

Student Abilities During the Expansion of US Education*

Lutz Hendricks[†]

Todd Schoellman[‡]

January 2014

Abstract

The US experienced two dramatic changes in the structure of education in a fifty year period. The first was a large expansion of educational attainment; the second, an increase in test score gaps between college bound and non-college bound students. This paper documents the impact of these two trends on the composition of school groups by ability and the importance of these composition effects for wages. The main finding is that there is a growing gap between the abilities of high school and college-educated workers that accounts for one-half of the college wage premium for recent cohorts and for the entire rise of the college wage premium between the 1910 and 1960 birth cohorts.

JEL: I2, J24.

Key words: Education. Ability. Skill premium.

*For helpful comments we thank our editor, an anonymous referee, Berthold Herrendorf, Richard Rogerson, Guillaume Vandenbroucke, and seminar participants at the Federal Reserve Banks of Atlanta and Cleveland, North Carolina State University, the University of Georgia, the University of Iowa, the University of Pittsburgh, the Clemson University Bag Lunch, the Triangle Dynamic Macro Workshop, the 2009 Midwest Macroeconomic Meeting, the 2009 NBER Macroeconomics Across Time and Space Meeting, the 2009 North American Summer Meeting of the Econometric Society, the 2009 Society for Economic Dynamics, and the 2010 Canadian Macro Study Group. The usual disclaimer applies.

[†]University of North Carolina, Chapel Hill. Address: UNC at Chapel Hill, Department of Economics, 6C Gardner Hall, CB 3305, Chapel Hill, NC 27599-3305. Phone: (919) 966-5328. E-mail: lutz@lhendricks.org.

[‡]Arizona State University. Contact: Department of Economics, W.P. Carey School of Business, 501 E. Orange Street, Tempe, AZ 85287. Phone: (480) 965-7995. E-mail: todd.schoellman@gmail.com.

1 Introduction

The twentieth century witnessed an extraordinary and well-documented expansion of education in the United States (Goldin and Katz, 2008). Figure 1a illustrates this trend. For the birth cohorts born every ten years between 1910 and 1960, it displays the fraction of white men in four exhaustive and mutually exclusive education categories: high school dropouts (<HS), high school graduates (HS), those with some college but not a four-year degree (SC), and college graduates with at least a four-year degree (C+). Of the men born in 1910, only one-third finished high school. By the 1960 cohort, high school graduation had become nearly universal and the median man attended at least some college.¹

At the same time that high school completion and college enrollment were expanding, there was also a systematic and less well-known change in who pursued higher education. The general trend was for education to become more meritocratic, with ability and preparation becoming better predictors of educational attainment. This paper builds on the earlier work of Taubman and Wales (1972) to provide systematic evidence by comparing the standardized test scores for those who stop their education with a high school degree (the HS group) and those who continue to college (the SC and C+ groups). Figure 1b plots the average percentile rank of these two groups against the birth cohort; as is explained in Section 2, each pair of data points represents the results of a separate study. The trend is striking. For the very earliest cohorts, college-bound high school seniors scored just ten percentage points higher than non-college-bound seniors; by the 1940s cohorts, that gap had grown to nearly thirty percentage points.

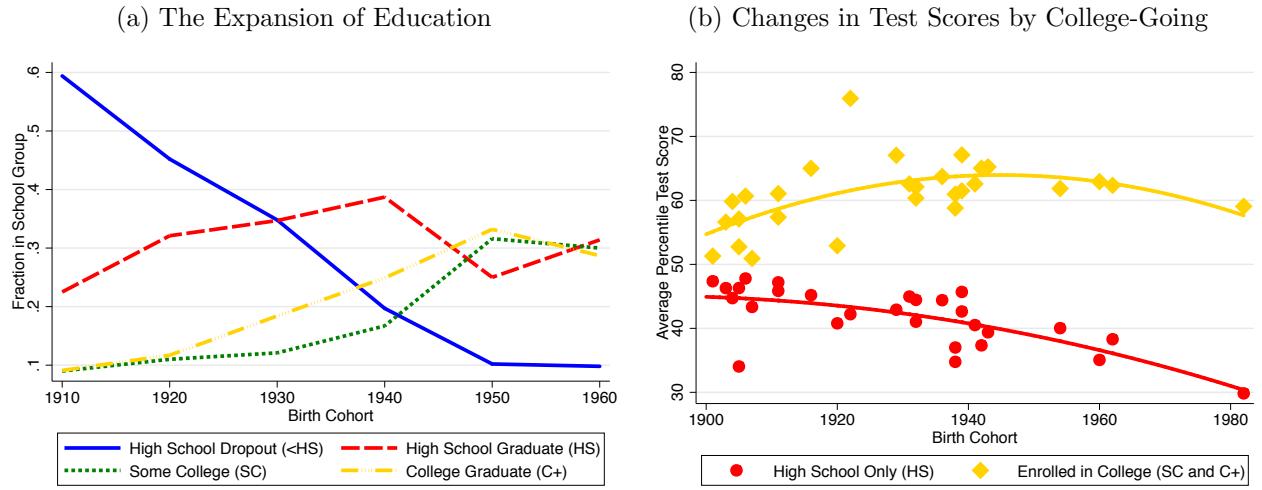
The main idea of this paper is that these trends are likely to have changed the ability composition of students by educational attainment. For example, it is unlikely that the ability of high school dropouts is the same for the 1910 and 1960 cohorts, given that more than half of the 1910 cohort dropped out but less than ten percent of the 1960 cohort did. Likewise, the ability of college graduates is likely to have changed given the large expansion of college enrollment and the changes in how college students are selected.

The primary motivation for studying composition effects is to understand their importance for the evolution of wage patterns over the course of the twentieth century. The main results concern two well-known features of the college wage premium. First, the college wage premium rose by 15 percentage points between the 1910 and 1960 cohorts.² Second,

¹These data are derived from forty-year olds in the 1950–2000 population censuses. Throughout this paper, “1910–1960 cohorts” can be read alternatively as “forty-year olds in years 1950–2000”. Descriptive facts use cohorts born at ten year intervals to match with the ten year intervals between censuses. For more details on the construction of the data in figure 1, see the appendix.

²The reported magnitude of the rise varies considerably in the literature, for at least two reasons. First,

Figure 1: Changes in US Education in the Twentieth Century



Note: Figure (a) shows the fraction of each cohort obtaining less than a high school degree (<HS), a high school degree only (HS), some college (SC), or a college bachelor's or more (C+). Figure (b) shows the mean percentile test score of those who obtain a high school degree (HS) as compared to those who start college (SC and C+, pooled).

the current college wage premium is 50 percentage points, which is difficult to reconcile with the low college completion rate in human capital models.³ This paper finds that changes in the composition of student abilities by educational attainment between the 1910 and 1960 cohorts can explain quantitatively the entire rise in the college wage premium while simultaneously making it easier to reconcile the current college wage premium with human capital theory.

To fix ideas, the average log-wage of workers with a particular educational attainment is considered to be a function of the price of skills specific to that education group and the quantity of those skills the average worker provides. The quantity is in turn determined by workers' cognitive abilities and the human capital they acquire over the course of their lives. Much of the previous literature seeking to explain the college wage premium holds the quantity of skills fixed and focuses on reasons why skill prices may have changed –

studies define the “high school” and “college” groups differently. The rise is larger if “college” includes advanced degrees, or if “high school” includes high school or college dropouts (Goldin and Katz, 2008; Heathcote et al., 2010). Second, the rise in the skill premium is affected by the age or potential experience at which it is measured. The rise was greater for younger or less experienced workers than for middle-aged workers (Autor et al., 2008).

³See for example Heckman et al. (2006) and Heckman et al. (2008), who also propose an alternative explanation.

for example, due to skill-biased technological change.⁴ The analysis in this paper allows either component of wages to change. The primary challenge is that while mean wages are observed directly, the other terms – skill prices, human capital, and ability – are not. This problem is addressed through the use of standardized test scores, which are treated as observed, noisy proxies for cognitive ability. Test scores make it possible to disentangle the role of cognitive ability from the other two factors. The methodology used does not allow one to separate skill prices from human capital.

A simple model of school choice with heterogeneous ability formalizes the challenge. The model also shows that the quantitative impact of composition effects on wages is controlled by two parameters. The first governs how strongly sorted the different school groups are by ability; more sorting means larger gaps in mean ability between school groups. The second parameter governs the mapping from ability to wages; a higher value for this parameter means that a given gap in mean abilities has larger implication for wages. The model is taken to the data in two steps.

First, the model is calibrated to the National Longitudinal Study of Youth 1979 (NLSY79) (Bureau of Labor Statistics; US Department of Labor, 2002). The NLSY79 is a representative sample of cohorts born around 1960 that includes information on their wages, education, and test scores. The NLSY79 provides two key moments: the relationship between wages and test scores, and the degree of educational sorting by test scores. An introductory analysis follows the previous economic literature and considers the special case where test scores measure cognitive ability exactly (Heckman et al., 1998; Garriga and Keightley, 2007). In this case the two moments from the NLSY79 identify the two key parameters of the model in a straightforward manner and the importance of composition effects can be computed directly. However, evidence from the psychometric literature establishes that test scores are likely a noisy measure of cognitive ability. This evidence is used to provide plausible bounds on the amount of noise in test scores and to calibrate the model. In this case the differences in mean ability between college and high school graduates likely account for half of the observed college wage premium.

The second step is to calibrate the model to fit the historical changes in schooling and test scores from figure 1. The main result of the paper is that the mean ability of college graduates relative to high school graduates rose by 14 percentage points. Thus, composition effects are sufficient to explain almost all of the rise of the college wage premium between the 1910 and 1960 cohorts (that is, between the years 1950 and 2000). Decompositions of

⁴This is the view espoused in Katz and Murphy (1992), Bound and Johnson (1992), Autor et al. (1998), and Goldin and Katz (2008). Bound and Johnson (1992) and the survey of Levy and Murnane (1992) propose other explanations including international trade or migration.

the results show that the expansion of education and the increase in sorting each explain about half of the total result. Finally, a number of robustness checks establish that these results are unlikely to be reduced by more than half with plausible changes in the model or the calibration targets.

