# Ability, Schooling and Wages: Going Beyond the National Longitudinal Surveys

**Fernando Borraz**[*]
**Universidad de Montevideo**

Abstract

This paper estimates returns to education in the US using information from two datasets, the National Longitudinal Surveys (NLS and NLSY79) and the Public Use Microdata Sample (PUMS). The high correlation between schooling and ability did not allow the separate identification of each effect. The PUMS dataset contains information on wages and education but not on ability and can therefore be exploited to improve the precision of the NLS and NLSY79 estimates. The results suggest a positive but not increasing over time wage gap only for the most able during the 80's, and between 1980 and 2000.

**Keywords**: Schooling, Wages, Ability, Bounds, Identification Region

[*] Address for correspondence: Prudencio de Pena 2440. CP: 11600. Montevideo, Uruguay. Phone 598—2-7074461. Fax: 598-2-7083842. E-mail: fborraz@um.edu.uy. I thank Philip Cross, Mathew Dey, Juan Dubra, Federico Echenique, Arik Levinson, Lucas Navarro and Marisa Sellanes for very helpful comments. Comments received during seminars at Georgetown University, the 2004 Midwest Economic Association meeting, the 2003 Latin American and Caribbean Economic Association meeting, the IDB and the 2004 Latin American meeting of the Econometric Society are gratefully acknowledged. All remaining errors are my own.

## I. Introduction

Since the 1980's, the United States has experienced considerable changes in the structure of wages being paid to different demographic and educational groups. The most significant of these is the increase in the wages of more educated workers relative to their less educated counterparts. Katz and Revenga (1992) find an increase in the wages of young male college graduates approximately 30 percent greater than the wages of young males with no more than twelve years of schooling between 1979 and 1987. They find also that wage inequality also increased within narrowly defined educational groups.

In this paper I analyze the determinants of the wage gap: how much is explained by an increase in the value of ability, and how much is due to an increase in the market value of education. Though efforts have been made to estimate these shares (Blackburn and Neumark 1993, Murnane, Willett and Levy 1995, Heckman and Vytlacil 2001, Taber 2001) the issue has not been resolved. A better understanding of the contributions of the two to the wage gap is important because it has important policy implications. If the widening gap is due to increasing returns to education over time there would exist justification for policies that enhance education. If, on the other hand, the rising gap is due to increasing returns to ability, given that ability is more difficult to change through policy interventions, there would be a smaller scope for policy.

Despite the existence of several competing theories for the increase in wage differentials by education, it is difficult to find a completely satisfactory explanation. The difficulty lies in the fact that education is highly correlated with ability. Some studies relate the increase in wage inequality to changes in international trade patterns or to changes in the industrial structure of the economy (Freeman 1997). Other studies point to the effects of changes in the relative supply of workers of different educational levels on wage inequality (Nickell and Bell 1996). However, none of the studies takes into account the effect of the underlying ability of individuals on wages. Recent studies (Heckman and Vytlacil 2001) suggest a relationship between a worker's inherent ability[1] and his level of schooling. Moreover, this relationship may have changed over time. Thus, estimates obtained from wage regressions are potentially biased by the presence of unobserved ability in the wage-equation error term.

Also, changes in the schooling-ability relationship could have led to changes in observed return to schooling over time. Another possibility is that actual returns to schooling have not changed over time and instead the observed increase in earning differentials is attributable to changes in the correlation between schooling and ability. Because of this, the omission of an ability covariate in a wage regression could lead to an "observed" rather than a real increase in the returns to schooling over time. A possible explanation is that the increase in the return to education could have occurred mainly for certain workers, for example those with higher levels of "academic" ability.

Disentangling the effects of ability and schooling with the available data is difficult. Workers with higher ability tend to acquire a higher level of education, giving a strong positive ability-education correlation, resulting in a sorting bias. Therefore, the two series tend to be indistinguishable and it is not possible to estimate precisely the effect of schooling on wages for all levels of ability. Based on information from the National Longitudinal Survey Youth Cohort of 1979 (NLSY79), Table 1 presents the distribution of schooling and ability. It shows that few workers who have completed sixteen years or more of education and have a college degree are positioned in the lowest third of ability[2]. Given the small number of individuals, the effect of a college degree on wages of individuals at such an ability level cannot be reliably estimated.

Most of the literature about ability, schooling and wages uses the NLSY79. Given the small sample size of the NLSY79 (and the sorting problem), it is not possible to estimate the effects of schooling on wages at all levels of ability. For example, Heckman and Vytlacil (2001) use the NLSY79 and find evidence that the effect of schooling on wages is not linear. However, they were not able to identify the effects of schooling on wages at all levels of ability given that the NLSY79 has too few

[1] i.e., ability not affected by acquisition of schooling, commonly measured by test scores

[2] Similarly, there are no individuals with less than 12 years of schooling with the highest level of ability.

observations in some regions of the joint distribution of ability and education (see Table 1).

This study will attempt to overcome the estimation problem that arises as a consequence of the strong positive correlation between schooling and ability by estimating the effect of schooling on wages for all levels of ability. In order to achieve this, information from two datasets are used. I supplement the National Longitudinal Surveys with the 1% PUMS (Public Use Microdata Sample)[3] from the U.S. Bureau of the Census. Although the PUMS does not include an ability covariate, it is added as an auxiliary sample to exploit the fact that a larger sample allows sharper inferences than can be obtained from the NLSY79 alone.

As a starting point, we use the NLS and NLSY79 to estimate average wages for workers with different ability and education levels. In a second step, the PUMS dataset, which containing information on wages and education but not on ability, can be exploited to improve the precision of the NLS and NLSY79 estimates. In essence, the PUMS data impose an adding-up restriction (which I explain below) on the parameters of interest, allowing for more efficient estimates than those yielded by the NLSY79 data alone. This is done by applying the non-parametric bounding technique described in Cross and Manski (2002). In fact, adding information from the PUMS dataset substantially sharpens the estimates of the returns to education at different ability levels- with confidence intervals for the returns to education substantially reduced. As a consequence of the use of the PUMS there is an important reduction in the confidence intervals (around 30-40%). This sharper inference is equivalent to doubling the sample size of the NLS. Our estimates show that the increase in returns, are concentrated among the most able.

## 2. Literature Review

The majority of the studies about wages, schooling and ability use the NLSY data. There are however, important methodological differences between them. Initially Mincer equations were estimated using OLS. Wages ere presumed to depend on

[3] The sample size is 1% of the U.S. population.

education and experience with ability as an additional explanatory variable. These estiates were followed by more studies that new ones introduce more advanced techniques.

Blackburn and Neumark (1993) estimate pooled samples across years and they estimate the OLS regression for a single experience group. They find that the returns to education are concentrated among the most able. In a logical extension, Murnane, Willett and Levy (1995) estimate the regression for a single age group in two different times (1978 and 1986). They find an important increase in the economic returns to education between 1978 and 1986.

