on Math Scores at Age 9

	First Stage – Attended Pre-k		2SLS – Math Score		2SLS – Demeaned Math Score	
	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity	-0.074**	-0.068**			65.9**	67.3**
	(0.021)	(0.018)			(25.3)	(21.1)
					[2.52]	[2.57]
Assigned entry age	0.069**	0.062**	234.2**	264.9**	57.4**	64.5**
	(0.023)	(0.018)	(77.9)	(79.6)	(26.5)	(23.9)
			[7.87]	[8.90]	[2.19]	[2.46]
Controls?	No	Yes	No	Yes	No	Yes

Notes· Each cell is a separate RD or IV regression with standard errors clustered by state-year-relative month of birth (RD) or state-year-assigned age of entry (IV) in parentheses and the estimates scaled as a fraction of a standard deviation in student test scores in brackets. The control variables are described in Table 3.3 notes See text for discussion of sample restrictions and further details. * significant at 10%; ** significant at 5%

Table 3.9: Student Heterogeneity in the Effects of an Additional Assigned Year of Schooling at Age 9

	All	Public	Private	Boys	Girls	White	Black	Other Race
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grade level (FS)	0 406**	0 423**	0 297**	0 342**	0 459**	0 373**	0 473**	0 504**
	(0 023)	(0 025)	(0 046)	(0 029)	(0 029)	(0 025)	(0 042)	(0 060)
Control mean	3 08	3 08	3 13	3 06	3 10	3 08	3 10	3 07
Standard deviation	0 37	0 37	0 37	0 37	0 38	0 35	0 44	0 42
N	18,961	16,907	2,054	9,320	9,641	13,383	3,332	2,246
Pre-kindergarten attendance (FS)	-0 068**	-0 075**	-0 029	-0 118**	-0 028	-0 087**	0 012	-0 093*
	(0 018)	(0 019)	(0 044)	(0 028)	(0 024)	(0 021)	(0 041)	(0 055)
Control mean	0 74	0 73	0 86	0 72	0 76	0 76	0 77	0 60
Standard deviation	0 44	0 44	0 35	0 45	0 42	0 43	0 42	0 49
N	17,284	15,342	1,942	8,490	8,794	12,372	3,005	1,907
Math scaled score (RF)	7 3**	7 8**	5 3**	3 4**	10 7**	6 8**	8 8**	8 1**
	(1 1)	(1 2)	(2 6)	(1 5)	(1 4)	(1 2)	(2 6)	(3 3)
Control mean	217 0	215 5	229 9	218 8	215 3	222 8	198 8	209 4
Standard deviation	28 7	28 6	26 0	28 9	28 4	26 5	28 0	28 9
N	18,961	16,907	2,054	9,320	9,641	13,383	3,332	2,246

Notes: Each cell is a first stage or reduced form discontinuity estimate from a separate first stage (FS) or reduced form (RF) RD regression (Equations (3.13) and (3.12) in the text) with standard errors clustered by state-year-relative month of birth in parentheses. The control variables are described in Table 3 3 notes. See text for discussion of sample restrictions and further details. * significant at 10%; ** significant at 5%.

Table 3.10: Student Heterogeneity in the Effects of an Additional Assigned Year of Entry Age at Age 9

	All	Public	Private	Boys	Girls	White	Black	Other Race
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Imputed age of entry (FS)								
Marginal effect at 5 1 years	0 239**	0 303**	-0 132	0 005	0 449**	0 063	0 740**	0 825**
	(0 061)	(0 065)	(0 155)	(0 090)	(0 075)	(0 070)	(0 119)	(0 171)
Marginal effect at 5 6 years	0 917**	0 890**	1 023**	1 001**	0 848**	0 912**	0 962**	0 653**
	(0 049)	(0 051)	(0 124)	(0 077)	(0 057)	(0 055)	(0 102)	(0 140)
Mean	5 55	5 55	5 53	5 59	5 51	5 54	5 59	5 55
Standard deviation	0 40	0 40	0 37	0 42	0 38	0 39	0 45	0 41
N	13,023	11,573	1,450	6,500	6,523	9,206	2,256	1,555
Pre kindergarten attendance (FS)	0 062**	0 063**	0 052	0 067**	0 063**	0 081**	0 000	0 057
	(0 018)	(0 019)	(0 048)	(0 030)	(0 026)	(0 022)	(0 043)	(0 062)
Mean	0 73	0 71	0 84	0 70	0 75	0 74	0 76	0 61
Standard deviation	0 45	0 45	0 36	0 46	0 43	0 44	0 43	0 49
N	11,861	10,485	1,376	5,903	5,958	8,503	2,024	1,328
Demeaned math scaled score (RF)								
Marginal effect at 5 1 years	0 0	0 3	3 9	-1 0	0 8	-4 0	13 9*	9 2
	(2 8)	(3 0)	(8 3)	(4 1)	(3 7)	(3 4)	(8 0)	(9 5)
Marginal effect at 5 6 years	8 9**	8 7**	15 8*	7 3*	11 0**	13 9**	-9 1	3 3
	(2 5)	(2 7)	(8 3)	(3 9)	(3 7)	(3 3)	(7 5)	(7 9)
Mean	-1 2	-1 2	-0 6	-1 4	-1 0	1 0	-1 3	1 4
Standard deviation	26 2	26 1	24 6	26 7	25 5	23 7	24 9	26 4
N	13,023	11,573	1,450	6,500	6,523	9,206	2,256	1,555

Notes Each cell is a separate first stage (FS) or reduced form (RF) OLS regression (Equation (3 15) or (3 17) in the text) with standard errors clustered by state-year-assigned age of entry in parentheses The control variables, included in all specifications, are described in Table 3 3 notes See text for discussion of sample restrictions and further details * significant at 10%, ** significant at 5%

Table 3.11: Effects of Pre-Kindergarten and Grade Level on Math Scores

Dependent Variable = Math Score Reduced Form Estimate	Age 9					Age 13				
	OLS		VWLS			OLS		VWLS		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Pre-kindergarten first stage estimate	11 1	9 1	10 7**	10 4**	15 7**	-1 0	6 1	9 5**	10 1**	6 9
	(7 4)	(6 5)	(3 9)	(4 5)	(5 5)	(7 5)	(6 9)	(3 5)	(4 3)	(5 0)
	[0 39]	[0 32]	[0 37]	[0 36]	[0 55]	[-0 03]	[0 22]	[0 34]	[0 36]	[0 25]
Grade level first stage estimate	33 0**	45 4**	33 3**	42 3**	40 7**	22 3**	20 8**	12 4**	13 9**	15 3**
	(6 0)	(5 9)	(2 8)	(3 8)	(3 9)	(6 7)	(6 8)	(3 0)	(4 1)	(4 3)
	[1 15]	[1 58]	[1 16]	[1 47]	[1 42]	[0 79]	[0 74]	[0 44]	[0 49]	[0 54]
Pre-k x grade level interaction					-16 9*					17 7
					(9 9)					(14 5)
					[-0 59]					[0 63]
Constant	-6 6**		-5 7**			-7.2**		-3 2**		
	(2 9)		(1 4)			(2 7)		(1 4)		
State and Year FE?	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
R-squared	0 419	0 685				0 175	0 553			
p-value from test of model restrictions			0 0000	0 0420	0 0537			0 0000	0 0129	0 0135
N	116	116	116	116	116	108	108	108	108	108

Notes: Each cell is a separate second stage OLS or VWLS regression (Equation 3.18 in the text) with the discontinuity estimate scaled as a fraction of a standard deviation in student test scores in brackets. Each observation represents a state-year. Weights in VWLS specifications are based on the reduced form variance estimate. See text for discussion of sample restrictions and further details. * significant at 10%; ** significant at 5%.

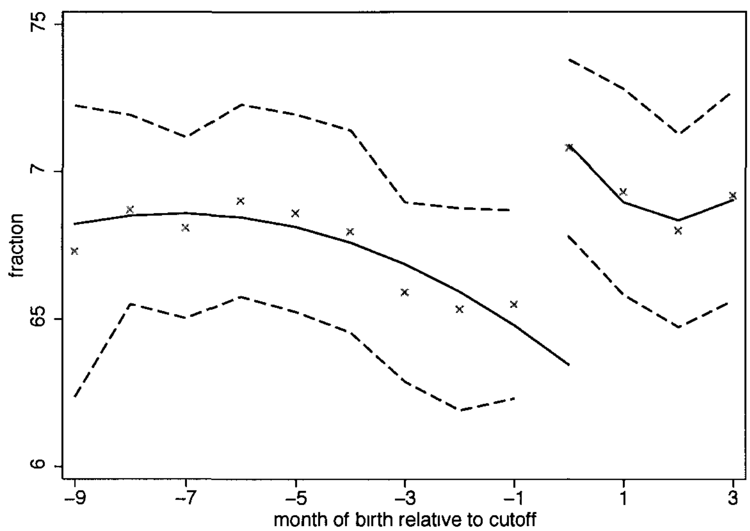
Table 3.12: Effects of Pre-Kindergarten and Age on Math Scores

Dependent Variable = Math Score Reduced Form Estimate (Demeaned)	Age 9					Age 13				
	OLS		VWLS			OLS		VWLS		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Pre-kindergarten first stage estimate	5 1	4 8	4 7	1 0	-13 4	-8 0	-5 8	-7 0**	-5 2	-6 4
	(5 2)	(5 4)	(3 1)	(3 7)	(8 3)	(6 2)	(6 2)	(3 0)	(3 6)	(9 2)
	[0 17]	[0 16]	[0 16]	[0 03]	[-0 45]	[-0 27]	[-0 19]	[-0 24]	[-0 17]	[-0 22]
Imputed entry age first stage estimate	-1 6	-7 9	4 9*	-6 5	-6 8	-16 5**	-18 0**	-6 6**	-12 5**	-12 5**
	(6 7)	(6 7)	(2 8)	(4 3)	(4 3)	(7 3)	(6 6)	(2 7)	(4 2)	(4 2)
	[-0 05]	[-0 26]	[0 16]	[-0 22]	[-0 23]	[-0 55]	[-0 60]	[-0 22]	[-0 42]	[-0 42]
Pre-k x age interaction					24 5*					1 8
					(12 7)					(12 2)
					[0 82]					[0 06]
Constant	6 1		0 9			13 5**		7 2**		
	(4 4)		(1 7)			(5 1)		(1 8)		
State and Year FE?	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
R-squared	0 012	0 369				0 135	0 483			
p-value from test of model restrictions			0 0000	0 0000	0 0000			0 0000	0 0006	0 0005
N	116	116	116	116	116	108	108	108	108	108

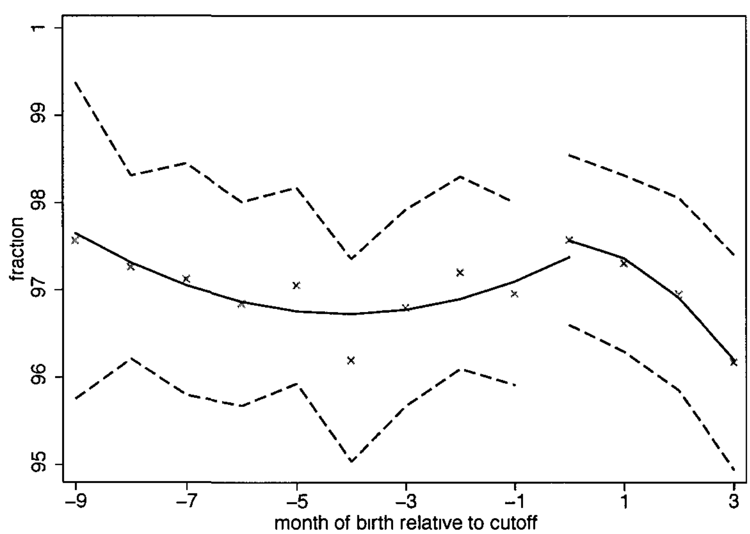
Notes Each cell is a separate second stage OLS or VWLS regression (Equation 3.19 in the text) with the discontinuity estimate scaled as a fraction of a standard deviation in student test scores in brackets. Each observation represents a state-year. Weights in VWLS specifications are based on the reduced form variance estimate. See text for discussion of sample restrictions and further details. * significant at 10%; ** significant at 5%.