This paper is most closely related to two existing literatures. First, the empirical findings on changes in the relationship between test scores and educational attainment over time builds on prior work by Finch (1946) and particularly Taubman and Wales (1972). The latter paper documented the spread in test scores between college-bound and non-college-bound high school seniors. This finding seems to have been largely forgotten, likely because it was published at a time when the college wage premium was declining, obscuring any possible link between test scores and wage patterns. In addition to returning attention to this important finding, this paper includes substantial work to increase the number of data points and the documentation of these trends.

Second, this paper is related to a literature that decomposes observed changes in educational wage differences into the underlying changes in skill prices and skill quantities. The fundamental challenge this literature faces is that neither skill prices nor skill quantities are directly observed. The literature has addressed this problem in a variety of ways.

A number of studies specify models of wage determination that motivate regressing wages or skill premiums on cohort education as a proxy for cohort quality (Juhn et al., 2005; Kaymak, 2009; Carneiro and Lee, 2011). Juhn et al. (1993) and Acemoglu (2002) use differences in wage growth between cohorts to eliminate cohort and age effects, thus identifying skill price changes. A final set of papers draws on models of human capital accumulation to disentangle skill prices from skill quantities. Laitner (2000) formulates a model that qualitatively generates predictions for relative wages and wage inequality consistent with post-war U.S. data. However, he does not attempt to quantify the implications of the model. Bowlus and Robinson (2012) estimate time series of skill prices for four school groups using the flat spot method developed by Heckman et al. (1998).⁵

In spite of the small number of studies, the approaches and findings are quite diverse. While a number of studies find that the expansion of education led to a modest reduction in the college wage premium (Juhn et al., 2005; Carneiro and Lee, 2011), other studies infer a sizable increase (Kaymak, 2009; Bowlus and Robinson, 2012). The diversity of findings is an indication that additional data may be needed to solve the identification problem associated with decomposing wages into skill prices and quantities. This paper

⁵Also related is Carneiro and Lee (2009) who estimate the effect of a counterfactual expansion of college enrollment among students born around 1960 using a local instrumental variable approach.

incorporates such evidence.

This paper provides new data measuring the cognitive abilities of cohorts born between 1901 and 1982. These data document a widening gap in test scores between college-bound and non-college-bound high school seniors. The implications of this widening gap for wage premiums is quantified using a transparent model. This approach conveys two benefits relative to the literature. First, test score data directly measure how at least one aspect of cohort quality changes over time. They suggest potentially important changes in the composition of education groups over time. Second, the new data cover a long period (the 1901–1982 birth cohorts) in a consistent way. The longer coverage is important because the data indicate that the largest changes in the test score gap occurred before the 1930 cohort, with the rate of change slowing over time.

This approach is designed primarily to quantify the importance of changing test score gaps as a proxy for changing cognitive ability gaps. It follows that the results do not quantify abilities that are uncorrelated with test scores or are unobserved altogether. Interested readers should consult the existing literature especially on changes in the price and quantity of unobserved abilities. That literature has not reached a consensus on whether these changes contribute to the rise in the college wage premium in an important way (Chay and Lee, 2000; Taber, 2001; Deschênes, 2006). It is possible that such changes may accentuate or partly undo this paper’s conclusion about composition effects.

The rest of the paper is organized as follows. Section 2 briefly gives details on the rising test score gap between college-bound and non-college-bound high school seniors. Section 3 introduces the model of school choice. Section 4 calibrates the model to the NLSY79 and derives cross-sectional results. Section 5 calibrates the model to the time series data and derives further results. Section 6 provides robustness checks and the final section concludes.

2 Test Scores and College Attendance

The first contribution of this paper is to provide extensive documentation on the divergence of test scores between college-bound and non-college-bound high school seniors. The main source of data is two dozen studies conducted by psychologists and educational researchers around the country. The goal of this section is to provide a brief overview of the content of these studies and how they are combined with results from the more recent, nationally representative samples such as the NLSY79 to generate Figure 1b. The appendix contains a longer description of the procedures along with references, detailed metadata on the different studies, and a number of robustness checks.

The starting point was to collect all known studies with data on the test scores of high school graduates who do and do not continue to college. The most useful studies are those that predate the availability of large, nationally representative data sets such as the NLSY. The first such studies were conducted shortly after World War I and tested students who were born just after the turn of the century.⁶ This paper's findings aggregate the results from more than two dozen such studies. The studies vary in terms of size, geographic scope, test instrument, and so on, but it is useful to describe a typical study, which comes in two parts. First, the researcher would arrange for a large number of high schools in a metropolitan area or a state (sometimes all such high schools) to administer an aptitude or achievement test to high school seniors. Second, the researcher would collect information on the college-going behavior of the students, either by asking them their plans as high school seniors, or by re-surveying the students, their parents, or schools a year or two later after their graduation. The key information of interest is cross-tabulations of test scores and college-going behavior.

Since many of these studies are quite old the original raw data do not exist. Instead, it is necessary to rely on the reported summary statistics and tabulations from published articles, mimeographs, books, and dissertations. One commonly reported table gives the number of students with scores in various ranges that did and did not continue to college. Following Taubman and Wales (1972), level scores are converted into percentiles. The average percentile rank of those who do and do not continue to college is then computed from these discretized distributions. This measure can be computed from most of the studies. Similar results can also be calculated using the raw micro data from the recent, nationally representative samples, including the NLSY79.

The resulting data are plotted in Figure 1b. The trend is striking. For cohorts born around the turn of the 20th century there was a very small test score gap between college-bound and non-college-bound high school seniors, on the order of 10 percentage points. The earliest studies expressed consistent surprise at how many low-scoring students continued to college and how many high-scoring students did not. The gap between the two groups grew steadily from the 1900 to the 1940 cohort, at which point it plateaued at nearly 30 percentage points.⁷ Contemporary sources pointed to two reasons why the gap was growing. First, it became increasingly common for universities to administer tests to applicants as

⁶The U.S. Armed Forces made heavy use of group intelligence tests in assigning recruits to positions during World War I. Their use in this context increased awareness and interest among the public and researchers, and provided an opening for their broader acceptance and adoption. Hence, the first studies were conducted immediately after the War (Cremin, 1961).

⁷This finding is closely related to that of Hoxby (2009), who documents a complementary trend of increasing sorting of students by test scores among colleges.

an admissions tool. Second, high schools administered tests to their students with an aim towards vocational guidance. Since test scores were often interpreted as measures of academic ability, students who scored well were encouraged to continue their education while those who did not were pushed towards vocational tracks.

The appendix includes documentation of the robustness of this basic finding. A similar pattern is demonstrated for alternative metrics of how strongly sorted college goers and non-goers are. References and evidence are provided to document that the tests used in early years appear to have been of quality similar to those from more recent years, as measured by inter-test correlations or the usefulness of tests for predicting subsequent college grades. Finally, the results are shown to be robust to the methodological details of the underlying studies, such as when they followed up with students or where the survey was conducted. The testing movement appears to have had a strong influence on who chose to stop their education with high school and who pursued college. The next section includes a model to analyze the importance of this composition change for wage patterns.

3 A Model of School Choice

The next goal is to specify a parsimonious model of school choice that formalizes the intuition from the introduction. The model is useful to clarify that the quantitative importance of composition effects for wage patterns depends on two key parameters. The model guides the subsequent empirical work.

The basic environment is a discrete time overlapping generations model. Each year a cohort of unit measure is born. Individuals are indexed by their year of birth τ as well as their age v , with the current period given by $\tau + v - 1$. Individuals live for a fixed T periods.

3.1 Endowments

Each person is endowed with a variety of idiosyncratic, time-invariant traits that affect their wages and schooling.⁸ These traits are captured by a two-dimensional endowment (a, p) that is drawn from a bivariate normal distribution. a represents ability. Ability is useful for both work and school, because it makes it easier to learn and process new information or perform new tasks. p represents the taste for schooling. It is a preference parameter that captures the relative disutility that a person derives from spending time in school instead of

⁸These traits may be malleable earlier in life. This paper addresses school choices made from young adulthood onward. Many of the relevant traits appear to difficult to change by age 16.

working. We impose several distributional assumptions: $\mathbb{E}\{a\} = \mathbb{E}\{p\} = 0$, $Var(a) = 1$, and $Cov(a, p) = 0$. Given our model structure, these assumptions are all without loss of generality. We denote the standard deviation of tastes for schooling by $\sigma_{p,\tau}$, which is allowed to vary over time.

3.2 Preferences

Let $c(q, v)$ denote the consumption of a person of type q at age v , and let $\beta > 0$ be the common discount factor. Then lifetime utility is given by:

$$\sum_{v=1}^T \beta^v \log[c(q, v)] - \exp[-(p + a)]\chi(s, \tau). \quad (1)$$

Workers value consumption in the standard way. They also place a direct utility value on their time spent in school, which is determined by the interaction between a worker-specific component $(p + a)$ and a cohort and school-specific component $\chi(s, \tau)$. The former term captures how enjoyable (p) and easy (a) a particular individual finds schooling to be. The functional form $-\exp[-(p + a)]$ assumes that school is distasteful, but less so for more cognitively able students or those with higher taste for schooling. The latter term captures how desirable school type s and its associated career paths are for cohort τ . It varies by cohort to capture changes in school and work, such as the amount of studying required to succeed in college or the career paths open to those with a particular educational attainment. χ is restricted to be positive and increasing in s . In this case, the preferences show complementarity between school and cognitive ability or taste for school. This complementarity is essential for the results. Alternative functional forms that preserve complementarity would yield similar results. This functional form is chosen as the simplest.