Heckman and Vytlacil (2001) exposed the sorting problem, showing that only the most able individuals go to college. Therefore they conclude that it is not possible to estimate the returns of schooling for all ability levels. Two papers overcome the sorting problem using more advanced estimation techniques. Taber (2001) estimated a dynamic programming selection model which shows that a very important part of the story is that we have experienced an increase in the returns to ability overtime. Tobias (2003) focuses his analysis in a support that is common to individuals with and without college degree. He concludes that the college high school wage gap is only increasing for the most able in the period 1984-1994.

## 3. Data Description

The NLS Young Men Cohort was initiated in 1966 with a sample of 5225 youth between the ages of 14 and 24. These men were interviewed from 1966 through 1981. The NLS Young Men Cohort of 1980 consists of a sample of 815 white male workers and contains information on their wages, their schooling along with the results of an IQ test administered to each.

The NLSY79 is a sample of 7429 white youth between the ages of 14 and 22 during 1979, the first year of the survey. Each respondent (3709 males and 3720 females) was interviewed yearly until 1994, and every two years after that. Our study takes advantage of scores on the ASVAB test (Armed Services Vocational Aptitude Battery). These are several tests measuring academic (or cognitive) and mechanical (arithmetic, numerical operations etc) ability. These test scores are used as (potentially error-prone) measures of ability.

The longitudinal nature of my data requires that I deflate wages data. As usual in the literature I use the national consumption expenditure deflator. Also, because of the presence of outliers, I restrict the sample to those receiving an hourly wage higher than 50% of the minimum wage in 1980 positive wage[4]. Also I excluded people who are still in school.

The PUMS survey is a 1% census sample. I use the 1980, 1990, and 2000 samples. They contain records representing 1-percent of the occupied and vacant housing units in the US. The number of observations is 130,000 in 1980 and approximately 200,000 in 1990 and 2000. The 1-percent sample give users the maximum amount of social, economic, and housing information available. It contains detail information about gender, sex, schooling and wages.

**3.1 Ability**

The NLS and NLSY79 both contain data on test scores of respondents. The NLS has information about an IQ test and the NLSY79 reports the Armed Based Services Vocational Aptitude Test (ASVAB). The ASVAB consists of ten standardized test which are use by the US Armed Forces to evaluate a variety of skills. The ten subsets are: general science, arithmetic reasoning, word knowledge, paragraph comprehension, numerical operations, coding speed, auto and shop information, math knowledge, mechanical comprehension and electronic information. The ASVAB test measures academic (or cognitive) and mechanical ability (arithmetic, numerical operations etc).

Following the pioneering work of Spearman (1927), ability is defined as a factor that explains test scores. Given the fact that intelligence test scores tend to increase with age and education, whereas the ability variable that is "innate" and "unobservable" should not, ability is defined as the first principal component, called "g", of the residuals of a regression of test scores with respect to age and education.

[4] A very small fraction of the white males have zero wage.

The ability information from NLSY79 is derived from test scores given to youth between the ages of 14 and 22 in 1979. These tests were administered in 1979 upon enrollment in the Army. In order to account for Keane and Wolpin (1997) comments, we define ability as the residual of a regression of the scores on age, education and parent schooling. That is we include the schooling of the mother in the regression in order to control for family environment.

A remarkable finding of the literature is that only one combination of tests, the first principal component, predicts test performance almost as well as the full battery of tests. Heckman and Vytlacil (2001) show that there is a little difference in terms of explained variability in wage regression using "g" or any other linear combination of tests. Also, they show that adjusting test scores to consider the fact that they increase with age and education, (referred to as estimation "adjusted" test scores) gives the best measure of general intelligence.

Keane and Wolpin (1997) claim that test scores are more a measure of endowment heterogeneity than of cognitive ability. They show that endowment heterogeneity is a mayor determinant of lifetime ability. It plays a much more important role than family income at 16 and other background variables in explaining variability in college attendance. They conclude that skill endowment differences are an important determinant of inequality. Therefore, it is important to obtain measures of investment in children before age 16.[5]

One limitation of the analysis is that because of data limitations we considered ability as a uni-dimensional variable based on test scores. However, there are different types of ability: social skills, trustworthiness, reliability and communications skills that can affect earnings.

### 3.2 Wages and education

[5] For example, child care, child nutrition, school expenditures.

Information on race, gender, age, wages and education are contained in the NLS, NLSY79 and PUMS surveys. I use hourly wages in the estimation because it comes closest to the transaction price of labor. In order to have significant variation in education and ability, I divide the population into three levels of schooling: individuals who attended high school but did not graduate, those with a high school diploma, and those with a college degree. I restricted my sample to white males because the returns to education are different across race and gender.

### 3.3 Stylized facts from NLSY79 data

This sub-section documents the increase in two types of variables: wages for groups with different education levels; and the wage *gap* among groups with different education levels.

Figures 1.a and 1.b show returns to schooling over time from the NLSY79. These plots show wages increasing for college graduates, high school diploma holders and non-graduate high school individuals from the mid-1980s to 2000. The most remarkable finding is the rapid (steep) increase in log wages since the mid-1980 for college graduates. These graphs are consistent with Blackburn and Neumark (1993) who claim that increasing returns to schooling are heavily concentrated among the most able. They are also consistent with Murnane, Willet and Levy (1995) who conclude that ability had a larger impact on wages for 24 year-old men in 1986 than in 1978.

In order to have a considerable variation in the ability across individuals, I split the population by considering three levels of ability. Figures 2.a and 2.b display wage increases for all levels of ability. It can be seen that the rise in wages is higher for the most able.

Moreover, Figure 3 shows an estimated increasing college wage gap starting in the mid-80's. There is a steady increase in the wage gap after the mid-80's. We can also observe a smaller increase in the wage gap between high school diploma holders and those non-graduates from high school.

From Figure 1a) and Figure 3 we can infer that the increase in returns to education is greater for people with higher levels of education.

Given these stylized facts, the question this paper will address is the following: What is the entire joint distribution of wages, schooling and ability and how did it evolve over time? This important question has remained un-answered due to the statistical shortcomings of the data. The strong correlation between schooling and ability results in the NLSY79 to have a scarcity of observations in some regions of the joint distribution of education and ability (see Table 1). In particular, we cannot make sharp inferences for non-high school graduates of high ability and college graduates of low ability.

## 4. Methodology

The objective of the empirical analysis is to estimate average wages given the level of schooling, ability and experience overtime for white males. I will focus on time effects to account for the fact that the relationship between ability and schooling may have changed over time.

The effect of these covariates is non-linear; therefore I cannot pool observations and use ordinary least squares as a method of estimation. For example, the returns to education are different depending on ability and experience. As in Heckman and Vytlacil (2001), I will conduct a non-parametric analysis. In order to accomplish this, I need to limit the number of explanatory variables.