Figures

Figure 3.1: Fraction Attended Pre-K and kindergarten by Month of Birth Relative to Cutoff



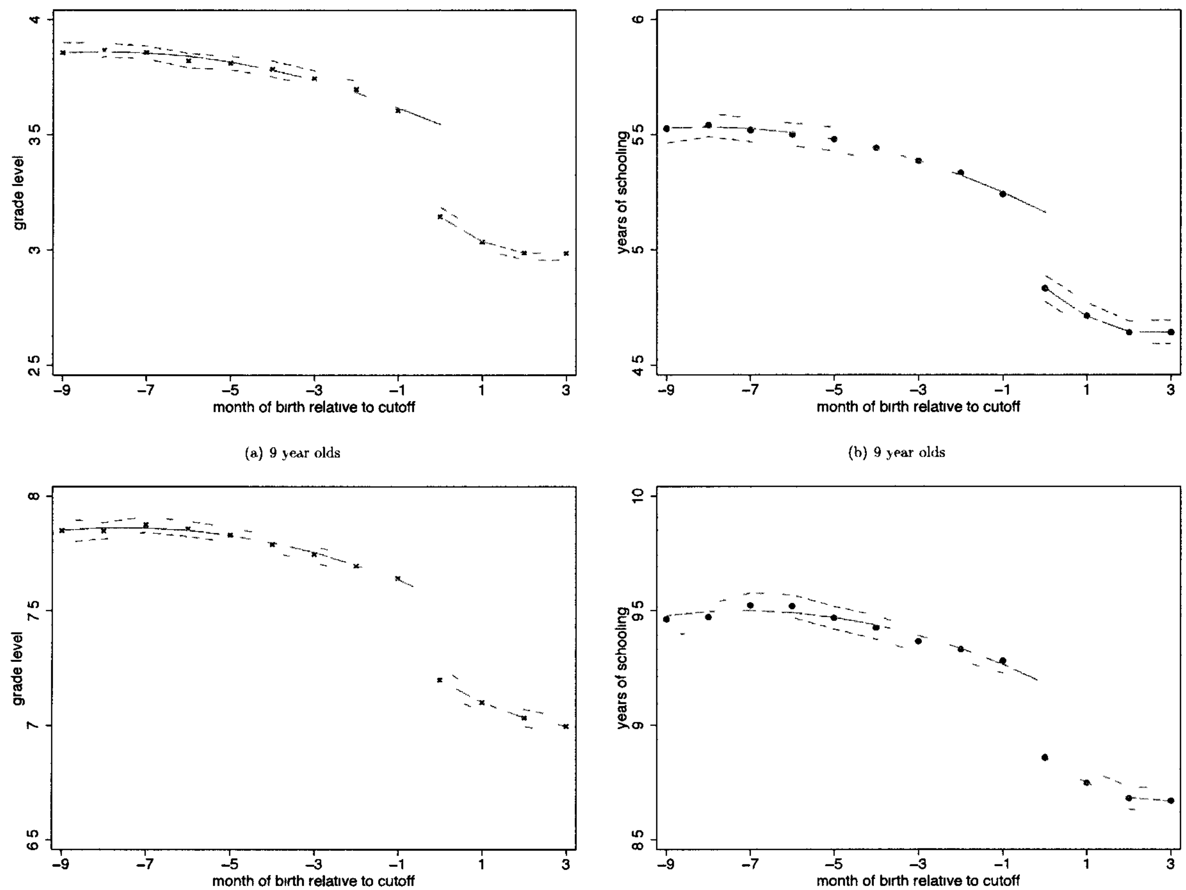
(a) Attended pre-k



(b) Attended kindergarten

Notes: Math sample, age 9. The indicators for attended pre-k and kindergarten are based on student self-reports. The means are estimated from a regression of the relevant measure of completed schooling on indicators for the month of birth relative to the cutoff. The dashed lines show 95 percent confidence intervals around the estimated means using standard errors robust to clustering at the state-year-relative month of birth level. The solid line gives a quadratic trend with the parameterization allowed to differ on either side of the cutoff. Specifications include survey year fixed effects. See text for discussion of sample restrictions and further details.

Figure 3.2: First Stage – Mean Grade Level and Years of Completed Schooling by Month of Birth Relative to Cutoff



(a) 9 year olds

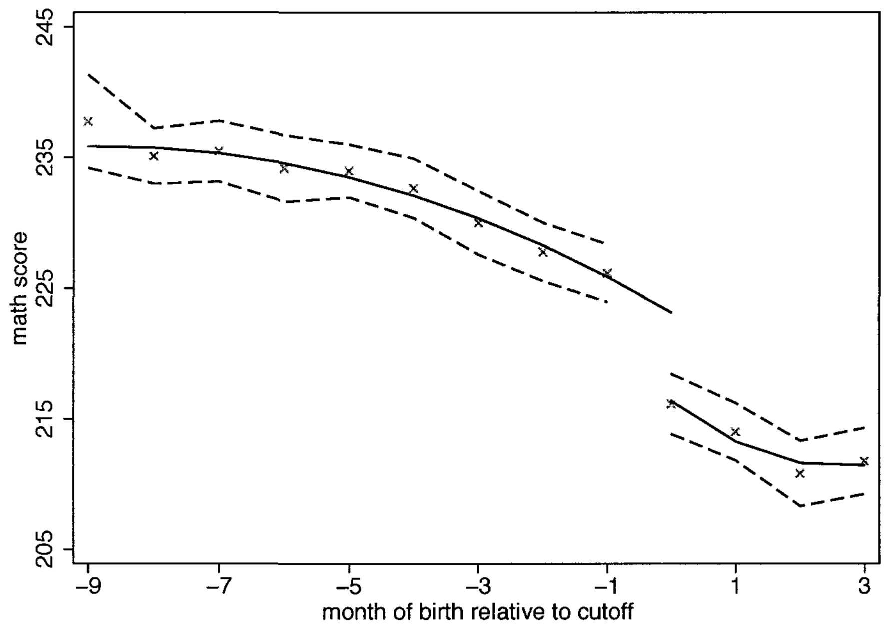
(b) 9 year olds

(c) 13 year olds

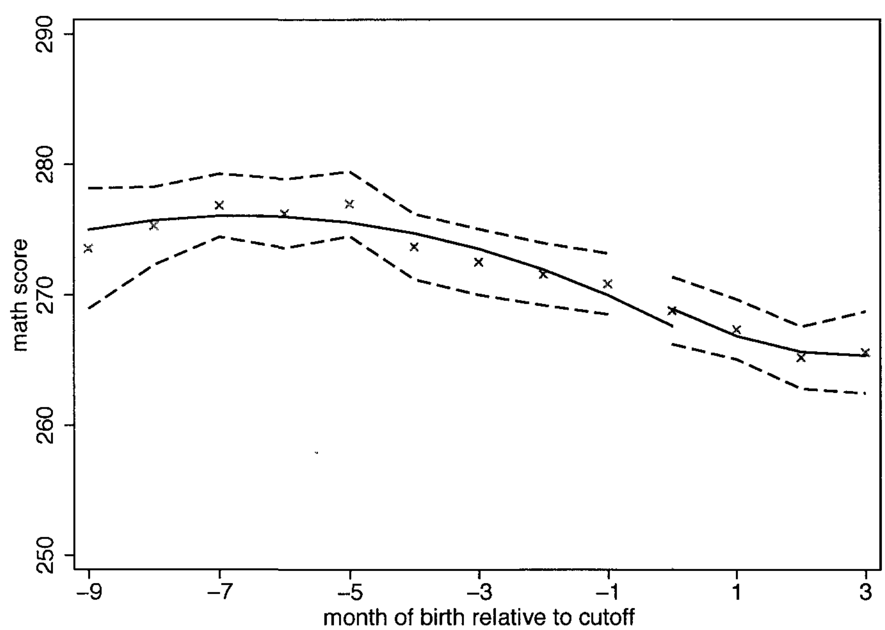
(d) 13 year olds

Notes: Math sample. See notes on Table 3.3. Grade level is child's reported grade level of enrollment. Years of completed schooling is reported grade level plus indicators for attended pre-k and kindergarten. Specifications include survey year fixed effects. See text for discussion of sample restrictions and further details.