3.3 Budget Constraint

School type s takes $T(s)$ years to complete. While in school, students forego the labor market. After graduation, workers receive earnings $w(s, q, v)$ that depend on their school attainment, age, and ability. Their budget constraint requires them to finance lifetime consumption through lifetime earnings,

$$\sum_{v=1}^T \frac{c(q, v)}{R^v} = \sum_{v=T(s)+1}^T \frac{w(s, q, v)}{R^v}, \quad (2)$$

where R is the exogenous interest rate.

In keeping with much of the literature, it is assumed that workers with different educational attainments provide different labor inputs.⁹ Wages are given by

$$\log[w(s, q, v)] = \theta a + z(s, \tau + v - 1) + h(s, v). \quad (3)$$

The three underlying determinants of wages are given on the right-hand side of the expression. As mentioned before, ability affects wages directly. Since ability is assumed to be distributed standard normal, θ is an important parameter. It measures the increase in wages that comes from a one standard deviation rise in ability. $z(s, \tau + v - 1)$ is the price per unit of type s labor supplied by cohort τ at age v . Finally $h(s, v)$ captures the human capital accumulated by workers of education s at age v through experience or learning-by-doing.

3.4 Characterization of School Choice

Workers choose their school attainment s and a consumption path $c(q, v)$ to maximize preferences (1) subject to their budget constraint (2). The solution is characterized in two steps: first, the optimal allocation of consumption over time given school choice; and second, the school choice that maximizes lifetime utility.

Consumption in this model satisfies the standard Euler equation, $c(q, v+1) = \beta R c(q, v)$. When combined with the budget constraint and plugged into the utility function, this yields an expression for lifetime utility:

$$\theta a \sum_{v=1}^T \beta^v + \sum_{v=1}^T \beta^v \log \left[\frac{R(\beta R)^{v-1}}{\sum_{u=1}^T \beta^{u-1}} \sum_{u=T(s)+1}^T \frac{e^{h(s, u) + z(s, \tau + u - 1)}}{R^u} \right] - \exp[-(p + a)] \chi(s, \tau). \quad (4)$$

This equation has three additive terms. The first term captures the effect of ability on lifetime utility: higher ability allows for higher lifetime consumption. The second term captures the impact of school attainment on lifetime utility: more schooling means fewer years in the labor market but also changes the skill price and the rate of human capital accumulation. Finally, the last term captures the direct utility effect of schooling.

A key property of the model is that school choices depend only on the sum $p + a$, and not

⁹For example, this setup is consistent with the literature that allows high school and college-educated workers to be imperfect substitutes in aggregate production. However, it is not necessary for the results of this paper to take such a stand. One channel that is ruled out implicitly is that the rising skill premium may reflect an increase in the rental price of high ability labor relative to low ability labor (Juhn et al., 1993; Murnane et al., 1995).

on other individual-specific attributes or on p or a independently.¹⁰ To see this, note that the first term of the indirect utility function depends on ability but does not interact with school choices, so that it drops out of the individual's optimization problem. The second term does not depend on p or a . So endowments interact with school choice only through the third term, which includes the linear combination $p + a$. This model includes the common property that ability does not affect school choice through the earnings channel, because it raises both the benefits of schooling (higher future wages) and the opportunity cost (higher foregone wages today) proportionally. Instead, ability, tastes, and school choice interact through preferences in the third term. Given the assumptions on $\chi(s, \tau)$, school attainment in the model is increasing in $p + a$. The individuals who have the highest combination of $p + a$ will choose college; those with intermediate values will choose high school graduation or some college; and those with the lowest values will choose to drop out of high school.

Since ability is one component of the sum $p + a$, the model generates positive but imperfect sorting by ability into school attainment. Further, since the standard deviation of ability is normalized to 1, the degree of sorting by ability into educational attainment is controlled by a single parameter, $\sigma_{p,\tau}$. As $\sigma_{p,\tau}$ rises, more of the variation in $p + a$ comes from variation in p . In this case, workers are less sorted by ability across school groups and mean ability gaps are smaller. In the limiting case of $\sigma_{p,\tau} = \infty$, educational choices are explained entirely by tastes for schooling. In this case, $E(a|s) = E(a) = 0$ for all school groups.

3.5 Implications for Mean Ability and Wages

Since the model allows for positive sorting by ability, it generates composition effects that matter for wages. The average wage of workers from cohort τ with education s at age v is given by:

$$E[\log(w)|s, \tau, v] = \theta E[a|s, \tau] + z(s, \tau + v - 1) + h(s, v). \quad (5)$$

Average wages are affected by three terms: by $\theta E[a|s, \tau]$, which governs the importance of composition effects for wages; by skill prices, z ; and by human capital, h . The goal is to separate out the role of composition effects in explaining wage patterns from the other two terms. No attempt is made to separate out skill prices from human capital endowments in

¹⁰This is the model property that makes it innocuous to assume that p and a are independent. If they were correlated, one could always re-define a as ability plus the correlated component of tastes, and p as the orthogonal component of taste; given that outcomes depend only on $p + a$, the results will be the same.

this paper.

The quantitative importance of composition effects for wages depends on two key model parameters. The first is $\sigma_{p,\tau}$, which determines the strength of sorting by ability into different school groups, which is reflected in $E[a|s, \tau]$ in the average wage equation. The second is θ , which determines the impact of ability on wages. In general, the smaller is $\sigma_{p,\tau}$ and the larger is θ , the larger is the quantitative role for mean ability in explaining observed wage patterns. Other parameters such as β or R matter little or not at all for the quantitative results. Perfect sorting by $p + a$ is critical for this simplification.

3.6 Model Discussion

The model admits other interpretations that yield similar results. One useful reinterpretation follows Manski (1989). Students still possess ability a , which makes school easier and raises wages, just as in the baseline interpretation. However, students have no tastes for schooling. If they knew their own ability, they would perfectly sort by ability into school attainment. Imperfect sorting in this model comes from the assumption that students are imperfectly informed about their ability, with p representing signal noise and $p + a$ representing their signal of their own ability. Students with better signals of ability further their education, because they anticipate that schooling will be relatively painless. This reinterpretation generates the same prediction of perfect sorting by $p + a$. Because of this the calibration and results from this alternative model would be *identical* to those derived from the baseline model.

The model does assume only a single stand-in friction that prevents perfect sorting by ability. An alternative approach taken elsewhere is to model multiple frictions in detail (Cunha et al., 2005; Navarro, 2008). Doing so would complicate the model and identification. However, the primary impediment is a lack of sufficient historical data to calibrate multiple frictions in detail.

An alternative friction to perfect sorting by ability that is not nested by this setup is borrowing constraints. Borrowing constraints differ from tastes because they are asymmetric: they prevent some high-ability students from furthering their education, but have no effect on low-ability students. By contrast, variation in tastes causes some high-ability students to drop out, but it also causes some low-ability students to attain high levels of education. The literature has not arrived at a consensus about the quantitative importance of borrowing constraints. Cameron and Taber (2004) and Stinebrickner and Stinebrickner (2008) find no evidence of borrowing constraints in the United States for recent cohorts of college attendees. There is little evidence as to whether credit constraints were quan-

titatively important for earlier cohorts. The data documented in the appendix show that low-ability students are becoming less likely to attend college over time. This information is consistent with a decline in the dispersion of tastes, but not a model featuring only a relaxation of borrowing constraints over time.

The analysis focuses on measured ability and abstracts from changes in the price or relative quantity of unmeasured ability across school groups. Whether such changes account for a large part of the rise in the college wage premium remains controversial (Taber, 2001; Chay and Lee, 2000). The return to measured ability is assumed to be constant over time, consistent with the evidence presented in Bowles et al. (2001). Allowing the return to rise over time would increase the effect of diverging ability gaps on relative wages and thus reinforce our main result.

4 Composition Effects in the Cross-Section

The model is a parsimonious formalization of the basic challenge. Mean wages are affected by skill prices, human capital, and mean ability, none of which are directly observable. In the model, two key parameters determine how important mean ability is for explaining wage patterns. The first is $\sigma_{p,\tau}$, which determines the size of mean ability gaps between school groups; the second is θ , which determines the impact of ability on wages. The next step is to show how standardized test scores can help calibrate these parameters and quantify the role of ability for wage patterns.

The primary data source is the NLSY79. The NLSY79 has two properties that make it ideal for this analysis. First, it is a representative sample of persons born between 1957 and 1964. Second, it includes information about the wages, school choices, and AFQT test scores of individuals in the sample. Most other data sets are deficient along one of these dimensions. For example, information on SAT scores are drawn from a non-representative sample, while common data sets such as the population census do not include information on test scores.

The sample is restricted to white men throughout. Women are excluded for the typical reason that only a selected sample of women work. Further, the selection process itself may be changing over time. Minorities are excluded because earlier cohorts of minorities likely faced discrimination that limited school attainment choices, occupational opportunities, and wages. Members of the supplemental samples are included, with weights used to offset the oversampling of low income persons. Since everyone born in the NLSY79 is from a narrow range of cohorts, they are grouped together and jointly referred to as the 1960

cohort. The focus of this section is on an analysis of the 1960 cohort and some initial cross-sectional results. The next section includes time series results.

The test score used from the NLSY79 is the Armed Forces Qualifying Test (AFQT) score. The AFQT is widely recognized as a cognitive test and AFQT scores are highly correlated with the scores from other aptitude tests. Students did not take the AFQT at the same age, which affects average scores. Regressions are used to remove age effects from AFQT scores in the standard way. The age-adjusted AFQT score is then standard normalized. Real hourly wages at age 40 and educational attainment are also constructed for each person. Details are available in the appendix.