The main objective is to estimate the expected log wages given workers characteristics (x) and ability (a) ($E[\log(w)|x,a]$), We let w represent wages deflated by the national consumption expenditure deflator, we include schooling, experience and time to control for workers characteristics (x) . We estimate time effect in 3 dimensions: schooling, ability and age. This flexible estimation allows us to measure the interactions between experience and time.

As mentioned above, NLSY79 is a sample from the probability distribution of wages given schooling and ability. I am able to estimate the return to education for people with high school diplomas and low ability from NLSY79. In Figure 4 we can observe increasing wages in time for all ability levels. Also, I can estimate the wages for people with college degrees and high ability. I need to see what happens with the wages of college graduates with low ability and with the wages for high school graduates of high ability.

NLSY79 has few observations in some regions of the joint distribution of ability and education. In particular, there is a small number of people with college degrees and low ability. Because of that, it is not possible to estimate returns to education for all schooling and ability levels.

A possible solution for this identification problem is to augment the NLSY79 information with a larger dataset, like the PUMS. Although the PUMS does not contain an ability variable, its addition as a second, larger sample allows for sharper inferences than would be available from using the NLSY79 itself.

NLSY79 allows us to estimate average wages for workers with different abilities and education levels. The PUMS dataset (which includes wage and education data, but excludes ability) contains information that improves the precision of the NLSY79 estimates. The source of the improved precision is the non-parametric bounding technique described in Cross and Manski (2002). Incorporating this marginal information available from the census, substantially allows for better estimation of the effects of ability and schooling on wages, at different ability levels. I will determine whether or not efficiency gains afforded by the marginal information are substantial.

In essence, I incorporate marginal information, the $E(w|x)$ from PUMS in order to improve inference about $E(w|x,a)$.

My approach is to divide the population into three ability levels, and treat data from the PUMS as if it were population data. In order to simplify the methodological exposition and to provide an intuitive understanding of the underlying methodology, I describe the situation with ability is divided into two levels instead of three. [6]

We need to keep in mind that the conditional probability of wages given workers characteristics, P(w|x) is a mixture of two distributions:

- P(w|x,Ability=High)
- P(w|x,Ability=Low)

Knowledge of P(w|x) places non-parametric bounds on E(w|x,a).

The lower bound for E(w|x,a=Low) is when Pr(a=Low|x) of the people have the lowest wage in the distribution of P(w|x). P(w|x,a=Low) is a right-truncated version of P(w|x) with the truncation at Pr(a=Low|x) of P(w|x) mass. The expectation of this distribution gives the lower bound on E(w|x,a=Low).

In a similar way I can construct an upper bound on E(w|x,a=Low).

Also, the vector {E(w|x,a=High), E(w|x,a=Low)} must satisfy:

$$E(w|x) = E(w|x,a=High)P(a=High|x) + E(w|x,a=Low)\,P(a=Low|x)$$

The bound on { E(w|x,a=High), E(w|x,a=Low) } is a subset of a line in $R^2$. Cross and Manski (2002) term this the identification region.

The next logical step is to combine these bounds with information from NLSY79. I can combine the identification region estimated from PUMS with a $(1-\alpha)$[7] confidence region for {E(w|x,a=High), E(w|x,a=Low)} from NLSY79.

For a complete description of the methodology used to construct the bound see Cross and Manski (2002).

[6] The case with three ability levels is slightly more complex. See Cross and Manki (2002) for a comprehensive explanation

[7] $\alpha$ is the significance level.

## 5. Estimation

### 5.1 Estimated wage regressions using solely NLSY79 data

In this section I present the returns to education and ability considering solely NLSY79 data. In this way I will be able to show clearly the estimation problem that I address using information from a second survey.

Following Heckman and Vytlacil (2001), in order to conduct a meaningful non-parametric analysis, I must limit the number of explanatory variables. To circumvent the curse of dimensionality, I include fewer variables than other papers that analyze wage, education and ability. Griliches (1977) shows that other variables, such as family background and location, affect wages through schooling and ability, and they do not have any direct impact on the wage equation.

Somewhat unexpectedly, Figure 4 shows increasing wages for high-school non-graduate with low ability. For high school graduates at the lowest ability level, we observe a wage increase only after 1992. Also, Figure 4 shows steady wage increases from the 1980s among college graduates.

One explanation for these results could be that wages increase with age (or experience), time, or maybe with both. Therefore, in order to test which is the best specification it is necessary to introduce experience and time effects.

**Is the effect of ability on wages nonlinear?**

Heckman and Vytlacil (2001) and Tobias (2003) show that the effects of time and age on wages are non-linear. Therefore, the common practice of estimating a linear wage equation with age and time covariates is unacceptable. A non-parametric analysis with respect to age and time will be more satisfactory in this particular case. With this in mind, I estimate wages across education-ability cells, at a certain age and for a particular year. I will also estimate experience-time interactions.

In order to test that the effect of ability on wages is nonlinear, I estimate a wage regression with dummy variables for ability groups in Table 2 using NLS 1980 wage data. At the 95% level of confidence the dummy variables for the highest ability levels is statistically significant.

I also estimated the wage equation in a semi-parametric way using the NLS80. In particular I estimated the following equation:

$$\log(w) = f(a) + \beta x + \varepsilon \qquad \text{with } E(\varepsilon | a,x) = 0 \quad \text{and} \quad E(\varepsilon^2 | a,x) = \sigma^2$$

where w is the hourly wage, a is ability as measured by an IQ test and x is a vector including schooling and experience, I estimated this equation following Yatchew's (1998) method. Estimation results are shown in Figure 5. In the middle of the distribution of test scores the impact of ability on wages is similar by OLS or by a semiparametric method. However, near the tails of the distribution there are significant differences indicating that the relationship between wages and ability is not linear.

I also performed the Yatchew (2003) specification test to assess the null hypothesis that there is a linear relationship between the variables. This test is based on differencing two series in order to eliminate the non-parametric part of the regression. The test statistics has standard normal distribution and is 23; therefore I reject a linear effect of ability on wages.

**Wages, ability and schooling**

Figure 6 displays the college graduates – high school diploma holders wage gap for all ages with respect to the most able. There is an increasing wage gap for white males between 25 and 33. This result showing an increasing wage gap for the most able is consistent with Blackburn and Newmark (1993).

A problem arises however while considering the use of NLSY79 and it is that the small number of observations in some regions of the ability-schooling distribution makes it difficult to estimate wages non-parametrically at all levels of ability. In order to increase the number of observations for these regions I will use a second and larger database: the PUMS from the U.S. Census.

### 5.2 Incorporating Census information

Given the fact that I will use information from two different surveys, a direct and obvious concern is how close the NLS and NLSY79 match the PUMS data.[8] In particular, the question is whether the NLSY79 and PUMS are samples from the same population.