Figure 3.3: Reduced Form – Mean Math Score by Month of Birth Relative to Cutoff



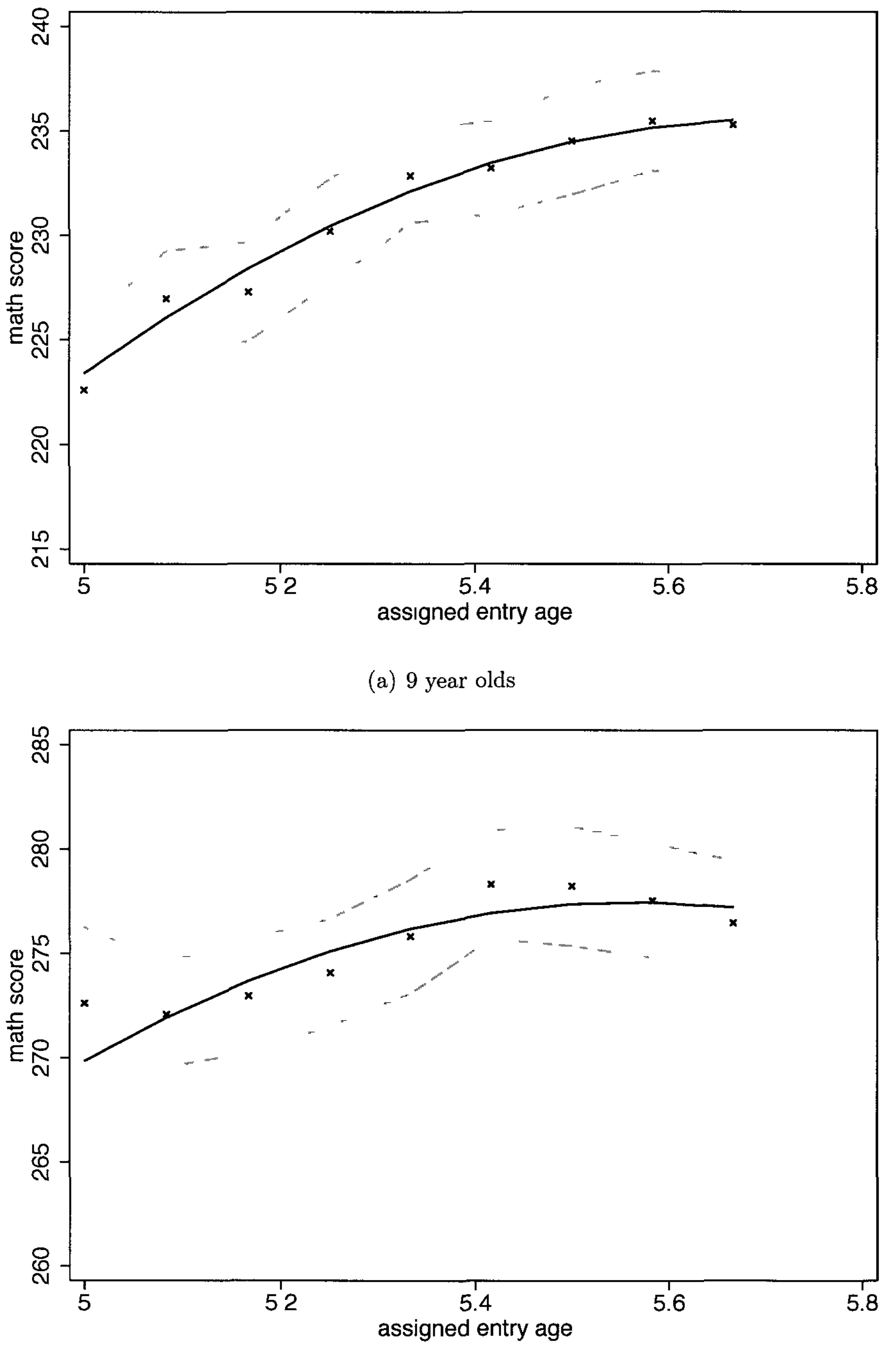
(a) 9 year olds



(b) 13 year olds

Notes: Math sample. Math score is IRT scaled score. Specifications include survey year fixed effects. See notes on Table 3.3. See text for discussion of sample restrictions and further details.

Figure 3.4: Reduced Form Effect of Assigned Entry Age on Math Scores

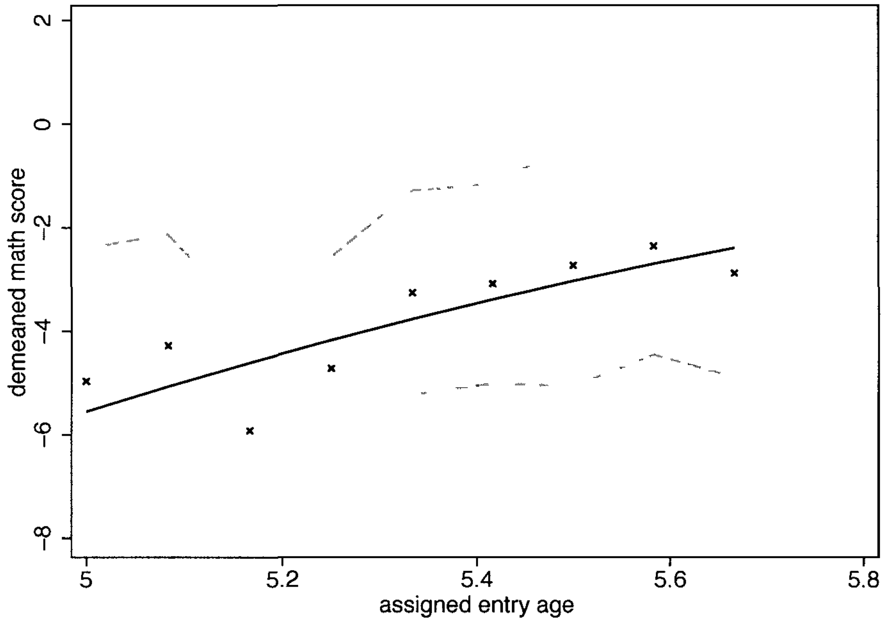


(a) 9 year olds

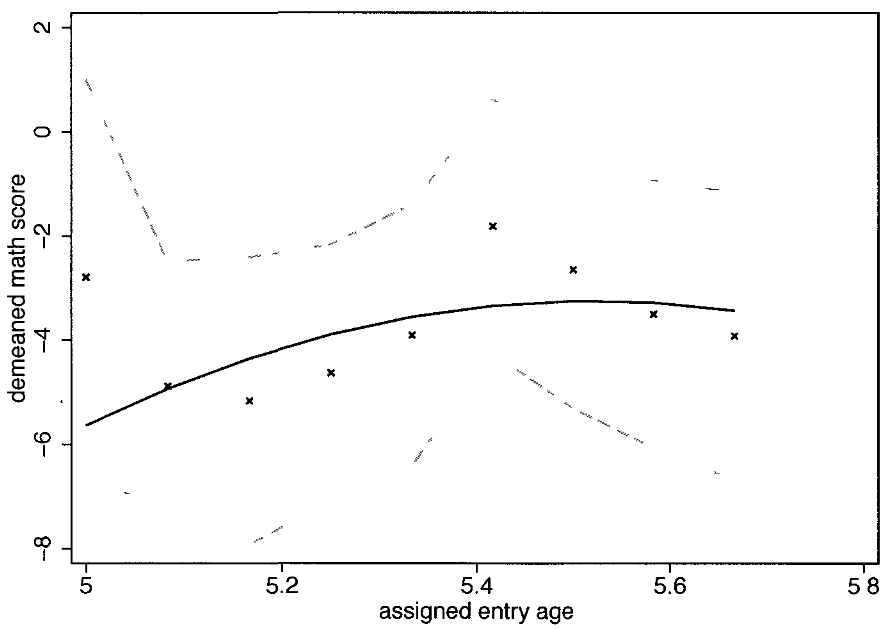
(b) 13 year olds

Notes: Math sample. Math score is IRT scaled score. See notes on Table 3.6. Assigned entry age is based on a September school start date and the student's month of birth relative to the relevant school entry cutoff. Specifications include survey year fixed effects. See text for discussion of sample restrictions and further details.

Figure 3.5: Reduced Form Effect of Assigned Entry Age on Demeaned Math Scores



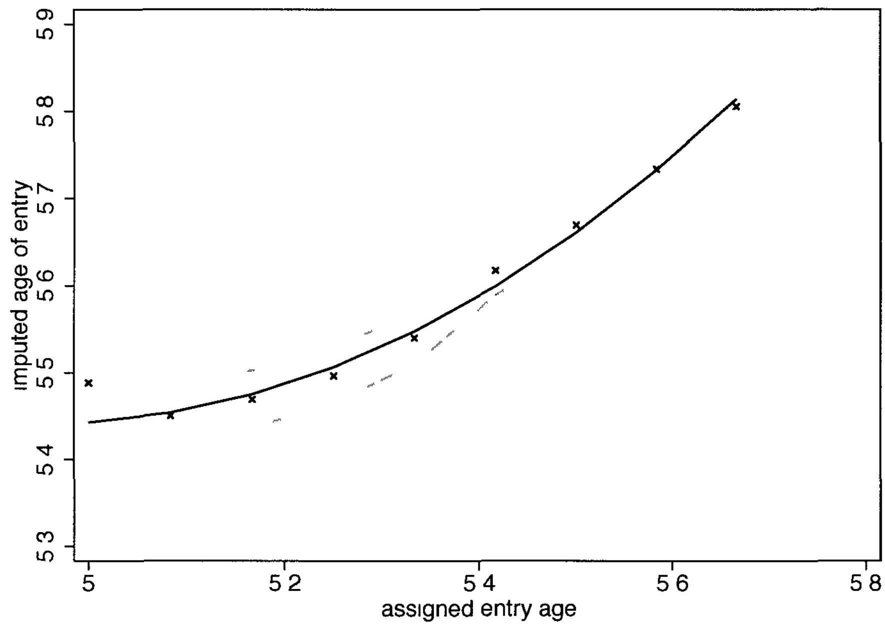
(a) 9 year olds



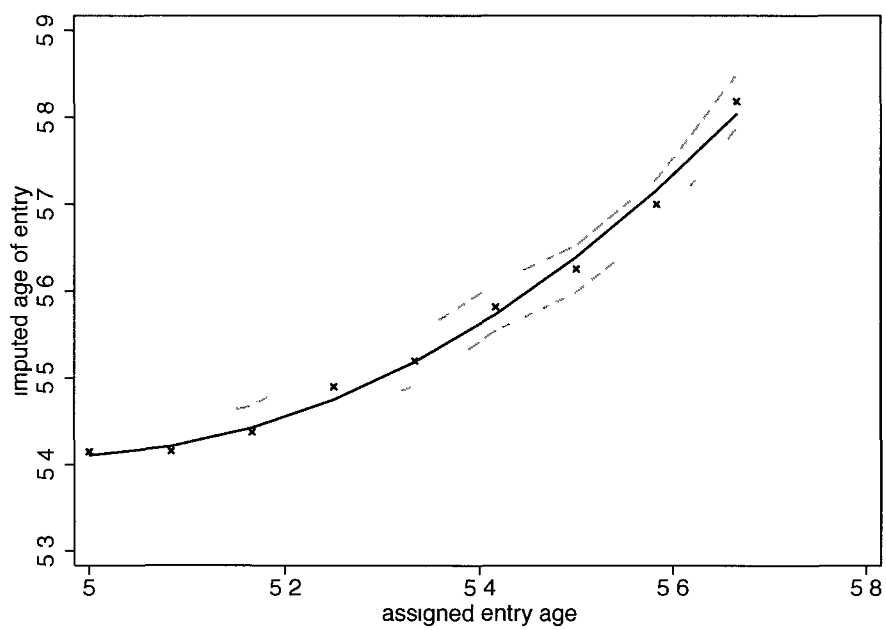
(b) 13 year olds

Notes: Math sample. The demeaned math score is the IRT scaled math score demeaned by grade level using all students in the full analysis sample, born on both sides of the entry cutoff. See notes on Table 3.6. Assigned entry age is based on a September school start date and the student's month of birth relative to the relevant school entry cutoff. Specifications include survey year fixed effects. See text for discussion of sample restrictions and further details.