Since test scores play a central role in the analysis, it is important to be precise about their interpretation. Test scores are assumed to be noisy, scaled proxies for cognitive ability, $\hat{a} = \eta(a + \tilde{\varepsilon}_a)$, where η is an unknown scaling factor and $\tilde{\varepsilon}_a$ is a normal random variable with mean 0 and standard deviation σ_a . Standard normalizing test scores removes the scaling factor. Once standard normalized, test scores and the noise term are given by:

$$\hat{a} = \frac{a}{\sqrt{1 + \sigma_a^2}} + \varepsilon_a \quad (6)$$

$$\varepsilon_a \sim \mathcal{N}\left(0, \frac{\sigma_a}{\sqrt{1 + \sigma_a^2}}\right) \quad (7)$$

The next step is to use test scores to quantify the role of composition effects in explaining observed wage patterns.

4.1 Calibration When Test Scores Measure Ability Exactly

Test scores provide useful information on the role of ability in school choices and wages. Intuitively, the degree of sorting by test scores into educational attainment is a proxy for the degree of sorting by ability, which helps identify $\sigma_{p,1960}$. Likewise, the effect of test scores on wages is a proxy for the effect of ability on wages, which helps identify θ . To illustrate this process it is useful to begin with the special case $\sigma_a = 0$. In this special case, test scores measure ability exactly, which greatly simplifies identification and the construction of the results.

The first step is to identify θ . The wage generating process in the model is:

$$\log[w(s, q, v)] = \theta a + z(s, \tau + v - 1) + h(s, v). \quad (8)$$

Generally, the econometrician does not have direct information on a . Instead, the available

Table 1: Estimated Wage Return to School Attainment and Test Score

Dependent variable: log-wages	
Test Score ($\beta_{\hat{a}}$)	0.104 (0.017)
High School Graduate (γ_{HS})	0.17 (0.06)
Some College (γ_{SC})	0.35 (0.06)
College Graduate (γ_{C+})	0.69 (0.07)
Observations	1942
R^2	0.24

Note: Table gives point estimates from a regression of log-wages on school dummies and test score (AFQT) from the NLSY79. The omitted school category is high school dropout (<HS). Standard errors are reported in parentheses.

information is measured test scores. The empirical counterpart is then to regress wages at age 40 on test scores and a full set of school dummies:

$$\log(w) = \beta_{\hat{a}}\hat{a} + \sum_s \gamma_s d_s + \varepsilon_w. \quad (9)$$

\hat{a} is the individual's standard normalized test score, and $\beta_{\hat{a}}$ is the coefficient associated with that score. d_s is an indicator variable that takes a value of 1 if the individual has school attainment s . γ_s captures the joint wage impact of skill prices and human capital; it is not possible in this setup to distinguish the two. ε_w is assumed to be a normal random variable that captures factors such as shocks or luck that affect wages but are not associated with test scores, skill prices, or human capital.

Table 1 shows the results of a regression of log-wages on test scores as implemented in the NLSY79. The return to test scores is $\beta_{\hat{a}} = 0.104$. This will be the baseline estimate of the return to test scores for the remainder of the paper; what will change is the associated interpretation. In the case where test scores measures ability exactly, the interpretation is straightforward: $\beta_{\hat{a}} = \theta$. A one standard deviation rise in ability (which is the same as test score) raises log-wages by 10.4 percentage points. This is the first key parameter for determining the importance of composition effects.

The second feature of the data that is important for the results is the degree of sorting by ability into educational attainment. Table 2 provides evidence that school groups are strongly sorted by test scores. Each row of the table corresponds to one of the four school groups. The four columns give the conditional probability of someone with that school level having a test score in each of the four quartiles of the distribution. The vast majority (86%) of high school dropouts are from the first test score quartile, while 76% of high school graduates have below-median test scores. On the other hand, 88% of college graduates have

Table 2: Conditional Distribution of Test Scores Given Schooling

School Attainment	Test Score Quartile			
	1	2	3	4
High School Dropout (<HS)	86%	12%	2%	0%
High School Graduate (HS)	42%	34%	19%	5%
Some College (SC)	18%	32%	31%	19%
College Graduate (C+)	1%	11%	29%	59%

Note: Table shows test score quartile distribution given school attainment. Each row adds to 100%. Calculated based on school attainment and AFQT quartile from the NLSY79.

above-median test scores.

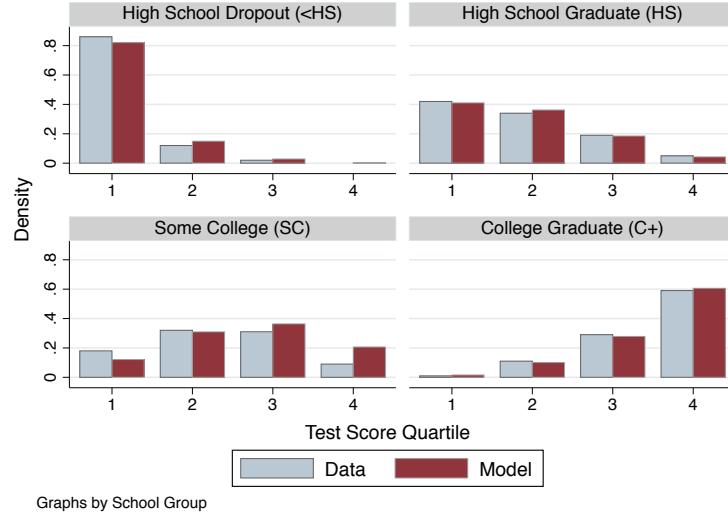
In the case where test scores measure ability exactly, these facts imply that school groups are strongly sorted by ability. It is again straightforward to use this information. In this special case, $E(a|s) = E(\hat{a}|s)$, which can be computed directly from the NLSY79. The role of composition effects in explaining wage premiums is given by $\theta [E(a|s) - E(a|s')] = \beta_{\hat{a}} [E(\hat{a}|s) - E(\hat{a}|s')]$.

An alternative approach is to calibrate the model to the NLSY79 and to report the composition effects implied by the calibrated model. At this point the exercise is not strictly necessary because, as highlighted above, estimates of $\beta_{\hat{a}}$ and $E(\hat{a}|s)$ are sufficient for these results. The goal of this exercise is to build some intuition for the general calibration procedure, which will be necessary for subsequent results. The exercise is also useful to show that in this case the calibrated model produces results very similar to the calculation above, which suggests that the calibration passes a basic test of reasonableness.

The calibration procedure is as follows. In this and all subsequent calibrations $\chi(s, 1960)$ are treated as a set of free parameters that can be varied to fit schooling by cohort exactly. This step is important because the quantitative results are very sensitive to getting educational attainment right. Thus it is useful to fit the attainment exactly and then to use the model as a measurement device to study the implied importance of composition effects. There are three remaining parameters that do not drop out of the model: $\sigma_{\hat{a}}$, θ , and $\sigma_{p,1960}$. It is assumed $\sigma_{\hat{a}} = 0$ throughout this section. The calibration chooses $\theta = 0.104$ so that the return to test scores in the model matches the same statistic in the data. Finally, $\sigma_{p,1960}$ is calibrated so that the model fits the sorting by test scores into educational attainment as closely as possible. The model is deliberately parsimonious, and yet this approach is quite successful. Figure 2 compares the sorting in the data and the sorting predicted by the model for the best fit of $\sigma_{p,1960} = 0.87$. The model is able to generate sorting quite comparable to the data. The only significant discrepancy is that the model-generated distribution for those with some college has too many people with above-average test scores and too few with below-average test scores. Otherwise the fit between model and data is quite close,

which suggests that the model will generate mean test score gaps comparable to the data.

Figure 2: Model-Predicted and Actual Distribution of Test Scores Given Schooling



Note: Figure shows the fraction of each school attainment group with test scores in quartiles 1–4. Data represents figures calculated using the AFQT in the NLSY79; model refers to the predictions of the calibrated model.

4.2 Results When Test Scores Measure Ability Exactly

Table 3 shows the cross-sectional results. Each row contains the comparison between the listed education group and the omitted group, high school graduates. Of these, the college-high school comparison is of greatest interest since the college wage premium has received so much attention in the literature. The second column gives the education wage premium measured from the US Census.¹¹ The third and fourth column give the contribution of composition effects to these wage premiums implied by the calculations described above and by the calibrated model. Differences in mean ability account for roughly one-quarter to one-third of the wage premiums, with slightly smaller results for the college wage premium.

These initial results help address the puzzle of Heckman et al. (2006) and Heckman et al. (2008). They find that in standard human capital models, the college wage premium for recent cohorts is difficult to reconcile with less than one-third of the recent cohorts graduating from college, unless one incorporates some substantial uncertainty or a large

¹¹The Census wage premiums are used to be consistent with the results of Section 5, which examines the changes in wage premiums over time and cohorts. It is not possible to calculate such changes using the NLSY79 because it lacks data from earlier cohorts. Details of the wage measurement are available in the appendix.

Table 3: Cross-Sectional Results when Test Scores Measure Ability Exactly

	Education Wage Premium Data	Contribution of Composition Effects	
		Direct Calculation	Calibrated Model
High School Dropout (<HS)	-0.24	-0.08	-0.08
Some College (SC)	0.18	0.06	0.07
College Graduate (C+)	0.52	0.14	0.15

Note: Table reports education wage premiums in the 1960 cohort and the fraction of each that can be attributed to composition effects. Data are taken from the census. Results are computed directly using NLSY79 data or derived from the calibrated model. Wage premiums are computed relative to high school graduates.

“psychic cost” of attending schooling. These results help reduce this puzzle modestly by pointing out that some of the apparently high college wage premium at age 40 is actually attributable to the gap in mean ability between college and high school graduates; the true private return to college is smaller than the observed wage gap.