Table 3 reports quartiles, means and standard errors for the log of hourly wages by schooling for two periods (1980 and 1990) at two different ages (28 and 30). I report the statistics for the NLS, NLSY79, and Census data. Table 3 shows that standard errors in the census data are much lower. This is due to the larger sample size of the PUMS survey and the high attrition rate for the NLS survey. Secondly, according to Table 3, differences in means are more pronounced in 1990.

However, we can not reject that both samples come from the same population. In eight of eight tests of differences in means presented in Table 3 there are no significant differences. The most significant for this paper is that the samples are closest for high school graduates and for individuals with a college degree.

In his seminal work, Mincer (1974) makes a critical distinction between experience and age. He concludes that experience rather than age should be used in the estimation of wage equation. Because of that, we use the experience variable in the estimation.

Table 4 presents the bounds for the log wage in 1980 1980. I will isolate the estimated wages from NLS, as well as information from PUMS that is used. The most remarkable

[8] See Hellerstein and Imbens (1999) for a discussion of the NLS and PUMS surveys.

fact is the important sharpening in the inferences. Also, because there is a small number of workers with college degrees and low ability we can not estimate economic return to education for that group. The use of the PUMS surveys allows for the estimation of a return between 1.91 and 2.20 at the 5% confidence level for the group with 6 to 10 years of experience and between 2.06 and 2.37 for individuals with more experience.

Considering that the sample size for NLS is only 815, for the subpopulation of interest there is a significant improvement in the estimation. The typical reduction in the confidence interval is equivalent to doubling the sample size.

Table 5 presents the results when information from NLSY79 is used. Since the sample size of the NLSY79 is more than three times the size of NLS, we do not observe significant reductions in the confidence interval width. This exercise is still useful however, because I am able to estimate a bound for the economic return of people with college degrees and low ability.

Table 6 presents the results for 2000, the last year with data in both samples, the NLSY79 and PUMS.

Because of the structure of the panel data, as individual's experience (or age) increases with time, we do not observe individuals with low levels of experience in the 2000 sample. Therefore, I can only analyze individuals with more than 11 years of experience in the 2000 sample.

In Table 7, I reject the hypothesis of an increasing wage gap only for the most able during the 80's for the individuals with 6 to 10 years of experience.

Similarly, we observe a positive wage gap for the most able for the years 1980 and 2000 for people with more than 11 years of experience. Therefore, there is no indication of an increasing wage gap over time.

## 6. Conclusion

I studied the wages-schooling-ability relationship using the Cross and Manki (2002) non parametrical analysis. The main outcome of this paper is that I am able to overcome the schooling-ability sorting problem therefore I estimate a bound for the returns to education for white males with college degrees at different ability levels.

The results of the estimation show that returns to education are concentrated among white males with the highest ability. In particular we observe a positive college degree - high school diploma wage gap only for individuals with the highest ability. This result is consistent with the claims of Blackburn and Neumark (1993), Taber (2001) and Tobias (2003) who found weak evidence that returns to education are concentrated among the most able. Our result confirms this finding. I also fail to find an increasing wage gap during the 1980 for individuals with medium experience, and between 1980 and 2000 for individuals with the most experience.

One limitation of the analysis is that because of data limitations we considered ability as a uni-dimensional variable. However, there are different types of ability: social skills, trustworthiness, reliability and communications skills that can affect earnings. A second limitation is that we have to restrict the analysis to white males. The small number of non whites in the NLSY79 prevents us from refining this XX. The study of the females wage gap present additional challenges due to the selection problem. It will be interesting further research to check if the results are similar for females or for other races.

## Bibliography


Blackburn, M. and D. Neumark (1993). "Omitted-Ability Bias and the Increase in the Return to Schooling," *Journal of Labor Economics*, 11: 521-544.

Cross, P. and C. Manski (2002). "Regressions, Short and Long Run," *Econometrica*, 70 (1): 357-368.

Freeman, R. (1997). When Earnings Diverge: Causes, Consequences, and Cures for the New Inequality in the U.S. Washington, DC: National Policy Association.

Griliches, E. (1977). “Estimating the Returns of Schooling: Some Econometric Problems ,” *Econometrica*. 45:1-22.

Grogger, J. and E. Eide (1995). ”Changes in College Skill and the Rise in the College Wage Premium,” *Journal of Human Resources*, 30(2): 280-310.

Heckman, J. and E. Vytlacil (2001). “Identifying the Role of Cognitive Ability in Explaining the Level of and Change in the Return to Schooling,” *Review of Economics and Statistics*, 83 (1):1-12.

Hellerstein, J.and G. Imbens (1999). “Imposing Moment Restrictions From Auxiliary Data by Weighting,” *The Review of Economics and Statistics* 81(1):1-14.

Juhn, C., Murphy K. and Pierce B. (1993). “”Wage Inequality and the Rise in Returns to Skill,” *Journal of Political Economy*. Vol. 101 no.3: 410:442.

Herrnstein, R., and C. Murray (1994). *The Bell Curve*, New York: Free Press.

Katz, L. and K. Revenga (1992). “Changes in Relatives Wages: 1963-1987: Supply and Demand Factors,” *The Quarterly Journal of Economics*. Vol 107. No 1: 35-78.

Keane, M. and K. Wolpin (1997). “The Career Decisions of Young Men,” *Journal of Political Economy.* Vol. 105. No. 3: 473-522.

Luven, E., H. Oosterbeek and H. van Ophem (2004). “Explaining International Differences in Male Skill Wage Differentials by Differences in Demand and Supply of Skill,” *The Economic Journal*, 114 (April) 466-486.

Manski, C. and J. Pepper (2000). “Monotone Instrumental Variables: With an Application to the Return to Schooling,” *Econometrica*, 68(4): 997-1010.

Mincer, J. (1974). “Schooling, Experience and Earnings”. New York: NBER Press.

Murnane, R., J. Willett and F. Levy (1995). "The Growing Importance of Cognitive Skills in Wage Determination," *The Review of Economics and Statistics*, 77 (2): 251-266.

Nickell, S. and B. Bell, (1996). "Changes in the Distribution of Wages and Unemployment in OECD Countries," *American Economic Review*, Papers and Proceedings, vol. 86, 302–308.

Spearman, C. (1927). The Abilities of Men: Their Nature and Their Measurement. New York: Macmillan.

Taber, C. (2001). "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability," *Review of Economic Studies* 68, 665-691.

Tobias, J. (2003), "Are Returns to Schooling Concentrated Among the Most Able? A Semiparametric Analysis of the Ability-earnings Relationships," *Oxford Bulletin of Economics and Statistic,* 65(1): 1-29.

Yatchew, A. (1998), "Nonparametric Regression Techniques in Economics," *Journal of Economic Literature*, XXXVI, 669-721.

Yatchew, A. (2003), *Semiparametric Regression for the Applied Econometrician.* Cambridge University Press.