Figure 3 6 First Stage Effect of Assigned Entry Age on Imputed Age of Entry



(a) 9 year olds



(b) 13 year olds

Notes Math sample See notes on Table 3 6 Assigned entry age is based on a September school start date and the student's month of birth relative to the relevant school entry cutoff Imputed age of entry is based on the student's current grade level and age Specifications include survey year fixed effects See text for discussion of sample restrictions and further details

APPENDICES

APPENDIX A

March Current Population Survey Data Appendix

I construct two sample from the 1964-2009 March Current Population Survey (CPS) cross-sections: a count sample for supply measures, and a wage sample used to estimate college-high school wage gaps. The count sample is broader than the wage sample to ensure that the aggregate counts accurately represent the market faced by workers. While the aggregate supply measures include all 1.8 million full-time, full-year employed workers, this is not the case for the cohort-specific and age-cohort specific supply measures. In constructing the measure of cohort-specific relative supply used in the main results of this paper, approximately 138,000 individuals are dropped from the count sample as a result of not being born between 1920 and 1979. A further 123,000 observations are dropped because of cohort-age group cells without full support of the ages, representing 120 of 1,805 cohort-age group cells. For example, I only observe workers through age 39 in the 1970 birth cohort, and as a result, the age 36-40 cell of the 1970 birth cohort must be dropped. Age group cells are defined as follows: 20-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-64.

For some robustness checks (in Appendix C), I instead calculate cohort-age group specific relative supply measures. These use age groupings to match the four-category decompositions: 26-30, 31-40, 41-50, 51-60. As a result, approximately 258,000 individuals not aged between 26 and 60 are dropped from the sample. 84,000 are dropped as a result of the birth cohort limitation, and 263,000 from cells without full age support, representing 214 of 1,464 coarse cohort-age group cells. The number dropped for age support concerns is larger here since the age categories are coarser.

Once the supply measures are merged onto the wage gap sample, I drop 74 (of 1,464) cells for which the cohort-specific relative supply of college labor cannot be constructed (but the wage gap can as a result of using single-year age-cohort cells). The final sample contains 1,365 single-year age-cohort cells, 273 fine age group-cohort cells, and 169 coarse age group-cohort cells.

I use the supplement weights in constructing all wage and supply measures used in the analysis. The weights ensure that the wage and supply measures are based on a nationally representative sample and allow aggregate counts of labor supply measures. In addition, since cohort-specific supply measures aggregate across sample years, it is important that the counts reflect the actual population in the US rather than a share of a sample.

APPENDIX B

Inference Appendix

As discussed in the text, because the main specifications are estimated on grouped data, VWLS will be inefficient in the presence of group error components (Dickens, 1990). There is reason to suspect within-cohort correlation in errors, in which case the heteroskedasticity-robust standard errors presented above may also be inconsistent. In this appendix, I present alternative standard error estimates robust to these concerns as well as the possibility of AR(1) serial correlation in the errors.

First, I use feasible generalized least squares (FGLS) to estimate models accounting for heteroskedasticity and cohort-specific error components. In the first step, I estimate equation (1.11), the main first stage decomposition, weighting by the variances from the micro data estimates of the cohort-year specific wage gaps. Next, I regress the log squared residuals from the weighted estimation of (1.11) on the micro data variance, the interaction between the variance and the same set of cohort, year, and age group fixed effects from the decomposition, and a constant term. The interaction terms allow for heteroskedasticity in the "individual" (cohort-year) error variance. The constant term represents the sum of the group error variances.

I then use the inverse of the exponentiated predicted values as weights in a new estimate of equation (1.11), providing efficient estimates of the model under the assumption that the model of the variance is correct. I iterate the procedure until the estimated (exponentiated) intercept is equal rounded to the fourth decimal place in consecutive iterations.

Finally, I use the new estimates of the decomposition fixed effects as the dependent variable in the main second stage models with weights constructed from the new variance estimates. I show the baseline second stage results in columns (1) and (2) of Table B.1, and the results of the FGLS estimates in columns (3) and (4). The standard errors are smaller than the baseline estimates, showing efficiency gains of the FGLS procedure, while the point estimates are not much affected (both sets of results provide consistent estimates). The FGLS standard error estimates could still be biased if the model of the error variances is mis-specified, though mis-specification would not affect consistency of the model coefficients.

In a separate procedure I estimate models robust to AR(1) serial correlation across birth cohorts using a standard FGLS procedure. Because I find evidence that the serial correlation parameter varies across age groups (specifically, I find more serial correlation in the younger age groups), I estimate models separately by age. These models are more flexible because they allow the cohort trend and relative supply different effects on the wage gap by age. I present the baseline specifications in columns (5) and (6). Estimates in each of these columns are based on three regressions (one for each age group). I do not present the cohort trend and relative supply coefficients. Despite the additional flexibility, the results do not differ much from those in columns (1) and (2). Accounting for serial correlation tends to reduce standard errors slightly (columns (7) and (8)). The minor differences across the columns justify ignoring serial correlation in the main results, though more general forms of serial correlation may be present.

Table B 1 Alternative Standard Errors

	Second Stage Estimates							
					Estimated Separately by Age Group			
	Baseline		FGLS		Baseline		AR(1) Serial Correlation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cohort	0 004**	-0 003	0 001**	-0 006**	**Interactions not reported**			
	(0 001)	(0 004)	(0 000)	(0 002)				
Cohort-specific relative supply	-0 316**	-0 277**	-0 317**	-0 235**	**Interactions not reported**			
	(0 031)	(0 065)	(0 015)	(0 046)				
Age 26-30 interactions:								
Unemployment at age 19	0 012**	0 017**	0 009**	0 014**	0 011**	0 018**	0 009**	0 019**
	(0 004)	(0 007)	(0 002)	(0 005)	(0 005)	(0 007)	(0 005)	(0 006)
Aggregate relative supply of labor at age 19		0 281*		0 172		0 395*		0 496**
		(0 162)		(0 130)		(0 203)		(0 173)
Log aggregate supply of high school labor at age 19		0 056		0 089		-0 018		-0 054
		(0 095)		(0 072)		(0 122)		(0 063)
Age 31-40 interactions:								
Unemployment at age 19	0 003	0 004	0 003	0 004	-0 001	0 003	0 002	0 004
	(0 003)	(0 006)	(0 002)	(0 005)	(0 004)	(0 007)	(0 005)	(0 006)
Aggregate relative supply of labor at age 19		0 222		0 176		0 081		0 008
		(0 164)		(0 129)		(0 160)		(0 148)
Log aggregate supply of high school labor at age 19		0 118		0 118*		0 211**		0 101
		(0 092)		(0 068)		(0 087)		(0 063)
Age 41-50 interactions:								
Unemployment at age 19	0 001	0 008	0 000	0 005	-0 002	0 009	-0 003	0 004
	(0 004)	(0 009)	(0 002)	(0 007)	(0 005)	(0 011)	(0 005)	(0 006)
Aggregate relative supply of labor at age 19		0 496**		0 359**		0 640		0 339**
		(0 191)		(0 155)		(0 430)		(0 115)
Log aggregate supply of high school labor at age 19		-0 048		0 012		0 076		-0 031
		(0 108)		(0 088)		(0 351)		(0 055)
Number of observations	124	85	124	85	124	85	124	85

Notes Each of columns (1)–(4) is an OLS regression Each of columns (5)–(8) shows the result of three regressions estimated separately by age groups 26-30, 31-40, and 41-50 All models include age group fixed effects and include samples limited to age 26-50 Standard errors robust to clustering by cohort in parentheses for columns (1) and (2) Standard errors on elasticities calculated using the delta method See Appendix B for discussion of standard errors and estimation of columns (3)–(8) * significant at 10%, ** significant at 5%

APPENDIX C

Alternate Parameterizations of Production and the Wage Gap

In this section I assess the importance of the changes to the production function to allow for more flexible parameterization of imperfect age substitution. The discussion and empirical work in Welch (1979) suggest several ways in which the model in equation (1.6) may be restrictive. First, η, the parameter related to the elasticity of substitution between age groups, might differ for college and high school workers. It is plausible that the skills of high school workers see larger changes with age than those of college workers. Second, conditional on education, workers closer in age to each other may be closer to perfect substitutes than those further apart. The model outlined above imposes a single elasticity regardless of the ages of the workers under comparison. Third, the ability of firms to substitute across ages could be "one-sided."

To take an extreme case, if old workers have all of the skills of young workers, but young workers do not have old worker skills, then the elasticity of substitution between ages for young workers will be small, while much larger for old workers.

All of these are concerns for estimation insofar as ignoring them would lead to bias, not only in the elasticity parameters, but also in the persistent effects of initial labor market conditions. The latter two concerns both predict a cohort-age interaction – the effect of cohort-specific relative supply will not be constant with age. It will be of particular importance to assess the empirical importance of the latter pair of theoretical concerns since it is possible that they alone can explain in the wage gap what would otherwise be attributed to wage persistence.

To estimate more flexible models of production, I include an interaction between cohort-specific relative supply and age to account for differences in the substitutability of skills between older and younger workers. Second, I include a coarser bin (across cohorts) used in the calculation of the relative supply measure to account for more perfect substitution between workers who are closer in age. Finally, I include college and high school supply separately in some specifications to account for differences in the elasticity of substitution between age within education group.

The results of this analysis are presented in Table C.1. The baseline specification is in column (1). Instead of the regression-adjusted cohort-specific relative supply used in the baseline, column (2) uses the raw cohort-age group specific relative supply. This measure could be important if there are changes in the fraction of workers who obtain college educations later in life across cohorts. In Figure C.1 the cohort-age group relative supplies can be seen to be nearly parallel, suggesting this may not be important. The similarity between results in columns (1) and (2) confirms this.[1]

Column (3) includes a measure of relative supply incorporating all workers within 5 years of age of a given cohort. If these workers are perfect substitutes for each other then this regressor should have a larger (in absolute value) coefficient than the baseline regressor. The results do not support this. Even so, in column (4) I include the broad age-specific relative supply along with the baseline measure. The total effect of a change in the cohort's own relative supply is not significantly altered (the sum of the two coefficients), and the results suggest that the neighboring cohorts are responsible for just under half of the effects of imperfect age substitution.