4.3 Calibration When Test Scores Measure Ability With Noise

The initial results can be derived without calibrating the model. However, the model makes it possible to undertake two additional exercises. The first is to consider the case where test scores measure ability with noise. In this case, the mean test score for different school groups is not the same as the mean ability, so mean ability gaps cannot be measured directly. Instead, the calibrated model can be used to quantify the role of ability.

Before discussing the exact calibration procedure, it is useful to see why allowing for noise in test scores is likely to be important. The reason is that the log-wage return to test scores is used to identify θ . In the case where test scores measure ability exactly, then in fact $\theta = \beta_a = 0.104$. However, if test scores measure ability with noise, then the empirical regression suffers from attenuation bias, which implies $\theta > \beta_a = 0.104$. Hence, a given gap in mean abilities will lead to larger composition effects that will account for more of the observed wage premiums. The next goal is to quantify this statement: how much noise is there likely to be in test scores, and how much more important are composition effects?

The primary challenge of implementing a model where test scores are noisy is a lack of direct evidence on how well test scores measure ability. The obvious reason is that ability itself is not measured; if it were, it would not be necessary to use test scores as a proxy for ability. However, it is possible to make inferences that usefully bound the noise in test scores.

The lower bound on the noise in test scores is derived from the well-known property that repeatedly administering similar or even identical tests to a group yields positively correlated but not identical results. In particular, the lower bound is chosen so that a

given test score is not a better predictor of ability than it is of other subsequent test scores. To quantify this statement, recall that test scores are assumed to be noisy, scaled proxies for ability. In this context it is natural to think of the noise in tests $\varepsilon_{\hat{a}}$ as being an independent, test-specific draw. Then the correlation between two different test scores for a given individual is $(1 + \sigma_{\hat{a}}^2)^{-1}$. Herrnstein and Murray (1994, Appendix 3) document the correlation between AFQT scores and scores from six other standardized tests taken by some NLSY79 individuals. The correlations range from 0.71 to 0.9, with a median score of 0.81.¹² Cawley et al. (1997) show that the correlation between AFQT scores and the first principal component of the ASVAB scores is 0.83.

These correlations suggest that $(1 + \sigma_{\hat{a}}^2)^{-1} = 0.8$, which in turn implies a lower bound $\sigma_{\hat{a}} \geq 0.5$.¹³ If test scores measured ability more precisely, then the correlation between scores from different tests should be higher. This result is used by fixing $\sigma_{\hat{a}} = 0.5$ in the model, then calibrating θ and $\sigma_{p,1960}$ to fit the log-wage return to test scores and the school-test score sorting as well as possible. The former moment can be replicated exactly. It has already been shown that even with only a single parameter $\sigma_{p,1960}$ the calibrated model can replicate the school-test score sorting closely (Figure 2); that continues to be the case here and throughout the remainder of the paper. The remaining figures are not shown to conserve space.

It is also useful to derive an upper bound on the plausible noise in test scores. The purpose of this bound is not to argue that the true results are at some midpoint of the lower and upper bounds. Instead, the goal is to show that a bounding argument in this case is effective in the sense that the range of results between the lower and upper bounds is fairly narrow. The implication is that it is innocuous to focus on the lower bound as the benchmark results. Further, it is the case that the results at the lower bound are already large relative to the wage patterns in the data.

The upper bound is derived by imposing plausible limits on the size of composition effects. More noise in test scores implies a larger attenuation bias in the regression of wages on test scores, a larger value for θ , and larger composition effects. At some point the implied composition effects become implausibly large. One natural benchmark is that composition effects should not account for more than the corresponding education wage premium. If

¹²A slight complication arises from the fact that Herrnstein and Murray compute correlations between percentile ranks rather than raw scores. Simulations were conducted to verify that this has only a minor quantitative effect on the resulting correlation.

¹³A similar approach is taken by Bishop (1989) to estimate the measurement error in the PSID's GIA score. Based on the GIA's KR-20 reliability of 0.652, Bishop's result implies $\sigma_{\hat{a}} = 0.73$, which would imply a larger role for ability than what is found here. In fact, it exceeds the upper bound for $\sigma_{\hat{a}}$ derived below, suggesting that Bishop's results may have counterfactual implications for wages.

Table 4: Cross-Sectional Results when Test Scores Measure Ability With Noise

	Education Wage Premium Data	Contribution of Composition Effects	
		Model Without Noise	Model With Noise
High School Dropout (<HS)	-0.24	-0.08	-0.14 – -0.22
Some College (SC)	0.18	0.07	0.11 – 0.18
College Graduate (C+)	0.52	0.15	0.25 – 0.39

Note: Table reports education wage premiums in the 1960 cohort and the fraction of each that can be attributed to composition effects. Data are taken from the census. Results are derived from the calibrated model, which produces a single estimate if test scores measure ability exactly or a range of plausible results if test scores measure ability with noise. Wage premiums are computed relative to high school graduates.

they did, this would imply a negative private return to going to school longer, which would seem inconsistent with simple optimization on the part of the students who achieve that attainment in the data.

Implementing the upper bound requires an iterative algorithm. The first step is to guess a particular value of $\sigma_{\hat{a}}$. The second step is to calibrate θ and $\sigma_{p,1960}$ to fit the log-wage return to test scores and the school-test score sorting as well as possible given $\sigma_{\hat{a}}$. The third step is to compute composition effects $\theta(E[a|s] - E[a|s - 1])$ and compare them to wage differences $w(s) - w(s - 1)$. If $\theta(E[a|s] - E[a|s - 1]) < w(s) - w(s - 1)$ for each school level, then the algorithm returns to the first step with a larger guess for $\sigma_{\hat{a}}$ and repeats the process; if $\theta(E[a|s] - E[a|s - 1]) > w(s) - w(s - 1)$ for any school level, then the algorithm returns to the first step with a smaller guess for $\sigma_{\hat{a}}$ and repeats the process. The algorithm iterates until it finds the $\sigma_{\hat{a}}$ so that $\theta(E[a|s] - E[a|s - 1]) = w(s) - w(s - 1)$ for one school level and $\theta(E[a|s] - E[a|s - 1]) < w(s) - w(s - 1)$ for the remaining school levels.

4.4 Results When Test Scores Measure Ability With Noise

Table 4 summarizes the cross-sectional results implied by this bounding exercise. Each row contains the comparison between the listed education group and the omitted group, high school graduates. The second column gives the education wage premiums. The third column repeats the contribution of composition effects to these wage premiums implied by the calibrated model for the case where test scores measure ability exactly. The fourth column gives the range of results implied by the lower and upper bound of the calibrated model for the case where test scores measure ability with noise. When test scores contain noise, then the calibration procedure chooses a higher value of θ to match the observed log-wage return to test score. This higher value of θ in turn yields a larger role for composition effects, 57–75% higher than in the case where test scores measure ability exactly. Composition effects account for at least 48% of observed wage premiums, and more than half of the

wage premium for high school dropouts and those with some college. These large results go further towards reducing the puzzle that it is hard to reconcile the high college wage premium with a low college completion rate in a human capital model (Heckman et al., 2006, 2008).

The results implied by the lower and upper bound are fairly similar. In particular, the results for the upper bound are less than twice those for the lower bound. The next section includes a time series calibration of the model that implies an even lower upper bound, so that the difference between the lower and upper bounds is even smaller.

5 Composition Effects in the Time Series

The calibrated model serves two purposes in this paper. First, it provides results for the case where test scores measure ability with noise. Second, it provides results for the time series. While the NLSY79 provides excellent data on schooling, wages, and test scores for the 1960 cohort, no comparable data set exists for earlier cohorts. At the same time, the dramatic expansion of education and the growing test score gap between those who enroll in college and those who do not suggest that composition effects may play a large role in the wage patterns of the twentieth century. In this section the calibrated model is used to quantify such composition effects.

5.1 Calibration

The time series calibration follows the same basic outline as the cross-sectional calibration. The model is calibrated for both a lower bound and an upper bound for $\sigma_{\hat{a}}$, with parameters and results provided for each case.

The lower bound is still $\sigma_{\hat{a}} = 0.5$. Given this moment, the remaining parameters are chosen to fit the model to the data. In particular, $\chi(s, \tau)$ are chosen to fit the expansion of schooling in figure 1a. Similarly, $\sigma_{p, \tau}$ is chosen to fit the estimated quadratic trend in the degree of sorting by test score into educational attainment shown in figure 1b. The data show that students are becoming more strongly sorted over time. The model can replicate this observation if the dispersion of tastes is declining over time, so that ability plays a larger role in school choices for later cohorts.¹⁴ Finally, θ is chosen so that the model-predicted return to schooling matches that of the data, $\beta_{\hat{a}} = 0.104$.

¹⁴An alternative interpretation is that students are imprecisely informed about their own ability, but that they are becoming more precisely informed over time; see section 2.6 for further discussion.

Table 5: Calibrated Parameters for the Lower Bound

Parameter	Value
Noise in Test Scores (σ_a)	0.50
Effect of Ability on Wages (θ)	0.155
Dispersion of Preferences, 1960 Cohort ($\sigma_{p,1960}$)	0.62
Dispersion of Preferences, 1950 Cohort ($\sigma_{p,1950}$)	0.80
Dispersion of Preferences, 1940 Cohort ($\sigma_{p,1940}$)	1.12
Dispersion of Preferences, 1930 Cohort ($\sigma_{p,1930}$)	1.10
Dispersion of Preferences, 1920 Cohort ($\sigma_{p,1920}$)	1.28
Dispersion of Preferences, 1910 Cohort ($\sigma_{p,1910}$)	1.44

Note: Table gives full set of calibrated parameters for the lower bound case, where the noise in test scores is fixed at $\sigma_a = 0.50$.