**Table 1. Frequency Distribution by Ability and Schooling**
**At Age 35, (14 to 16 years-old in 1979), White Males**
**Sample Size = 319**

| Highest Grade Completed | Ability: Low | Ability: Medium | Ability: High |
|---|---|---|---|
| 7 | 2 | 0 | 0 |
| 8 | 10 | 1 | 0 |
| 9 | 12 | 6 | 0 |
| 10 | 8 | 6 | 0 |
| 11 | 13 | 4 | 0 |
| 12 | 47 | 58 | 41 |
| 13 | 3 | 8 | 10 |

| | | | |
|---|---|---|---|
| 14 | 1 | 10 | 18 |
| 15 | 2 | 2 | 5 |
| 16 | 1 | 4 | 18 |
| 17 | 0 | 1 | 4 |
| 18 | 0 | 1 | 3 |
| 19 | 0 | 0 | 1 |
| 20 | 0 | 0 | 3 |

I divide the population into three equals groups according to low, medium and high ability levels.

Low ability is defined as being below the 33th percentile of the first principal component of the (residualized) ASVAB test score distribution.

Source: NLSY79.

**Table 2. OLS Log Wage Equation**

| Variable | Coefficient | Standard Error |
|---|---|---|
| Constant | 0.,456 | 0.223 |
| Schooling | 0.073 | 0.008 |
| Schooling Square | -0.002 | 0,009 |
| Experience | 0.085 | 0.025 |
| Experience Square | -0.002 | 0,001 |
| Dummy Middle Ability | 0.043 | 0.034 |
| Dummy Highest Ability | 0.097 | 0.038 |

Source: NLS80. Sample Size = 815
$R^2$=0.14. DW=2.01. F-Stat.=25.4

**Table 3. 95% Confidence Intervals for Quartiles and Means of the Log of Hourly Wage (1980 dollars) by Age and Schooling.**

| | Age 30 | | | |
|---|---|---|---|---|
| | NLS 1980 | PUMS 1980 | NLSY 1990 | PUMS 1990 |
| High School | | | | |
| First Quartile | 1.58 | 1.66 | 1.54 | 1.70 |
| | [ 1.47 , 1.79 ] | [ 1.63 , 1.67 ] | [ 1.49 , 1.70 ] | [ 1.65 , 1.76 ] |
| Mean | 1.94 | 1.94 | 1.89 | 2.02 |
| | [ 1.81 , 2.08 ] | [ 1.93 , 1.96 ] | [ 1.77 , 2.03 ] | [ 1.99 , 2.07 ] |
| Median | 1.90 | 1.98 | 1.86 | 2.04 |
| | [ 1.74 , 2.05 ] | [ 1.96 , 1.98 ] | [ 1.74 , 2.01 ] | [ 2.01 , 2.09 ] |
| Third Quartile | 2.25 | 2.24 | 2.16 | 2.33 |
| | [ 2.02 , 2.47 ] | [ 2.23 , 2.26 ] | [ 2.08 , 2.28 ] | [ 2.28 , 2.36 ] |
| College | | | | |
| First Quartile | 1.88 | 1.79 | 1.87 | 1.89 |
| | [ 1.74 , 2.19 ] | [ 1.75 , 1.81 ] | [ 1.72 , 2.06 ] | [ 1.76 , 1.99 ] |
| Mean | 2.16 | 2.07 | 2.25 | 2.34 |
| | [ 1.99 , 2.30 ] | [ 2.05 , 2.08 ] | [ 2.06 , 2.49 ] | [ 2.27 , 2.42 ] |
| Median | 2.36 | 2.09 | 2.17 | 2.37 |
| | [ 2.11 , 2.33 ] | [ 2.06 , 2.10 ] | [ 2.04 , 2.35 ] | [ 2.27 , 2.46 ] |
| Third Quartile | 2.40 | 2.34 | 2.47 | 2.73 |
| | [ 2.30 , 2.53 ] | [ 2.31 , 2.36 ] | [ 2.26 , 2.75 ] | [ 2.69 , 2.82 ] |

| | Age 32 | | | |
|---|---|---|---|---|
| | NLS 1980 | PUMS 1980 | NLSY 1990 | PUMS 1990 |
| High School | | | | |
| First Quartile | 1.59 | 1.71 | 1.62 | 1.76 |
| | [ 1.39 , 1.92 ] | [ 1.69 , 1.75 ] | [ 1.55 , 1.79 ] | [ 1.70 , 1.84 ] |
| Mean | 1.98 | 1.99 | 2.01 | 2.07 |
| | [ 1.80 , 2.18 ] | [ 1.98 , 2.01 ] | [ 1.82 , 2.09 ] | [ 2.03 , 2.11 ] |
| Median | 1.95 | 2.04 | 1.97 | 2.10 |
| | [ 1.61 , 2.35 ] | [ 2.02 , 2.04 ] | [ 1.84 , 2.10 ] | [ 2.08 , 2.15 ] |
| Third Quartile | 2.42 | 2.28 | 2.19 | 2.40 |
| | [ 2.16 , 2.61 ] | [ 2.26 , 2.30 ] | [ 2.10 , 2.41 ] | [ 2.34 , 2.43 ] |
| College | | | | |
| First Quartile | 2.11 | 1.85 | 2.11 | 2.03 |
| | [ 1.74 , 2.36 ] | [ 1.83 , 1.88 ] | [ 1.77 , 2.28 ] | [ 1.87 , 2.11 ] |
| Mean | 2.36 | 2.15 | 2.39 | 2.46 |
| | [ 2.20 , 2.53 ] | [ 2.13 , 2.17 ] | [ 2.24 , 2.54 ] | [ 2.39 , 2.53 ] |
| Median | 2.35 | 2.16 | 2.37 | 2.47 |
| | [ 2.11 , 2.55 ] | [ 2.14 , 2.18 ] | [ 2.26 , 2.60 ] | [ 2.42 , 2.56 ] |
| Third Quartile | 2.65 | 2.43 | 2.72 | 2.88 |
| | [ 2.40 , 2.91 ] | [ 2.40 , 2.45 ] | [ 2.56 , 2.87 ] | [ 2.79 , 2.97 ] |

Note: bootstrap 95% confidence intervals based on 10000 simulations in parentheses.