Column (5) allows the elasticity to vary between college and high school workers. If the simple model is correct (a single value for η), then the cohort-specific log supplies of college and high school workers should enter the model with equal magni-

[1]The sample size falls between columns (1) and (2) as a result of needing to drop cohort-age group cells without full age support, for which the cohort-age group specific relative supply would be biased.

tude, but opposite signs. The results, -0.271 and 0.341, indicate a larger elasticity of substitution between ages for college workers (3.7 versus 2.9), significantly different with p=0.0001). Results using the broader measures of age-specific supply are similar. Column (7) again includes both sets of supplies together, and the own-cohort's relative supply dominates, at least for college workers.

Finally, I also estimate changes in the substitution parameters across age group. The results in column (8) suggest that older workers have higher elasticities of substitution than younger workers, meaning that younger workers may have skills employers cannot easily obtain from older workers.[2]

It should be emphasized that the presence of persistent effects on the wage gap from omitted labor market conditions could easily lead to bias in results discussed in this section. However, the effect of unemployment rates faced by cohorts at age 19 does not vary significantly across specifications in Table C.1. The more flexible models considered in this section have no impact on the estimated effect of initial unemployment rates on the wage gap.

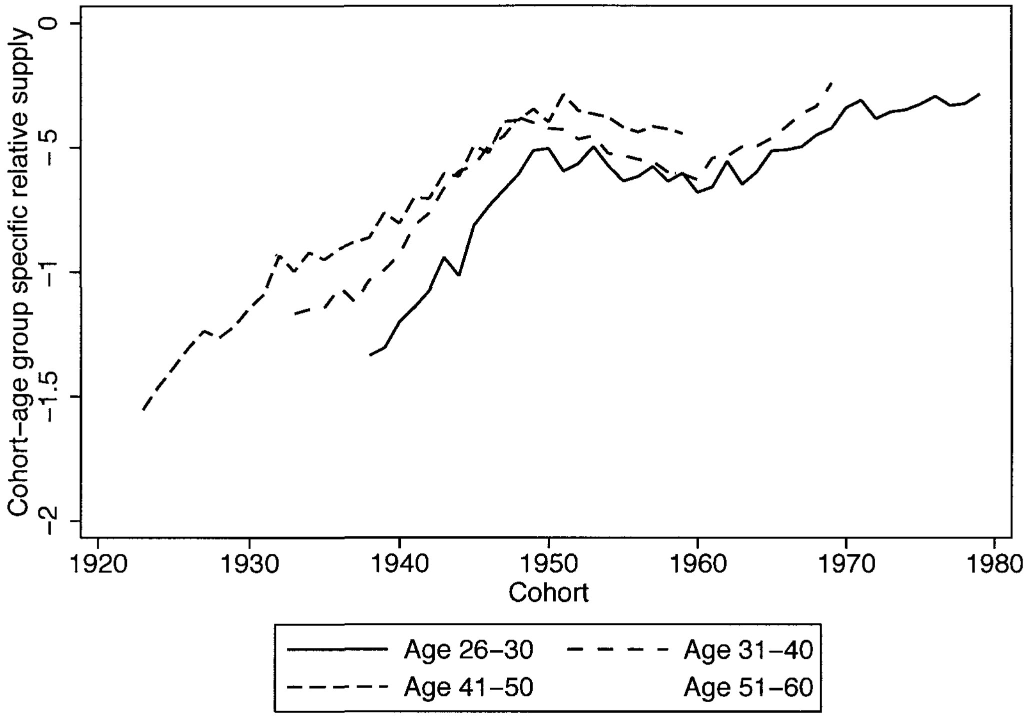
[2]In an unreported specification I allow the elasticity of substitution to vary with age and education. I find marginal evidence for an age interaction for both education groups (I obtain p=0.0221 from a test of the college age interaction restriction, and p=0.0881 for high school), though the cohort-specific relative supply effect for the young age group does not vary across education (p=0.9128).

Table C.1: Flexible Parameterizations of Imperfect Age Substitution

	Second Stage Estimates							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cohort	0.004**	0.004**	0.002**	0.003**	0.001**	0.001**	0.001**	0.005**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Cohort-specific relative supply (age 35-40)	-0.316**			-0.142*				-0.410**
	(0.028)			(0.074)				(0.034)
Cohort-age group specific relative supply		-0.271**						
		(0.022)						
Cohort-specific relative supply, MA(11)			-0.279**	-0.163**				
			(0.023)	(0.070)				
Cohort-specific supply of college labor					-0.271**		-0.216**	
					(0.020)		(0.082)	
Cohort-specific supply of high school labor					0.341**		0.162	
					(0.032)		(0.110)	
Cohort-specific supply of college labor, MA(11)						-0.320**	-0.083	
						(0.025)	(0.101)	
Cohort-specific supply of high school labor, MA(11)						0.410**	0.241	
						(0.047)	(0.151)	
Cohort-specific relative supply x age 31-40								0.035
								(0.030)
Cohort-specific relative supply x age 41-50								0.123**
								(0.029)
Unemployment at age 19 x age 26-30	0.012**	0.014**	0.016**	0.014**	0.012**	0.013**	0.010**	0.010**
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)
Unemployment at age 19 x age 31-40	0.003	0.002	0.010**	0.006*	0.003	0.007**	0.004	0.002
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
Unemployment at age 19 x age 41-50	0.001	-0.002	0.007*	0.004	0.001	0.003	0.001	-0.006
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Number of observations	124	110	124	124	124	124	124	124
R-squared	0.800	0.826	0.803	0.809	0.820	0.816	0.824	0.822

Notes: Each column is an OLS regression. Standard errors robust to clustering at the cohort level in parentheses. All models include age group fixed effects. The cohort-specific supply at age 35-40 is normalized in the manner discussed in the text. The cohort-age group specific supply is the raw relative supply in each cohort-age group cell. I cannot construct this measure for 23 age group-cohort cells in the regression because the cell does not have full age support in the March CPS between 1964 and 2009. Limiting the regression in column (1) to the sample in column (2) does not significantly affect the results. The 11 period moving average is constructed by totaling both the total cohort specific supply of college and high school labor (regression adjusted to age 35-40) across the 11 cohorts centered in the own cohort, and then taking the log ratio. * significant at 10%; ** significant at 5%.

Figure C.1: The Cohort-Age Group Specific Relative Supply of College Labor



Notes The sample includes all full-time, full-year employed men and women age 26-60 in the CPS March Supplements between 1964 and 2009 The cohort-age group specific relative supply of college labor is the log ratio of the supply of effective college labor to the supply of effective high school labor in a birth cohort and in a given age group See text The figure shows three-year moving averages

APPENDIX D

Additional Robustness Checks

D.1 Alternative choices for age of entry

Table D.1 explores the robustness of the main results to alternative choices for the age of labor market entry for high school workers.[1] I find support for the choice of age 19 as the high school entry age, though results using age 20 as the entry age are similar. I do not find an effect of unemployment at age 18 on the wage gap.

Columns (1) and (2) give the baseline specification with and without controls for aggregate supply at entry. The remaining columns show these same specifications but assuming high school market entry at age 18 or 20 as an alternative to 19. The unemployment results in the columns without controls for initial aggregate supply

[1]The choice for age of entry for college workers in specifications including controls for initial conditions at college entry does not affect the results.

are not significantly changed by choice of entry ages, though the magnitude of the marginal effect does fall from 1.2 to 0.8 percent between the age 19 and age 18 high school entry columns. The specifications with controls for initial aggregate supply results in columns (2) and (6) appear similar, but those in column (4) are a clear outlier.[2] Though differences across specifications could be interpreted as support for using the age 19 unemployment rate, the standard errors on the effects of unemployment at age 18 and 20 are large enough to include the baseline specification estimate in the 95% confidence interval.

D.2 Effects of initial unemployment rates by single-year age groups

I estimate two versions of a single-stage regression model of the single-year cohort-time specific wage gaps. I estimate specifications similar to those in columns (2) and (3) of Table 1.3, but with single-year age interactions with the age 19 unemployment rate. The first version of the regression includes observable contemporaneous regressors: a linear time trend, the aggregate relative supply of labor, and the unemployment rate. The second version instead includes a full set of year fixed effects. I plot the coefficients on the age 19 unemployment rate age interactions in Figure

[2]The sample size in column (4) is smaller as a result of dropping cohorts for whom the aggregate supply of high school labor at age 18 is unavailable, though this is not driving the difference in the results (dropping the 1945 cohort from column (2) does not affect the results).

D.1. The model which relies on observable time controls (circles) shows no evidence of persistent effects on the wage gap. The results of the model with time effects, on the other hand, show statistically significant effects of 1-2 percentage points, fading after age 35. These results are consistent with the results in Table 1.3 and support the grouping of ages into the 26-30 category,

Also, the differences between the models with observable contemporaneous labor market conditions and with year fixed effects are consistent with the results in columns (2) and (3) in Table 1.3, which indicated the presence of unobserved contemporaneous determinants of the wage gap correlated with initial unemployment rates. Including year fixed effects resolves this concern, but Oreopoulos et al. (2008), Kahn (2010), and Genda et al. (2010), for example, use regional variation in unemployment rates at graduation, but include just year fixed effects (instead of region x year), leaving open the concern that past unemployment rates in the region are correlated with future regional labor market conditions.

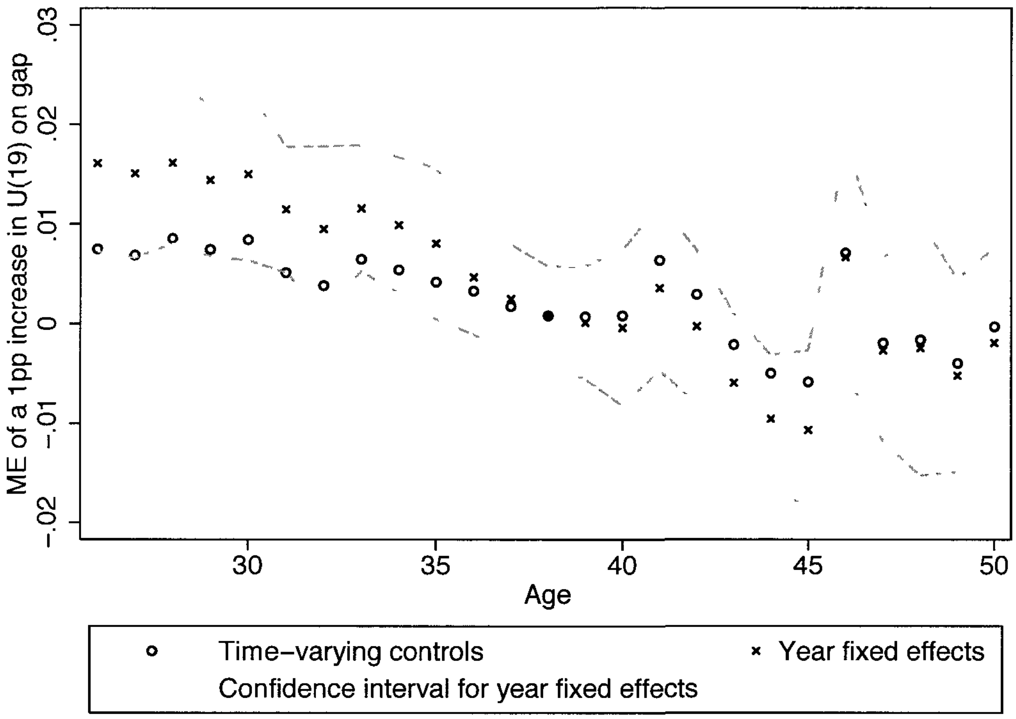
Thus, it is imperative to account for correlations between initial conditions and unobserved contemporaneous conditions. Simply controlling for the contemporaneous unemployment rate, as is done in many of the works cited above, is not sufficient for identification of wage persistence. The evidence in Figure D.1 suggests that the bias will tend to lead to an understatement of the magnitude of the persistence, indicative of positive correlations between initial unemployment rates and unobserved contemporaneous conditions, so the economy several years after a recession exhibits recovery (benefitting high school workers relative to college workers) beyond what is captured by the contemporaneous unemployment rate.