An iterative procedure is used to find the upper bound again. That procedure works the same as in the previous section. In general, the upper bound is likely to be closer to the lower bound than was the case in the previous section. The reason is that the cross-sectional calibration only checks whether the implied composition effects are plausible for the 1960 cohort, whereas the time series calibration checks whether the implied composition effects are plausible for all cohorts between 1910 and 1960. The larger number of moments gives more wage premiums that may potentially bind.

5.2 Results for Lower Bound

The results for the lower bound are presented first and taken to be the benchmark findings. They are compared to the upper bound in the next section. Table 5 shows the full set of calibrated parameters for the lower bound. The value for θ is roughly fifty percent larger than in the case when test scores measure ability perfectly. The other main point is that the calibrated dispersion of tastes declined substantially between the 1910 and 1960 cohorts, indicating that ability played a much greater role in determining who continued to college for the 1960 cohort. Workers in this model sort perfectly by $p + a$ and the variance of a is set at 1 throughout. Ability variation accounted for just 32% of the variation in $p + a$ for the 1910 cohort, but 72% for the 1960 cohort.

These changes in sorting, along with the expansion of education, imply large changes in the mean ability of the four school groups. Figure 3 shows the model-implied evolution of the distribution of ability conditional on schooling. Figure 3a illustrates the degree of sorting found for the 1960 cohort in the NLSY79. There are clear differences in the mean of the ability distribution between each of the four school groups, and almost no overlap between the distributions for high school dropouts and college graduates.

Figure 3b illustrates a particular counterfactual: it shows the distribution of ability conditional on schooling that would have applied with the 1960 cohort's actual school at-

tainment but the counterfactual dispersion of tastes implied by the model for the 1910 cohort. Comparison of figures 3a and 3b illustrates the effect of the increase in sorting isolated from the expansion of education. The distributions in figure 3b have small mean differences, particularly for those who at least graduate high school. Further, the distributions overlap substantially.

Finally, figure 3c illustrates the model-implied distributions for the 1910 cohort. Comparison of figures 3b and 3c illustrates the effect of the expansion of education. The mean of each distribution is shifted left in figure 3b by the rise in schooling. To see why this happens, consider the distribution for high school graduates. Over time, attainment rises. In the model, this happens because high school graduates with relatively high levels of $p+a$ in later cohorts start attempting college. At the same time, some people with relatively low $p+a$ in later cohorts will complete high school instead of dropping out. Both of these effects act to reduce the average ability of high school graduates.

Comparison of figures 3a and 3c shows the combined effect of the expansion of education and the change in sorting. The leftward shift in ability for high school dropouts is particularly pronounced because both effects move in the same direction, toward a decline in ability. On the other hand there is hardly any change in the peak of the distribution for college graduates, as the expansion of college is in large part offset by the change in sorting. Intuitively, it is possible to expand college enrollment without lowering the mean ability of college graduates if stronger sorting by test scores makes it possible to identify high-ability students who in earlier cohorts did not attend college. This point is central to the findings.

The model produces two main sets of time series results. The first describes changes in the student abilities conditional on educational attainment for different cohorts. These results are presented in table 6. The measured time series for wages and model-implied time series for mean ability by school group are smooth, so it is sufficient to describe only the total change between the first and last cohorts. The second column gives the total log-wage growth for each school group between the 1910 and 1960 cohorts. It may be useful to recall that our empirical results are drawn from age 40 observations of each cohort, so these results correspond also to changes between years 1950 and 2000. The third column gives the model-implied role of composition effects. Given the observed wage growth and the model-implied change in mean ability, it is possible to back out the implied growth in $h+z$, which is given in the fourth column. Growth in $h+z$ is the growth in skill prices and human capital, which is also the wage growth that would have been observed if mean ability had remained constant. Changing ability had the largest effect for high school dropouts: the 17 percentage point decline in effective ability caused observed log wage growth to be

Table 6: Decomposition of Wage Growth by School Group, 1910–1960 Cohorts

	Wage Growth	Model-Implied Decomposition	
		Composition Effects	Skill Price and Human Capital Growth
High School Dropout (<HS)	0.22	-0.17	0.40
High School Graduate (HS)	0.29	-0.14	0.42
Some College (SC)	0.30	-0.08	0.38
College Graduate (C+)	0.43	0.00	0.44

Note: Table presents the change in wages by education level between the 1910 and 1960 cohorts and the fraction of each that can be attributed to composition effects versus changes in skill prices and human capital. Data are taken from the censuses. Results are derived from the calibrated model.

Table 7: Decomposition of Education Wage Premium Changes, 1910–1960 Cohorts

	Wage Premium Change	Model-Implied Decomposition	
		Composition Effects	Relative Skill Price and Human Capital Growth
High School Dropout (<HS)	-0.06	-0.03	-0.03
Some College (SC)	0.02	0.06	-0.04
College Graduate (C+)	0.15	0.14	0.01

Note: Table presents the change in skill premiums between the 1910 and 1960 cohorts and the fraction of each that can be attributed to composition effects versus changes in relative skill prices and human capital. Data are taken from the censuses. Results are derived from the calibrated model. Wage premiums are computed relative to high school graduates.

roughly one-half of the growth in $h+z$. The effect for high school graduates was smaller and for those with some college smaller still. For college graduates ability remained constant. Hence the model generates a wage slowdown that affects the less educated groups more.

The second set of results describe the importance of composition effects for the evolution of education wage premiums. These results are presented in table 7. The second column gives the change in each education wage premium between the 1910 and 1960 cohorts. The third column gives the model-implied role of composition effects, which here means changes in the relative ability of the given education group as compared to high school graduates between the 1910 and 1960 cohorts. Given these, it is possible to back out the implied growth in $h+z$ gaps, which is given in the fourth column. Changing relative ability had the largest effect for college graduates. In fact, almost the entire rise in the college wage premium can be attributed to composition effects, because the mean ability of college graduates remained roughly unchanged over this time while the mean ability of high school graduates declined. Half of the change in the high school dropout-high school graduate premium can be attributed to changes in mean ability for the two groups.

To summarize, the results suggest that changes in ability have slowed observed wage growth for most school groups as mean ability has declined. Further, the most important result is that the entire rise in the college wage premium can be explained by changes in the relative ability of college and high school graduates. This result is derived from a different methodology but arrives at a similar conclusion as Bowlus and Robinson (2012), who find

Table 8: Selected Results for Lower and Upper Bounds of Test Score Noise

Statistic	Data	Contribution of Composition Effects	
		Model Without Noise	Model With Noise (Baseline)
1960 College Wage Premium	0.52	0.15	0.25 – 0.37
1910–1960 High School Wage Change	0.29	-0.08	-0.14 – -0.21
1910–1960 College Wage Premium Change	0.15	0.08	0.14 – 0.20

Note: Table presents select wage statistics and the fraction of each that can be attributed to composition effects. Data are taken from the censuses. Results are derived from the calibrated model, which gives a range of plausible results.

that 72% of the rise in the college wage premium between the years 1980 and 1995 can be attributed to changes in the quantity of labor services provided by college relative to high school graduates. The next step is to compare these results to the upper bound and consider the full range of plausible results.

5.3 Range of Results

Application of the iterative procedure outlined above yields an upper bound for $\sigma_{\hat{a}} = 0.66$, as well as the associated remaining parameters and model results. Table 8 gives selected results for the lower and upper bound. These results are for the college wage premium for the 1960 cohort; the growth in wages for high school graduates between the 1910 and 1960 cohorts; and the change in the college wage premium between the 1910 and 1960 cohorts. The second column repeats the wage change observed in the data for each. The third column repeats the contribution of composition effects to each case implied by the model where test scores measure ability exactly. The fourth column gives the range of results implied by the lower and upper bound for the case where test scores measure ability with noise. The main message from this column is that the upper bound is now even closer to the lower bound; in each case, the implied upper bound is roughly fifty percent larger than the lower bound. This fact suggests that the bounding argument restricts the range of potential results successfully. The next step in the analysis is to decompose these results.

5.4 Decomposition of Results

The next experiment decomposes the relative role of the increase in sorting and the expansion of education in driving the results. To hold sorting fixed, the model is recalibrated with $\sigma_{p,\tau} = \sigma_{p,1960}$ for all cohorts. Other than holding $\sigma_{p,\tau}$ fixed, the details of the calibration are as in the baseline experiment. The lower bound is still given by $\sigma_{\hat{a}} = 0.50$, but the upper bound has to be recalibrated because fixed sorting implies different ability gaps for early cohorts.

Table 9: Decomposition of Selected Results: Role of Changes in Sorting

Statistic	Data	Contribution of Composition Effects	
		Baseline Model	Model with Constant Sorting
1960 College Wage Premium	0.52	0.25 – 0.37	0.25 – 0.27
1910–1960 High School Wage Change	0.29	-0.14 – -0.21	-0.16 – -0.18
1910–1960 College Wage Premium Change	0.15	0.14 – 0.20	0.08 – 0.09

Note: Table presents select wage statistics and the fraction of each that can be attributed to composition effects. Data are taken from the censuses. Results are derived from the calibrated model under the baseline case where sorting improves between cohorts or the alternative case where sorting is held fixed. Each case gives a range of plausible results.

The results are presented in table 9 in the same format as table 8. The range of results for the baseline model are repeated for comparison. There are two key findings. First, the model with constant sorting generates smaller time series results. The quantitative reduction is modest for the change in wage levels and stronger for the change in the college wage premium; for the latter, the results are roughly one-half of those in the baseline model. This finding indicates that half of the model’s predictions for the time series of the college wage premium stems from changes in sorting and half from the expansion of education; each is important.

The second main finding of this table is that the upper bound collapses to lie almost exactly at the lower bound for the case with constant sorting. This happens because the constant sorting experiment assumes more sorting and larger composition effects in earlier cohorts than does the baseline experiment. Because of this the model with constant sorting hits its upper bound for much smaller values of σ_a . The range of plausible results in this case is extremely narrow.