| Table 4. 95% Confidence Intervals on Log Wages - 1980 - | | | | |
|---|---|---|---|---|
| **6 to 10 years of experience** | | | | |
| Education | Ability | From NLS | Using PUMS | Reduction in Confidence Intervals (%) |
| High School | Low | [ 1.79 , 2.68 ] | [ 1.79 , 2.01 ] | 75.8 |
| High School | Middle | [ 2.04 , 2.39 ] | [ 2.04 , 2.26 ] | 36.4 |
| High School | High | [ 1.77 , 2.23 ] | [ 1.77 , 2.14 ] | 21.2 |
| College | Low | | [ 1.91 , 2.20 ] | ∞ |
| College | Middle | [ 2.10 , 2.42 ] | [ 2.10 , 2.25 ] | 53.4 |
| College | High | [ 2.22 , 2.40 ] | [ 2.22 , 2.40 ] | 0 |
| **11 to 16 years of experience** | | | | |
| Education | Ability | From NLS | Using PUMS | Reduction in Confidence Intervals (%) |
| High School | Low | [ 1.93 , 2.12 ] | [ 1.93 , 2.12 ] | 0 |
| High School | Middle | [ 2.04 , 2.23 ] | [ 2.04 , 2.23 ] | 0 |
| High School | High | [ 1.97 , 2.25 ] | [ 1.97 , 2.14 ] | 37.7 |
| College | Low | | [ 2.06 , 2.37 ] | ∞ |
| College | Middle | [ 2.36 , 2.65 ] | [ 2.36 , 2.60 ] | 15.8 |
| College | High | [ 2.32 , 2.58 ] | [ 2.32 , 2.58 ] | 0 |

| Table 5. 95% Confidence Intervals on Log Wages - 1990 - 6 to 10 years of experience | | | |
|---|---|---|---|
| Education | Ability | From NLS | Using PUMS |
| High School | Low | [ 1.66 , 1.76 ] | [ 1.66 , 1.76 ] |
| High School | Middle | [ 1.85 , 1.99 ] | [ 1.85 , 1.99 ] |
| High School | High | [ 1.83 , 2.03 ] | [ 1.83 , 2.03 ] |
| College | Low | | [ 1.40 , 2.18 ] |
| College | Middle | [ 1.84 , 2.11 ] | [ 1.84 , 2.11 ] |
| College | High | [ 2.17 , 2.36 ] | [ 2.17 , 2.36 ] |

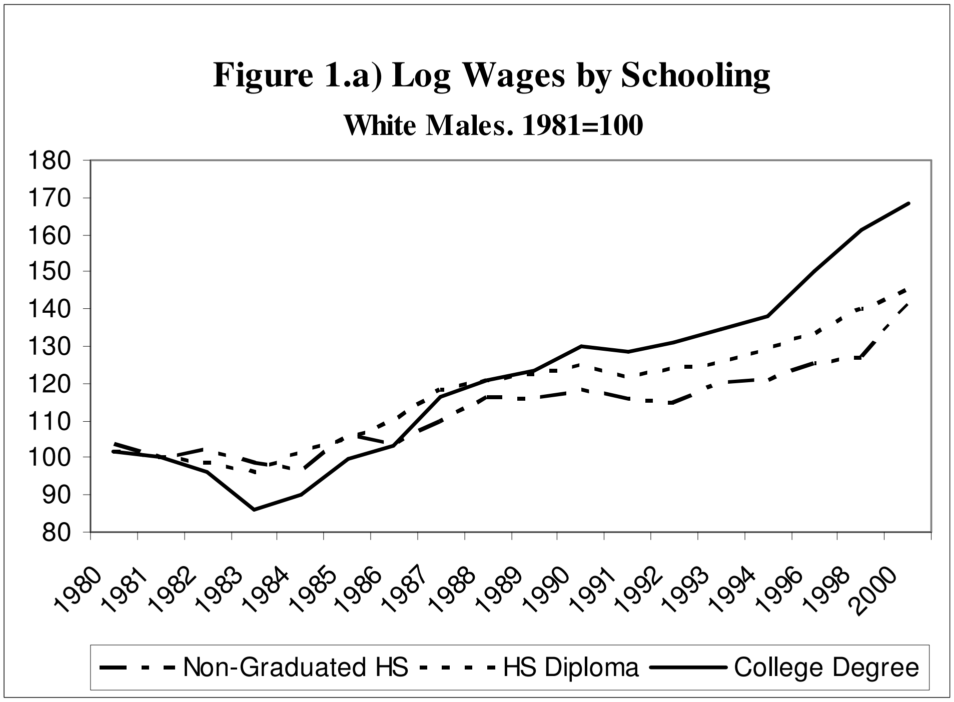
| Table 6. 95% Confidence Intervals on Log Wages - 2000 - 11 to 16 years of experience | | | | |
|---|---|---|---|---|
| Education | Ability | From NLS | Using PUMS | Reduction in Confidence Intervals (%) |
| High School | Low | [ 1.70 , 2.12 ] | [ 1.70 , 2.12 ] | 0 |
| High School | Middle | [ 1.66 , 2.19 ] | [ 1.66 , 2.19 ] | 0 |
| High School | High | | [ 1.53 , 2.10 ] | ∞ |
| College | Low | | [ 2.35 , 2.38 ] | ∞ |
| College | Middle | [ 2.25 , 2.62 ] | [ 2.25 , 2.54 ] | 22.1 |
| College | High | [ 2.49 , 2.69 ] | [ 2.49 , 2.64 ] | 0 |

| Table 7. Log Wages and Log Wage Gap by Schooling-Ability. 1980 - 1990 6 to 10 years of experience | | | | | | |
|---|---|---|---|---|---|---|
| | | 1980 | 1990 | 1980 | 1990 | 1990 with respect to 1980 |
| | | 95 % Confidence Intervals | | Wage Gap | | Wage Gap Increase |
| High School | Low | [ 1.79 , 2.01 ] | [ 1.66 , 1.76 ] | | | |
| High School | Middle | [ 2.04 , 2.26 ] | [ 1.85 , 1.99 ] | | | |
| High School | High | [ 1.77 , 2.14 ] | [ 1.83 , 2.03 ] | | | |
| College | Low | [ 1.91 , 2.20 ] | [ 1.40 , 2.18 ] | 0 | 0 | 0 |
| College | Middle | [ 2.10 , 2.25 ] | [ 1.84 , 2.11 ] | 0 | + | 0 |
| College | High | [ 2.22 , 2.40 ] | [ 2.17 , 2.36 ] | + | + | 0 |

Note: Statistically significant changes in the Wage Gap are indicated with "+"; declines with "-"; a zero denotes no change.

| Table 8. Log Wages and Log Wage Gap by Schooling-Ability. 1980 -2000 11 to 16 years of experience | | | | | | |
|---|---|---|---|---|---|---|
| | | 1980 | 2000 | 1980 | 2000 | 2000 with respect to 1980 |
| | | 95 % Confidence Intervals | | Wage Gap | | Wage Gap Increase |
| High School | Low | [ 1.93 , 2.12 ] | [ 1.70 , 2.12 ] | | | |
| High School | Middle | [ 2.04 , 2.23 ] | [ 1.66 , 2.19 ] | | | |
| High School | High | [ 1.97 , 2.14 ] | [ 1.53 , 2.10 ] | | | |
| College | Low | [ 2.06 , 2.37 ] | [ 2.35 , 2.38 ] | 0 | 0 | 0 |
| College | Middle | [ 2.36 , 2.60 ] | [ 2.25 , 2.54 ] | 0 | + | 0 |
| College | High | [ 2.32 , 2.58 ] | [ 2.49 , 2.64 ] | + | + | 0 |

Note: Statistically significant changes in the Wage Gap are indicated with "+"; declines with "-"; a zero denotes no change.