Table D.1: Robustness of Persistence Results to Alternative Choices for Age of Entry

	Second Stage Estimates					
High School Entry at	Age 19		Age 18		Age 20	
	(1)	(2)	(3)	(4)	(5)	(6)
Cohort	0 004**	-0 003	0 005**	-0 008**	0 004**	-0 004
	(0 001)	(0 004)	(0 001)	(0 004)	(0 001)	(0 004)
Cohort-specific relative supply	-0 316**	-0 277**	-0 326**	-0 199**	-0 311**	-0 218**
	(0 031)	(0 065)	(0 033)	(0 079)	(0 030)	(0 053)
Age 26-30 interactions:						
Unemployment at H entry	0 012**	0 017**	0 008**	-0 002	0 012**	0 010
	(0 004)	(0 007)	(0 004)	(0 007)	(0 004)	(0 009)
Aggregate relative supply of labor at H entry		0 281*		-0 054		0 125
		(0 162)		(0 174)		(0 193)
Log aggregate supply of high school labor at H entry		0 056		0 361**		0 145
		(0 095)		(0 101)		(0 104)
Age 31-40 interactions:						
Unemployment at H entry	0 003	0 004	0 001	0 007	0 004	0 005
	(0 003)	(0 006)	(0 004)	(0 006)	(0 004)	(0 005)
Aggregate relative supply of labor at H entry		0 222		0 358**		0 194
		(0 164)		(0 161)		(0 150)
Log aggregate supply of high school labor at H entry		0 118		0 163*		0 141*
		(0 092)		(0 091)		(0 084)
Age 41-50 interactions:						
Unemployment at H entry	0 001	0 008	-0 001	-0 009	-0 000	0 003
	(0 004)	(0 009)	(0 004)	(0 009)	(0 004)	(0 010)
Aggregate relative supply of labor at H entry		0 496**		0 212		0 292
		(0 191)		(0 198)		(0 208)
Log aggregate supply of high school labor at H entry		-0 048		0 229*		0 082
		(0 108)		(0 125)		(0 109)
Elasticity of substitution b/w age	3 2	3 6	3 1	5 0	3 2	4 6
	(0 31)	(0 84)	(0 31)	(1 99)	(0 32)	(1 11)
Number of observations	124	85	122	82	126	88
R-squared	0 800	0 796	0 785	0 799	0 798	0 762

Notes Each column is an OLS regression. Standard errors robust to clustering at the cohort level in parentheses. Standard errors on elasticities calculated using the delta method. All models include age group fixed effects. Sample is limited to age 26-50 cohort effects. * significant at 10%; ** significant at 5%.

Figure D.1: Marginal Effect of Unemployment at Age 19 on the Wage Gap by Age



Notes: Estimates obtained using a single-stage procedure. Both models include age group fixed effects, a linear cohort trend, and the cohort-specific relative supply of college labor. Time-varying controls include the aggregate relative supply of college labor and a linear year trend. The confidence interval is the 95% interval for each age-specific marginal effect. See text.

APPENDIX E

Chicago Public School Administrative Data Appendix

The teacher absence data files obtained from the Chicago Public Schools provide us with a daily attendance record for all teachers. Each day is coded for a teacher as one of the following: an attended work day, a holiday, an unscheduled day (so the teacher is not expected at work), a sick day, a personal day, or an unpaid absence as a result of using all sick and personal days. For the 2002-2003 and 2003-2004 school-years we have complete data, but in 2004-2005 and 2005-2006, due to payroll reporting procedures in the CPS, many 2-week fiscal periods are not reported. Specifically, fiscal periods in which a teacher recorded only regular work days, unscheduled days, and unpaid days are not observed in the two later school-years, though the rule is not perfectly applied, and we cannot rule out that periods containing other types of absences were also cut from the data. For the purposes of this analysis determining the exact nature of the absence is not a major concern. We are primarily interested in obtaining a measure the annual total number of sick days, personal days, and unpaid days for each teacher.

In cleaning the data we keep only fiscal periods during the regular school-year (September-June), and which either include or are after the reported hire date of the teacher.

For the missing fiscal periods we impute the number of holidays, work days, unscheduled days, and unpaid days. First, we look within period (across teachers) to determine the number of holiday days in the period. Next, for the remaining non-holiday days we use the ratio of work, unscheduled, and sick days in the teacher's nearby periods to impute the missing days. Specifically, we divide the school-year into three trimesters and compute the ratio using only periods within the trimester.

APPENDIX F

National Assessment of Educational Progress Long-Term Trend Data Appendix

A key component of my empirical strategy requires knowing the school entry cutoff date for states over time. These dates are obtained from state law, when available.[1] For CO, MA, NJ, and WA, state cutoff dates are unavailable because these cutoffs are determined locally, while I observe student residence only at the state level. Additionally, for IN prior to 1989, I have no information on the cutoff date. For the sample of 9 (13) year-olds, missing cutoff dates necessitate dropping 2,870 (3,069) students, roughly nine percent of the available sample.

Several additional states must be excluded from the analysis due to the timing of their entry cutoff relative to the LTT sampling frame. The LTT samples students on

[1] I would like to thank Kelly Bedard and Dimitry Lubotsky, who generously shared the entry date cutoffs they collected from state statutes and empirical investigation.

the basis of the calendar year in which they were born, so that student birth months range from January through December of a given year.[2] For this reason, states with entry cutoffs in early January or late December will have no students in the data on one side of the entry date cutoff. Most entry cutoffs are in the month of September, so this sample restriction is not a major concern: I drop 47 of 357 state-year-age cells.

In order to ensure that I observe students born in at least the three birth months on either side of the cutoff date within each state-year, I also drop state-year observations with entry cutoffs in late October, November, early December, or February (no states have school-entry cutoffs set in March). In addition, I drop a small number of state-years that are not well represented in the LTT, leading to spotty coverage of students across birth month. Thus, I drop an additional 86 state-year-age cells.

The fact that I observe month but not exact date of birth further complicates the analysis. The majority of state-years have school entry cutoffs that fall at or near the start or end of the month (e.g., the 1st, 2nd, 3rd, 30th, or 31st). For the state-years with entry cutoffs near the beginning (end) of a month, I assume that all students who were born in that month did not (did) make the cutoff. This introduces a small bit of measurement error since for states with a cutoff on the 2nd of the month roughly

[2]This describes the sampling for 9 and 13 year olds. The LTT sampling frame for 17 year olds is based on an October-September calendar year.

three percent of students born in the month will have actually made the cutoff.[3]

Private schools need not abide by the same school entry policies as public schools, which might be a reason to exclude these students from analysis. However, it is possible that a child's birth date relative to the cutoff induces selection out of public schools. The LTT samples public and private schools, allowing me to test for this kind of selection. In the main results, I include both public and private school students, treating them as non-compliers similar to students who are redshirted. In the heterogeneity results, I show results excluding the approximately 10-11 percent of students reported to attend private schools.

Table F.1: State-Year Entry Cutoffs in LTT Analysis Sample

	Age 9		Age 13	
Cutoff Date	State-Years	Students	State-Years	Students
June 1	3	159	1	34
July 1	4	487	3	397
August 31	2	90	1	17
September 1	67	12,322	64	11,999
September 2	3	360	2	86
September 3	11	1,298	8	749
September 15	5	455	4	708
September 30	11	2,453	10	2,034
October 1	10	1,337	15	1,916
Total	116	18,961	108	17,940

[3]In two of the remaining state-years in the sample (IA and ID), school entry cutoffs fall in the middle of the month. Including students born in the month of the cutoff date for these states would lead to the misclassification of a large number these students. I drop the 111 students who were born in the month of the cutoff date for these state-years.

Table F 2: State-Years in LTT Analysis Sample

State	Age 9						Age 13					
	1990	1992	1994	1996	1999	2004	1990	1992	1994	1996	1999	2004
Alabama	x	x	x	x	x	x	x	x	x	x	x	x
Arkansas	x	x	x		x		x	x	x		x	
Arizona	x	x	x		x	x	x	x	x		x	x
Deleware					x							
Florida	x	x		x		x	x	x		x		x
Georgia	x	x	x	x	x	x	x	x	x	x	x	x
Iowa	x			x	x	x	x			x	x	x
Idaho			x									
Illinois		x	x	x	x	x				x	x	x
Indiana				x	x	x					x	x
Kansas			x	x	x	x			x	x	x	x
Kentucky		x	x		x	x		x	x	x	x	x
Louisiana						x						x
Minnesota	x	x	x	x	x		x	x	x	x	x	
Missouri	x	x	x	x	x		x	x	x	x	x	
Mississippi	x		x		x	x	x		x		x	x
Montana	x						x					
New Mexico	x	x	x	x	x		x	x	x	x	x	
Nevada		x		x	x	x		x		x	x	x
Ohio		x	x	x	x	x	x	x	x	x	x	x
Oklahoma	x		x		x	x	x		x		x	x
Oregon	x	x		x	x	x		x		x	x	x
South Carolina						x						x
South Dakota						x						x
Tennessee	x	x	x	x		x			x	x		x
Texas	x	x	x	x	x	x	x	x	x	x	x	x
Utah	x		x	x		x	x		x	x		x
Virginia	x	x		x	x	x			x	x	x	x
Wisconsin	x	x	x			x	x	x	x			x
West Virgina				x	x	x				x	x	x

Notes Each "x" indicates a state-year-age cell is included in the final analysis sample

APPENDIX G

Investment Effects of Entry Policies

In Table G.1, I explore the effects of entry policies on alternative outcomes to test scores, relating to the human capital investment decisions of parents and students. Rows (1) and (2) isolate the effects of entry policies on kindergarten and pre-kindergarten enrollment from the first stage effect on total years of schooling reported in Table 3.3. The policies do not affect kindergarten enrollment rates (near universal on both sides of the cutoff), but assigned early entry lowers pre-kindergarten attendance rates by about seven percentage points (nine percent).[1]

In addition to entry policy effects on pre-school investments of students and parents, the policies may also affect human capital investments while enrolled in school.