6 Robustness

The previous section introduced three key results. At the lower bound of the range, composition effects explain about half of the college wage premium for the 1960 cohort and the entire rise of the college premium between the 1910 and 1960 cohorts (years 1950 to 2000). Finally, composition effects generate a slowdown in wages conditional on schooling that has a stronger effect on less educated groups. The goal of this section is to demonstrate the robustness of these results. The focus is on two experiments. The first allows for $\beta_a < 0.104$. The model’s predictions are most sensitive to this target, so it is worthwhile to consider smaller values. The second experiment explores the results from an alternative model where schooling and cognitive ability complementarity comes through wages rather than preferences. Finally, brief consideration is given to the possibility that the ability

Table 10: Robustness of Selected Results: Lower Wage Return to Test Score

Statistic	Data	Contribution of Composition Effects	
		Baseline Model	Model with Lower Return to Test Score
1960 College Wage Premium	0.52	0.25 – 0.37	0.17 – 0.37
1910–1960 High School Wage Change	0.29	-0.14 – -0.21	-0.09 – -0.21
1910–1960 College Wage Premium Change	0.15	0.14 – 0.20	0.09 – 0.20

Note: Table presents select wage statistics and the fraction of each that can be attributed to composition effects. Data are taken from the censuses. Results are derived from the calibrated model under the baseline estimate of the return to test score or an alternative, more conservative calibration target. Each case gives a range of plausible results.

distribution may have changed over time.

6.1 Lower Log-Wage Return to Test Scores

The key moment for the calibration is the log-wage return to test scores in the NLSY79, which is measured in this paper as $\beta_{\hat{a}} = 0.104$. This estimate is similar to other estimates in the literature that use the NLSY79 (see for example Mulligan (1999) table 6, or Altonji and Pierret (2001) table I). However, estimates based on other data sources differ. Bowles et al. (2001) collect 24 studies using different data sources. The mean return across studies was 7%, with substantial dispersion. This section provides results for using a more conservative value $\beta_{\hat{a}} = 0.07$ as a calibration target.

The calibration strategy is the same as the baseline case, except that for each possible $\sigma_{\hat{a}}$, θ is calibrated to match $\beta_{\hat{a}} = 0.07$. Table 10 gives the results in the same format as the previous two tables. The main finding is that the results for the lower bound are about one-third smaller than in the baseline case, while the results for the upper bound are the same. Note that even for the lower bound of the robustness check, composition effects still account for nearly two-thirds of the rise in the college wage premium and nearly one-third of the current college wage premium.

The results do not change at the upper bound because of the bounding methodology. The upper bound is chosen so that ability gaps are as large as plausible. To do so, the algorithm chooses a calibrated value of θ similar to the upper bound in the baseline. It rationalizes the low measured return to IQ as being the result of higher levels of calibrated noise in test scores and a larger degree of attenuation bias in the regression. Hence, the essentially unchanged results at the upper bound are driven by the way the upper bound is constructed.

Table 11: Robustness of Selected Results: Ability School Complementarity in Wages

Statistic	Data	Contribution of Composition Effects	
		Baseline Model	Model with Complementarity
1960 College Wage Premium	0.52	0.25 – 0.37	0.25 – 0.40
1910–1960 High School Wage Change	0.29	-0.14 – -0.21	-0.11 – -0.18
1910–1960 College Wage Premium Change	0.15	0.14 – 0.20	0.10 – 0.17

Note: Table presents select wage statistics and the fraction of each that can be attributed to composition effects. Data are taken from the censuses. Results are derived from the calibrated model under the baseline case where ability-school complementarity enters through preferences or the alternative where it enters through wages. Each case gives a range of plausible results.

6.2 Ability-School Complementarity in Wages

Empirically, high test score students tend to go to school longer (see for example figure 2). The baseline model assumes complementarity between cognitive ability and school through the utility function to match this fact. A common alternative in the literature is to use complementarity through wages instead. In this case, more able workers go to school longer because their wage payoff to doing so is higher, not because they find it less distasteful. The goal of this section is to show that the exact form of complementarity is not essential for the results.

A generalized period log-wage function that allows for ability-school complementarity is:

$$\log[w(s, q, v)] = \theta_s a + z(s, \tau + v - 1) + h(s, v). \quad (10)$$

Complementarity requires that θ_s be weakly increasing in s . It is useful to focus on the alternative interpretation of the model discussed in Section 3.6. In this interpretation a is still a worker's cognitive ability but p represents noise in the worker's signal about that ability and $p+a$ is the signal. This interpretation offers the convenient feature that workers sort perfectly by $p+a$, as in the baseline model. Hence, the basic mechanics of this wage complementarity model will be the same as the mechanics of the baseline model, which makes it easy to focus on whether there are any important quantitative differences in their implications.

To provide evidence on the extent of wage complementarity the NLSY sample is split into two groups: high school dropouts plus high school graduates, and some college plus college graduates. The log-wage return to test score is estimated separately for each group. The corresponding estimates are $\beta_{a,HS} = 0.076$ and $\beta_{a,C} = 0.120$, as opposed to $\beta_a = 0.104$ when all four groups are pooled. The model is then recalibrated to match the two estimated returns to test scores, with the other calibration targets the same as above.

Table 11 shows the results in the same format as for the previous robustness check. Overall, a model with complementarity that comes through wages produces results quite similar to the baseline model, but slightly smaller in the time series. The main reason for the smaller results can be understood by comparing the importance of cognitive ability for wages across school types and models, $\theta_{HS} < \theta < \theta_C$, with θ referring to the shared importance of ability for wages in the baseline model. The alternative model implies that cognitive ability is more important for wages for college students but less important for high school students. This in turn leads to smaller results since across all of the calibrations the model finds a much larger decline in mean cognitive ability for high school dropouts and high school graduates than for those who attend college; see for example table 6. In fact, many of the calibrations imply almost no change in the mean ability of college students between the 1910 and 1960 cohorts. Hence, the main impact of allowing for complementarity through wages is that it amplifies the relatively small changes in mean ability for students who attend college and compresses the relatively large changes in mean ability for students who do not, with a net result of modestly lower results. Still, even the lower bound implies that changes in mean ability between high school and college graduates account for two-thirds of the rise in the college wage premium. The exact form of school-cognitive ability complementarities is not central to the conclusions.

6.3 The Flynn Effect

All the results so far have assumed that the distribution of abilities is time-invariant. There is, however, substantial evidence of a sustained rise in test scores since World War II, a phenomenon known as the Flynn effect (Flynn, 1984, 2009). There is disagreement in the psychometric literature as to whether the Flynn effect represents real gains in skills, improvements in test-taking skills, or some other possibility (Flynn, 2009). This section explores the implications for the results if the Flynn Effect captures actual rises in ability.

Although it is still somewhat controversial, the evidence now seems to suggest that the rise in ability is a mean shift that affects all parts of the distribution more or less equally. Flynn (2009) documents that average test scores on the WISC, a broad-based IQ exam, rose 1.2 standard deviations between 1947 and 2002, which corresponds roughly to the period of interest. He also conjectures (based on incomplete evidence) that test scores on the Raven's Progressive Matrix Exam, a test of spatial recognition, rose 1.83 standard deviations over the same years.

The Flynn effect has modest implications for the key results. An increase in the entire distribution of ability changes the mean ability of workers $E(a|s)$ by a constant amount,

but does not affect the mean ability gaps $E(a|s) - E(a|s')$. In particular, a rise in ability of 1.2–1.83 standard deviations implies a rise in $\theta E(a|s)$ by 19–28 percentage points for the benchmark calibrated value of $\theta = 0.155$. Any of these results imply that the mean ability of all school groups actually rose between the 1910 and 1960 cohorts, so that the model does not help explain the wage slowdown. Otherwise, the Flynn Effect has no important implications for the results about wage premiums because it affected different school groups equally.