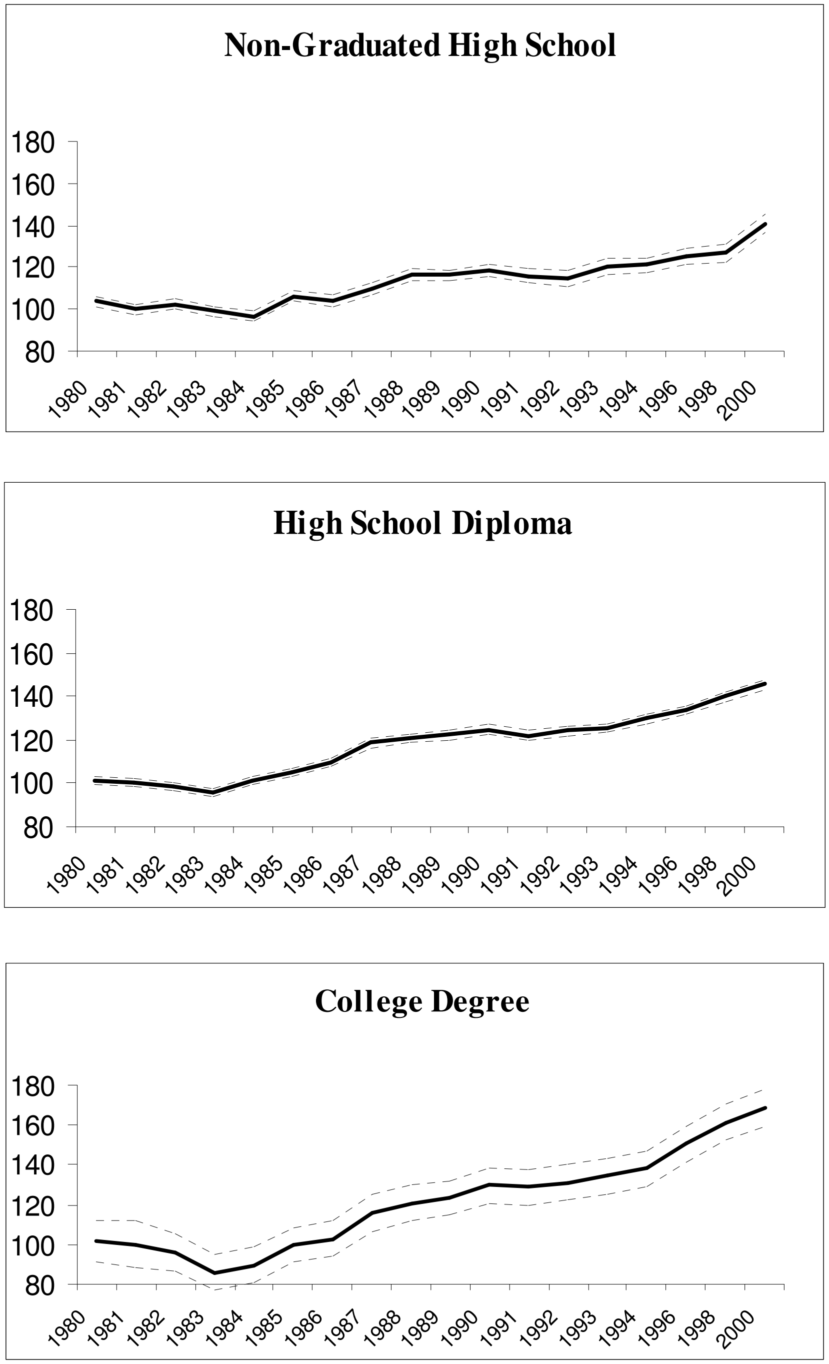



Source: NLSY79

**Figure 1.b) Log Wages by Schooling**

White Males. 1981 =100

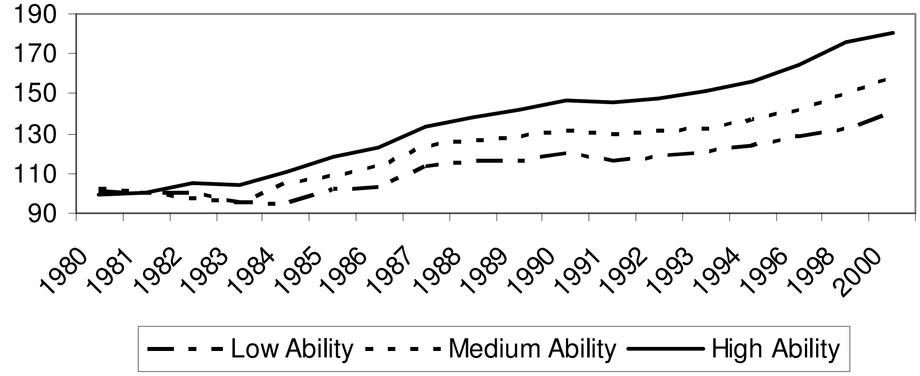
Dashed lines represent the 95% confidence levels