[1]The age 13 pre-kindergarten discontinuity point estimate is less than half as large as the age 9 discontinuity. In principal, the effect of early entry on pre-kindergarten attendance should not change between age 9 and 13, except for sampling error. Though the difference is not statistically different, one possible reason for a difference is increased reporting error for older children.

Evidence is mixed on the effects of early entry on investment while in school. Table G.1 shows the effects of entry policies on reported time spent doing homework, daily hours of television, and the frequency the parent asks about homework. 9 year-old students assigned early entry report doing about 4 additional minutes of homework daily, and watching an additional 11 minutes of television ($11 \approx 0.189 * 60$). Discontinuities in time spent on homework and television for 13 year olds are not statistically significant. Figures G.1a and G.1c illustrate the discontinuity estimates of the effect of assigned early on homework for 9 and 13 year-olds.

One possible explanation for the differences in the effects of entry policies on investment between age 9 and age 13 is that the effects on homework and television for 9 year-olds represent peer or curriculum effects. Students assigned early entry interact with older children and are in a higher grade level on average than late entering students. Differences in curriculum or age-based student norms may be larger for younger students. It is plausible, for example, that the relative difficulty of 4th grade curriculum to 3rd grade curriculum is significantly larger than the difference in 8th versus 7th grade curriculum. Larger peer or curriculum effects on homework and television behavior for younger students would produce larger discontinuities for younger students.

If the investment effects of assigned early entry for 9 year olds are a result of changes in behavior unrelated to peer or curriculum effects, then students assigned early entry should invest more even when compared to the typical investments of students in the same grade level. To test this hypothesis, I demean the homework and television outcomes within grade level (but for the entire sample of students on both sides of the cutoff) prior to estimating the discontinuity. This ensures that students only get "credit" for investment over and above that of students enrolled in the same grade.

I report results using these "normalized" outcomes in rows (7) and (8) of Table G.1, and in Figures G.1b and G.1c. While the positive effect of assigned early entry on homework for 9 year olds disappears, the effect on television viewing is essentially unchanged. This suggests that the effects on homework time for 9 year olds can be explained by peer and/or curriculum effects, while assigned early entry increases television viewing even when compared to same grade peers. The normalized results for 13 year olds are similar to the raw outcome results, though a negative effect of early entry on time spent doing homework is now marginally statistically significant.

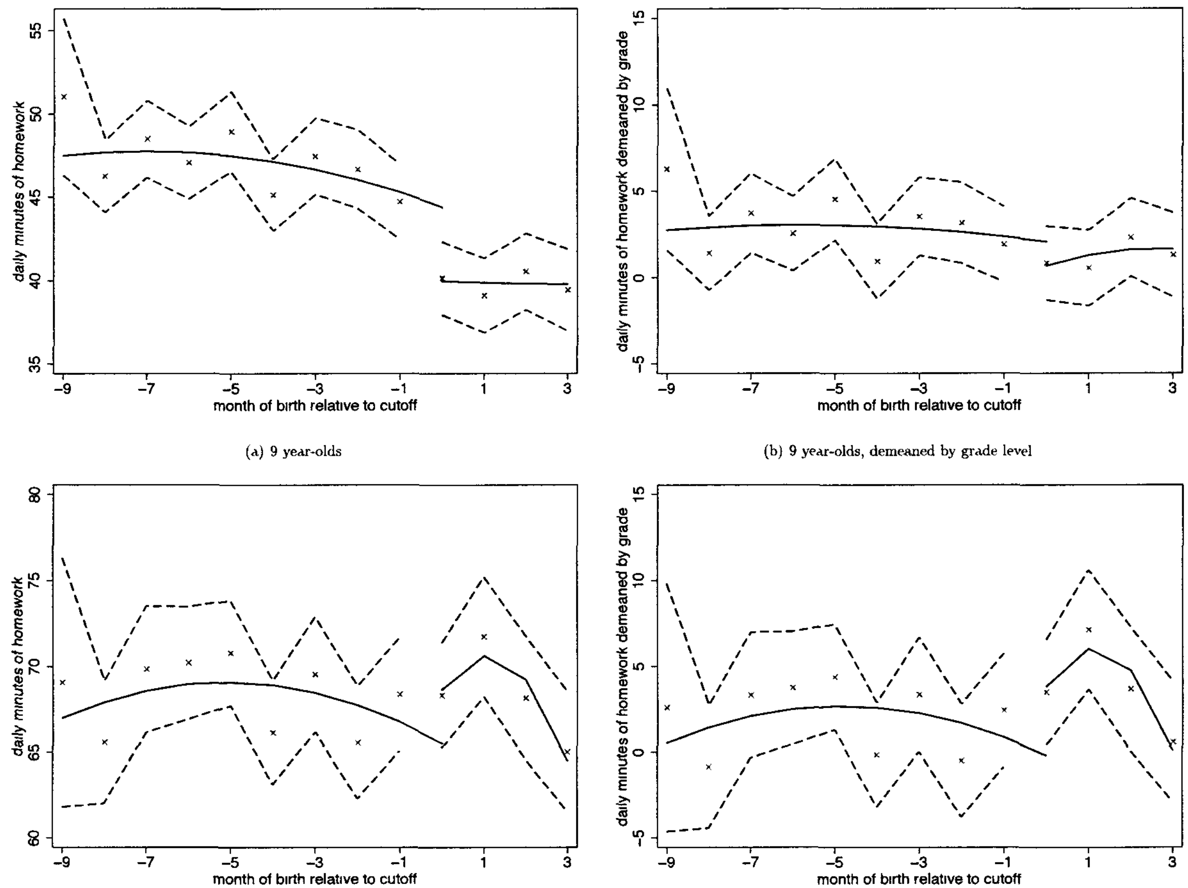
Together, the results of regressions using grade-normalized investment outcomes indicate that students assigned early entry may be less inclined to invest in their education than those assigned late entry. Lower investment for the early entry cohort could be the result of a compounding effect in which not attending pre-kindergarten leads students to invest less in their education throughout school.

Table G.1: Effects of School Entry Policies on Parent and Student Investment

		Age 9				Age 13			
		Discontinuity	Mean	SD	N	Discontinuity	Mean	SD	N
Attend pre-kindergarten	(1)	-0.068**	0.73	0.44	17,284	-0.033*	0.72	0.45	16,531
		(0.018)				(0.018)			
Attend kindergarten	(2)	-0.002	0.98	0.15	18,686	-0.006	0.98	0.14	17,549
		(0.005)				(0.006)			
Missing attend pre-kindergarten	(3)	0.015	0.09	0.28	18,961	0.028**	0.08	0.27	17,940
		(0.011)				(0.010)			
Missing attend kindergarten	(4)	0.010**	0.01	0.12	18,961	0.007	0.02	0.15	17,940
		(0.005)				(0.006)			
Minutes of daily homework	(5)	3.882**	42.62	34.93	18,881	-2.593	65.74	47.90	17,839
		(1.261)				(1.833)			
Hours of daily television	(6)	0.189**	3.22	1.82	18,873	0.028	3.23	1.60	17,877
		(0.068)				(0.062)			
Minutes of daily homework (norm.)	(7)	0.809	0.00	34.71	18,881	-3.501*	0.00	47.89	17,839
		(1.249)				(1.834)			
Hours of daily television (norm.)	(8)	0.175**	-0.00	1.82	18,873	0.055	-0.00	1.60	17,877
		(0.068)				(0.062)			
Parent asks about homework daily	(9)	-0.016	0.77	0.42	18,885	-0.008	0.80	0.40	17,812
		(0.016)				(0.016)			
Parent asks about homework monthly or less	(10)	0.014	0.14	0.34	18,885	0.013	0.09	0.28	17,812
		(0.013)				(0.012)			

Notes: Each cell is a discontinuity estimate from a separate RD regression (Equation (3.12) in the text) with standard errors clustered by state-year-relative month of birth in parentheses. All models include the control variables described in Table 3.3 notes. All outcomes are based on student self-reports. See text for discussion of sample restrictions and further details.

Figure G.1: Effect of Entry Policies on Daily Time Spent on Homework



(a) 9 year-olds

(b) 9 year-olds, demeaned by grade level

(c) 13 year-olds

(d) 13 year-olds demeaned by grade level

Notes: Math sample. See Table 3.3 notes. The dependent variable in panels (b) and (d) is demeaned within student grade level of enrollment prior to estimation. See text for discussion of sample restrictions and further details.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aliprantis, Dionissi, "Redshirting, Compulsory Schooling Laws, and Education Attainment," Working Paper 10-12, Federal Reserve Bank of Cleveland, Cleveland, OH September 2010.

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *The Journal of the American Statistical Association*, 2008, *103* (484), 1481–1495.

Angrist, Joshua D. and Alan B. Krueger, "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, 1991, *106* (4), 979–1014.

_ and _ , "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," *Journal of the American Statistical Association*, June 1992, *87*, 328–336.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney, "Trends in U.S. Wage Inequality: Revising the Revisionists," *The Review of Economics and Statistics*, May 2008, *90* (2), 300–323.

Baker, George, Michael Gibbs, and Bengt Holmstrom, "The Wage Policy of a Firm," *Quarterly Journal of Economics*, November 1994, *109* (4), 921–955.

Barnett, W. Steven, Cynthia Lamy, and Kwanghee Jung, "The Effects of State Prekindergarten Programs on Young Children's School Readiness in Five States," Working Paper, The National Institute for Early Education Research, Rutgers University December 2005.

Barua, Rashimi and Kevin Lang, "School Entry, Educational Attainment and Quarter of Birth: A Cautionary Tale of LATE," Working Paper 15236, National Bureau of Economic Research, Cambridge, MA August 2009.

Beaudry, Paul and David A. Green, "Cohort Patterns in Canadian Earnings: Assessing the Role of Skill Premia in Inequality Trends," *The Canadian Journal of Economics*, November 2000, *33* (4), 907–936.

_ and John DiNardo, "The Effect of Implicit Contracts on the Movement of Wages Over the Business Cycle: Evidence from Micro Data," *Journal of Political Economy*, August 1991.

Bedard, Kelly and Elizabeth Dhuey, "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects," *The Quarterly Journal of Economics*, November 2006, *121* (4), 1437–1472.