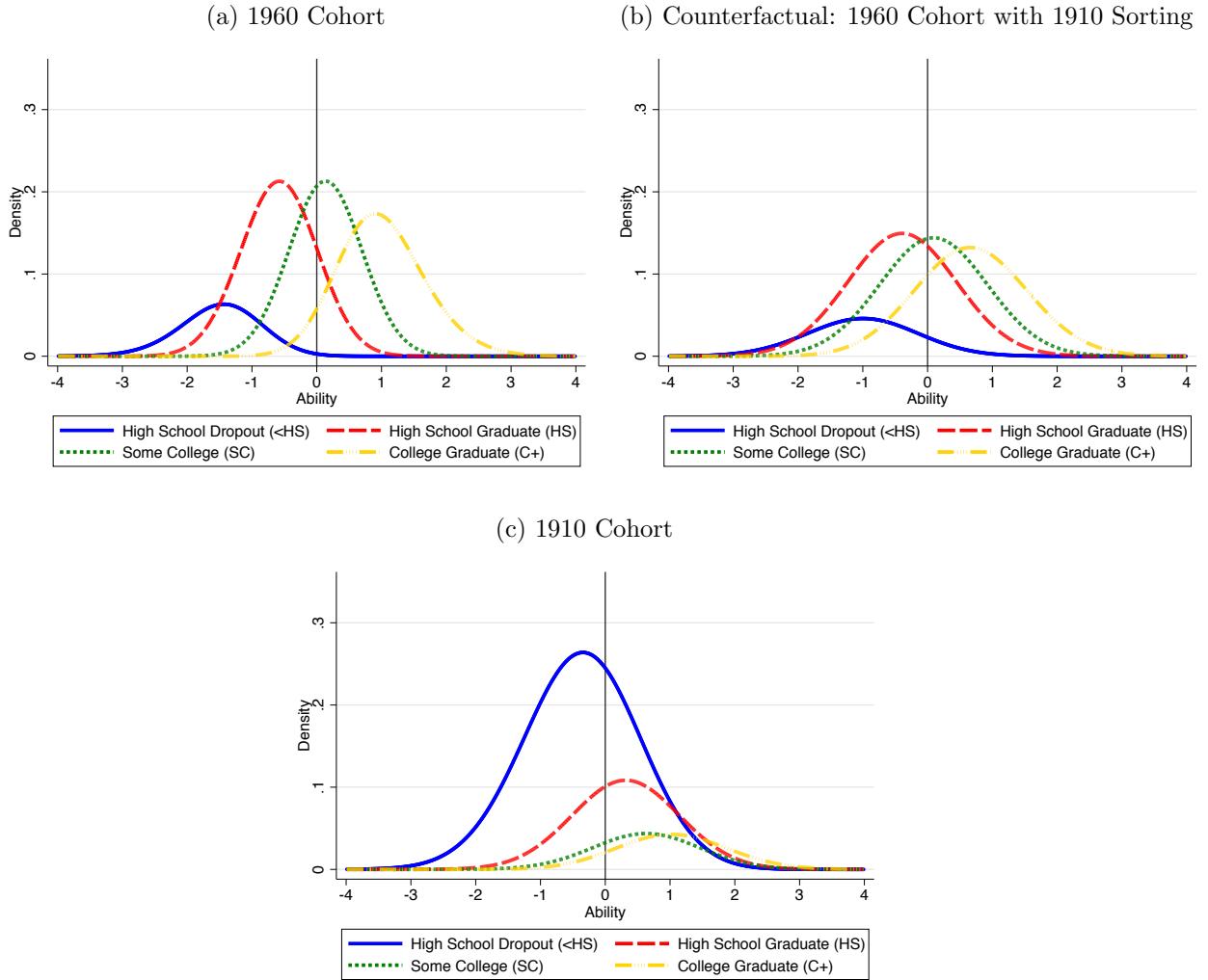
7 Conclusion

Between the 1910 and 1960 cohorts (years 1950–2000), the college wage premium widened substantially. Today the college wage premium is sufficiently large that it may be difficult to reconcile with a model of individual human capital investment. Most papers have tried to understand these movements as the result of changes in skill prices, roughly the wage per unit of labor. The analysis in this paper breaks with most of the literature by asking instead whether changes in the units of labor per worker may be responsible. Large changes in the school attainment of workers and the degree of sorting by test scores into educational attainment suggest that the mean ability of workers with different education levels may have changed. The main purpose of this paper is to quantify these composition effects and their impact on wages.

The results suggest that much of the most important wage questions can be attributed to changes in the mean ability of students by school attainment. The benchmark calibration can explain all of the rise in the college wage premium as well as half of the currently high college wage premium. Additionally, the model can help explain some of the wage slowdown as the result of declining mean ability conditional on schooling. Robustness checks indicate that roughly half of the time series results come from the expansion of education and half from the increase in sorting. These results would still be economically significant even if test scores measured ability exactly or if more conservative moments were used in the calibration procedure.

The results rely on a simplified model with one dimension of ability and one generic friction to sorting, the tastes of workers. Future work could make progress by developing a more detailed model of ability or the frictions that act to prevent stronger sorting by ability, and by finding more historical data on these forces for empirical use.

Figure 3: The Distribution of Ability Conditional on Schooling



Note: Figure shows the model-implied distributions of ability conditional on schooling. Figures (a) and (c) show the distributions for the 1960 and 1910 cohorts. Figure (b) shows the hypothetical distribution that would have prevailed in 1960 if the expansion of education had occurred but not the increase in sorting of students by test score. Comparison of (a) to (b) isolates the effect of increased sorting, while comparison of (b) to (c) isolates the effect of the expansion of education.

References

Acemoglu, D., 2002. Technical change, inequality, and the labor market. *Journal of Economic Literature* 40, 7–72.

Altonji, J.G., Pierret, C.R., 2001. Employer learning and statistical discrimination. *Quarterly Journal of Economics* 116, 313–350.

Autor, D.H., Katz, L.F., Kearney, M.S., 2008. Trends in US wage inequality: Revising the revisionists. *Review of Economics and Statistics* 90, 300–323.

Autor, D.H., Katz, L.F., Krueger, A.B., 1998. Computing inequality: Have computers changed the labor market? *Quarterly Journal of Economics* 113, 1169–1213.

Bishop, J.H., 1989. Is the test score decline responsible for the productivity growth decline? *American Economic Review* 79, 178–197.

Bound, J., Johnson, G., 1992. Changes in the structure of wages in the 1980s: An evaluation of alternative explanations. *American Economic Review* 82, 371–392.

Bowles, S., Gintis, H., Osborne, M., 2001. The determinants of earnings: A behavioral approach. *Journal of Economic Literature* 39, 1137–1176.

Bowlus, A.J., Robinson, C., 2012. Human capital prices, productivity, and growth. *American Economic Review* 102, 3483–3515.

Bureau of Labor Statistics; US Department of Labor, 2002. National longitudinal survey of youth, 1979 cohort. Produced and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH.

Cameron, S.V., Taber, C., 2004. Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy* 112, 132–182.

Carneiro, P., Lee, S., 2009. Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics* 149, 191–208.

Carneiro, P., Lee, S., 2011. Trends in quality adjusted skill premia in the US, 1960-2000. *American Economic Review* 101, 2309–2349.

Cawley, J., Conneely, K., Heckman, J., Vytlacil, E., 1997. Cognitive ability, wages, and meritocracy, in: Devlin, B., Fienberg, S., Resnick, D., Roeder, K. (Eds.), *Intelligence, Genes, and Success: Scientists Respond to The Bell Curve*. New York: Springer.

Chay, K.Y., Lee, D.S., 2000. Changes in relative wages in the 1980s: Returns to observed and unobserved skills and black–white wage differentials. *Journal of Econometrics* 99, 1–38.

Cremin, L.A., 1961. *The Transformation of the School: Progressivism in American Education, 1876-1957*. Knopf, New York.

Cunha, F., Heckman, J.J., Navarro, S., 2005. Separating uncertainty from heterogeneity in life cycle earnings. *Oxford Economic Papers* 57, 191–261.

Deschênes, O., 2006. Unobserved ability, comparative advantage, and the rising return to education in the United States 1979-2002. Mimeo, University of California at Santa Barbara.

Finch, F.H., 1946. *Enrollment Increases and Changes in the Mental Level*. Stanford University Press, Stanford, CA.

Flynn, J.R., 1984. The mean IQ of americans: Massive gains 1932 to 1978. *Psychological Bulletin* 95, 29–51.

Flynn, J.R., 2009. *What Is Intelligence? Beyond the Flynn Effect*. Cambridge University Press.

Garriga, C., Keightley, M.P., 2007. A general equilibrium theory of college with education subsidies, in-school labor supply, and borrowing constraints. Mimeo. Federal Reserve Bank of St Louis.

Goldin, C., Katz, L.F., 2008. *The Race between Education and Technology*. Harvard University Press, Cambridge, MA.

Heathcote, J., Perri, F., Violante, G.L., 2010. Unequal we stand: An empirical analysis of economic inequality in the United States, 1967-2006. *Review of Economic Dynamics* 13, 15 – 51.

Heckman, J.J., Lochner, L., Taber, C., 1998. Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics* 1, 1–58.

Heckman, J.J., Lochner, L.J., Todd, P.E., 2006. Earnings functions, rates of return and treatment effects: The Mincer equation and beyond, in: Hanushek, E., Welch, F. (Eds.), *Handbook of the Economics of Education*. Elsevier. volume 1. chapter 7, pp. 307–458.

Heckman, J.J., Lochner, L.J., Todd, P.E., 2008. Earnings functions and rates of return. *Journal of Human Capital* 2, 1–31.

Herrnstein, R.J., Murray, C.A., 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hoxby, C.M., 2009. The changing selectivity of american colleges. *Journal of Economic Perspectives* 23, 95–118.

Juhn, C., Kim, D.I., Vella, F., 2005. The expansion of college education in the United States: Is there evidence of declining cohort quality? *Economic Inquiry* 43, 303–315.

Juhn, C., Murphy, K., Pierce, B., 1993. Wage inequality and the rise in the return to skill. *Journal of Political Economy* 101, 410–442.

Katz, L.F., Murphy, K.M., 1992. Changes in relative wages, 1963-1987: Supply and demand factors. *Quarterly Journal of Economics* 107, 35–78.

Kaymak, B., 2009. Ability bias and the rising education premium in the United States: A cohort-based analysis. *Journal of Human Capital* 3, 224–267.

Laitner, J., 2000. Earnings within education groups and overall productivity growth. *Journal of Political Economy* 108, 807–832.

Levy, F., Murnane, R.J., 1992. U.S. earnings levels and earnings inequality: A review of recent trends and proposed explanations. *Journal of Economic Literature* 30, 1333–1381.

Manski, C.F., 1989. Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon. *Economics of Education Review* 8, 305–312.

Mulligan, C.B., 1999. Galton versus the human capital approach to inheritance. *Journal of Political Economy* 107, S184–S224.

Murnane, R.J., Willett, J.B., Levy, F., 1995. The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics* 77, 251–266.

Navarro, S., 2008. Using observed choices to infer agent's information: Reconsidering the importance of borrowing constraints, uncertainty and preferences in college attendance. Mimeo: University of Wisconsin.

Stinebrickner, R., Stinebrickner, T.R., 2008. The effect of credit constraints on the college dropout decision: A direct approach using a new panel study. *American Economic Review* 98, 2163–2184.

Taber, C.R., 2001. The rising college premium in the eighties: Return to college or return to unobserved ability? *Review of Economic Studies* 68, 665–691.

Taubman, P., Wales, T., 1972. Mental ability and higher educational attainment in the 20th Century. Berkeley: National Bureau of Economic Research and Carnegie Commission on Higher Education.