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

Blackburn, McKinley, David Bloom, and Richard Freeman, "Changes in Earnings Differentials in the 1980s: Concordance, Convergence, Causes and Consequences," NBER Working Paper 3901, National Bureau of Economic Research, Cambridge, MA. November 1991.

Bound, John and George Johnson, "Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations," *American Economic Review*, June 1992, *82*, 371–392.

_ , David A. Jaeger, and Regina M. Baker, "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, June 1995, *90*, 443–450.

Cahan, Sorel and Daniel Davis, "A Between-Grade-Levels Approach to the Investigation of the Absolute Effects of Schooling on Achievement," *American Educational Research Journal*, Spring 1987, *24* (1), 1–12.

_ and Nora Cohen, "Age versus Schooling Effects on Intelligence Development," *Child Development*, October 1989, *60* (5), 1239–1249.

Card, David and John DiNardo, "Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles," *Journal of Labor Economics*, 2002, *20*, 733–783.

_ and Thomas Lemieux, "Dropout and Enrollment Trends in the Post-War Period: What Went Wrong in the 1970s?," NBER Working Paper 7658, National Bureau of Economic Research, Cambridge, MA. April 2000.

_ and _ , "Can Falling Supply Explain The Rising Return To College For Younger Men? A Cohort-Based Analysis," *Quarterly Journal of Economics*, May 2001, *116* (2), 705–746.

Cascio, Elizabeth U. and Ethan G. Lewis, "Schooling and the Armed Forces Qualifying Test: Evidence from School-Entry Laws," *The Journal of Human Resources*, Spring 2006, *41* (2), 294–318.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan, "Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR," Working Paper 16381, National Bureau of Economic Research, Cambridge, MA September 2010.

Clotfelter, C. T., Helen F. Ladd, and Jacob L. Vigdor, "Are Teacher Absences Worth Worrying About in the U.S.?," Working Paper 13648, National Bureau of Economic Research, Cambridge, MA November 2007.

Cunha, Flavio, James J. Heckman, Lance Lochnet, and Dimitriy V. Masterov, "Interpreting the Evidence on Life Cycle Skill Formation," in E. Hanushek and F. Welch, eds., *Handbook of the Economics of Education, Volume I*, 2006.

Currie, Janet and Duncan Thomas, "Does Head Start Make a Difference?," *The American Economic Review*, June 1995, *85* (3), 341–364.

Datar, Ashlesha, "Does Delaying Kindergarten Entrance Age Give Children A Head Start," *Economics of Education Review*, February 2006, *25* (1), 43–62.

Deming, David and Susan Dynarski, "The Lengthening of Childhood," *The Journal of Economic Perspectives*, Summer 2008, *22* (3), 71–92.

Devereux, Paul J. and Robert A. Hart, "The Spot Market Matters: Evidence on Implicit Contracts from Britain," *Scottish Journal of Political Economy*, December 2007, *54* (5), 661–683.

Dhuey, Elizabeth and Stephen Lipscomb, "Disabled or Young? Relative Age and Special Education Diagnoses in Schools," *Economics of Education Review*, December 2010, *29* (5), 857–872.

Dickens, William T., "Error Components in Grouped Data: Is It Ever Worth Weighting?," *The Review of Economics and Statistics*, May 1990, *72* (2), 328–333.

Dickert-Conlin, Stacy and Todd Elder, "Suburban Legend: School Cutoff Dates and the Timing of Births," *Economics of Education Review*, December 2010, *29* (5), 826–841.

Dynarski, Susan, Joshua Hyman, and Diane Whitmore Schanzenbach, "The Effect of Class Size on Postsecondary Enrollment, Persistence and Completion," *(Unpublished manuscript)*, 2011.

Elder, Todd E. and Darren H. Lubotsky, "Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers," *The Journal of Human Resources*, 2009, *44* (3), 641–683.

Freeman, Richard B., "How Much Has De–Unionization Contributed to the Rise in Male Earnings Inequality?," NBER Working Paper 3826, National Bureau of Economic Research, Cambridge, MA. August 1991.

Genda, Yuji, Ayako Kondo, and Souichi Ohta, "Long-Term Effects of a Recession at Labor Market Entry in Japan and the United States," *The Journal of Human Resources*, Winter 2010, *45* (1), 157–196.

Gibbons, Robert and Michael Waldman, "Enriching a Theory of Wage and Promotion Dynamics inside Firms," *Journal of Labor Economics*, 2006, *24* (1), 59–107.

Gormley, Jr., William T. and Ted Gayer, "Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program," *The Journal of Human Resources*, Summer 2005, *40* (3), 533–558.

Grossman, Pan, Donald Boyd, Michelle Brown, Julie Cohen, Hamilton Lankford, Susanna Loeb, Dan Mindich, Sinead Mullen, and James Wyckoff, "Measure for Measure: A Pilot Study Linking English Language Arts Instruction and Teachers' Value-Added to Student Achievement," Working Paper 16015, National Bureau of Economic Research, Cambridge, MA May 2010.

Hagedorn, Marcus and Iourii Manovskii, "Spot Wages over the Business Cycle?," *(Unpublished Manuscript)*, August 2009.

Harris, D. N., S. Rutledge, W.K. Ingle, and C. Thompson, "When Supply Meets Demand: Principal Preferences and Teacher Hiring," in "Paper Prepared for the 2006 Annual Meeting of the American Educational Research Association" 2006.

Harris, Milton and Bengt Holmstrom, "A Theory of Wage Dynamics," *The Review of Economic Studies*, July 1982, *49* (3), 315–333.

Heckman, James J. and Robert Robb, "Using Longitudinal Data to Estimate Age, Period and Cohort Effects in Earnings Equations," in S. Feinberg and W. Mason, eds., *Analyzing Longitudinal Data for Age, Period and Cohort Effects*, New York: Academic Press, 1985.

_ , **Seong Hyeok Moon, Rodrigo Pinto, Peter Savalyev, and Adam Yavitz**, "A New Cost-Benefit and Rate of Return Analysis for the Perry Preschool Program: A Summary," Working Paper 16180, National Bureau of Economic Research, Cambridge, MA July 2010.

Hershbein, Brad, "Graduating in a Recession: High School Women's Labor Supply, Education and Home Production," *(Unpublished Manuscript)*, April 2010.

Jacob, Brian A., "Do Principals Fire the Worst Teachers?," Working Paper 15715, National Bureau of Economic Research, Cambridge, MA February 2010.

_ , "The Effect of Employment Protection on Worker Effort: Evidence from Public Schooling," Working Paper 15655, National Bureau of Economic Research, Cambridge, MA February 2010.

_ **and Lars Lefgren**, "Principals As Agents: Subjective Performance Measurement in Education," *Journal of Labor Economics*, 2008, *26* (1), 101–136.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan, "Earnings Losses of Displaced Workers," *The American Economic Review*, September 1993, *83* (4), 685–709.

Kahn, Lisa B., "The Long-Term Labor Market Effects of Graduating from College in a Bad Economy," *Labour Economics*, April 2010, *17* (2).

Kane, Thomas J., Eric S. Taylor, John H. Tyler, and Amy L. Wooten, "Identifying Effective Classroom Practices Using Student Achievement Data," Working Paper 15803, National Bureau of Economic Research, Cambridge, MA March 2010.

Katz, Lawrence and Kevin Murphy, "Changes in Relative Wages, 1963-1987 – Supply and Demand Factors," *Quarterly Journal of Economics*, February 1992, *107* (1), 35–78.

King, Miriam, Steven Ruggles, J. Trent Alexander, Sarah Flood, Katie Genadek, Matthew B. Schroeder, Brandon Trampe, and Rebecca Vick, "Integrated Public Use Microdata Series," in "Current Population Survey: Version 3.0. [Machine-readable database]" 2010.

Lee, David S., "Wage Inequality in the United States During the 1980s: Rising Dispersion or Falling Minimum Wage?," *Quarterly Journal of Economics*, August 1999, *114* (3), 977–1023.

Lemieux, Thomas, "Increasing Residual Wage Inequality: Composition Effects, Noisy Data or Rising Demand for Skill?," *American Economic Review*, June 2006, *96* (3), 461–498.

Ludwig, Jens and Deborah A. Phillips, "The Benefits and Costs of Head Start," Working Paper 12973, National Bureau of Economic Research, Cambridge, MA March 2007.

Magnuson, Katerine A., Christopher Ruhm, and Jane Waldfogel, "Does Prekindergarten Improve School Preparation and Performance," *Economics of Education Review*, 2007, *26*, 33–51.

McEwan, Patrick J. and Joseph S. Shapiro, "The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates," *The Journal of Human Resources*, Winter 2008, *43* (1), 1–29.

Miller, R. T., R.J. Murnane, and J. B. Willet, "Do Worker Absences Affect Productivity? The Case of Teachers," *International Labour Review*, 2008, *147* (1), 71–89.

Moulton, Brent R., "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, 1986, *32*, 385–397.

Oreopoulos, Philip, Till von Wachter, and Andrew Heisz, "The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates," IZA Discussion Paper 3578, The Institute for the Study of Labor, Bonn, Germany June 2008.

Oyer, Paul, "Initial Labor Market Conditions and Long-Term Outcomes for Economists," *The Journal of Economic Perspectives*, Summer 2006, *20* (3), 143–160.

_ , "The Making of an Investment Banker: Stock Market Shocks, Career Choice, and Lifetime Income," *The Journal of Finance*, December 2008, *63* (6), 2601–2628.

Puma, Michael, Stephen Bell, Ronna Cook, Camilla Heid, and Michael Lopez, "Head Start Impact Study: First Year Findings," Technical Report, U.S. Department of Health and Human Services, Administration for Children and Families, Washington, DC June 2005.

Rockoff, Jonah, "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *The American Economic Review*, 2004, *94* (2), 247–252.

_ , Brian A. Jacob, Thomas J. Kane, and Douglas O. Staiger, "Can You Recognize and Effective Teacher When You Recruit One?," Working Paper 14485, National Bureau of Economic Research, Cambridge, MA November 2008.

Rogers, A. M. and J. J. Stoeckel, "NAEP Mathematics and Reading (1971-2004) Long-Term Trend Restricted-Use Data Files Data Companion," Technical Report, U.S. Department of Education and Institute of Education Sciences and National Center for Education Statistics, Washington, DC December 2007.

Schmieder, Johannes F. and Till von Wachter, "Does Wage Persistence Matter for Employment Fluctuations? Evidence from Displaced Workers," *American Economic Journal: Applied Economics*, July 2010, *2* (3), 1–21.

The New Teacher Project, "The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness," 2009.

Welch, Finis, "Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Boom," *The Journal of Political Economy*, October 1979, *87* (5), S65–S97.