Source: NLSY79

**Figure 2.a) Log Wages by Ability**

**White Males. 1981=100**

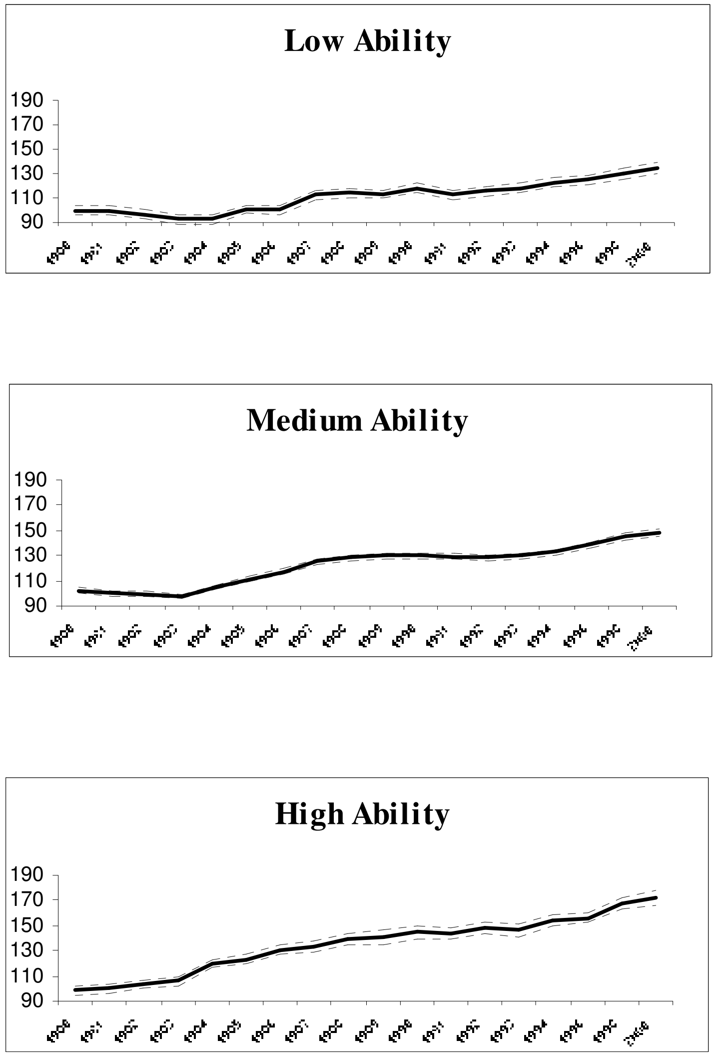



Source: NLSY79

**Figure 2.b) Log Wages by Ability**

White Males. 1981 =100

Dashed lines represent the 95% confidence levels

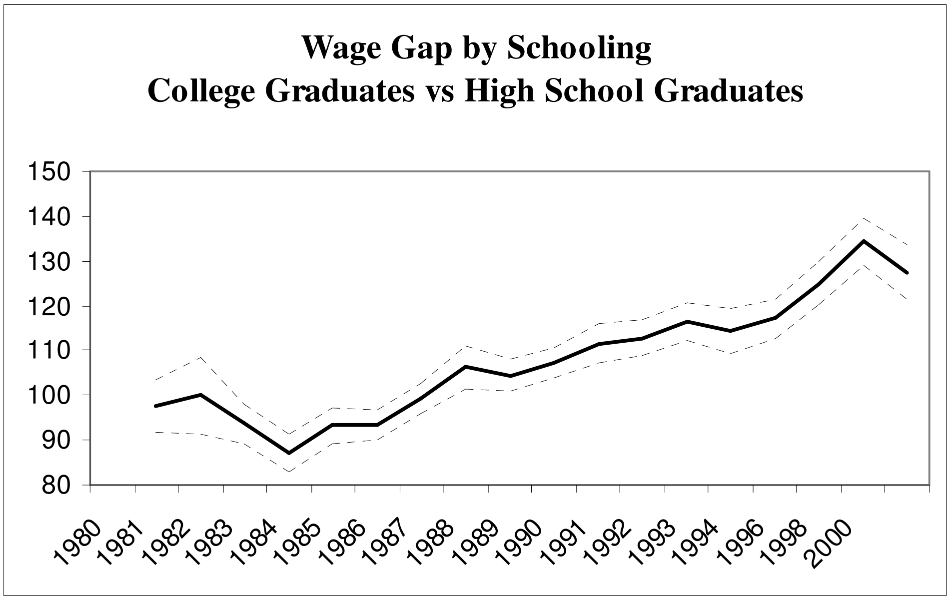


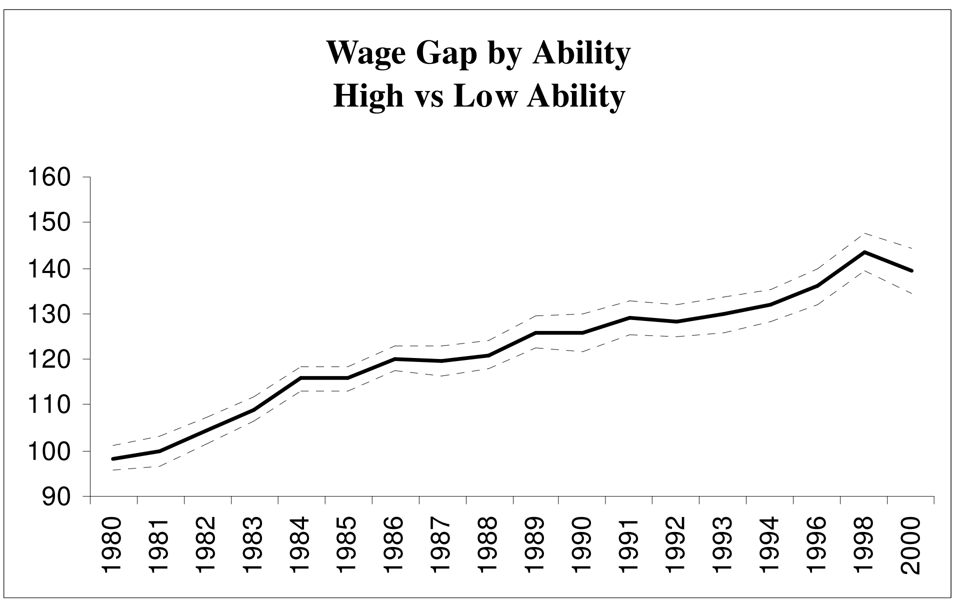

Source: NLSY79

## Figure 3. Wage Gaps

White Males. 1981 = 100

Dashed lines represent the 95% confidence levels



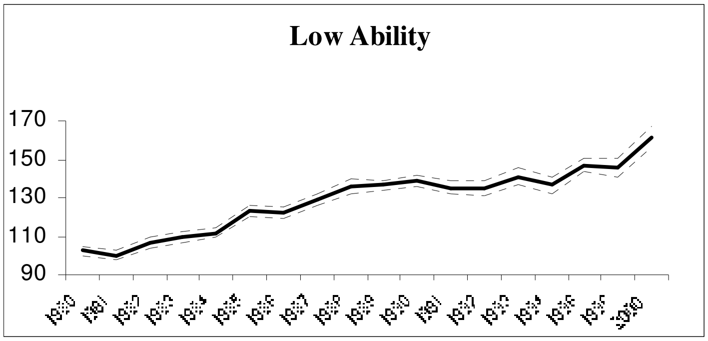



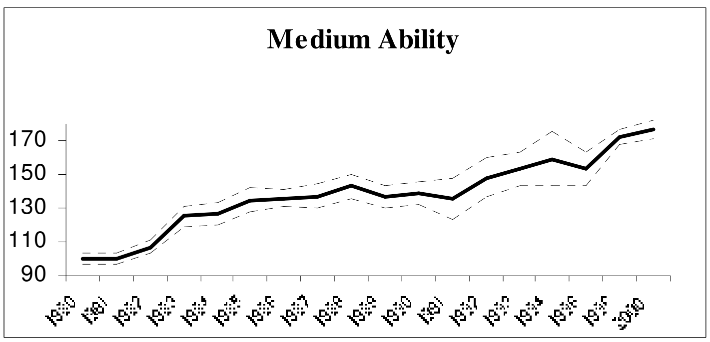

Source: NLSY79

## Figure 4.a) Log Wages: Non-High School Graduates by Ability

White Males. 1981 = 100

Dashed lines represent the 95% confidence levels



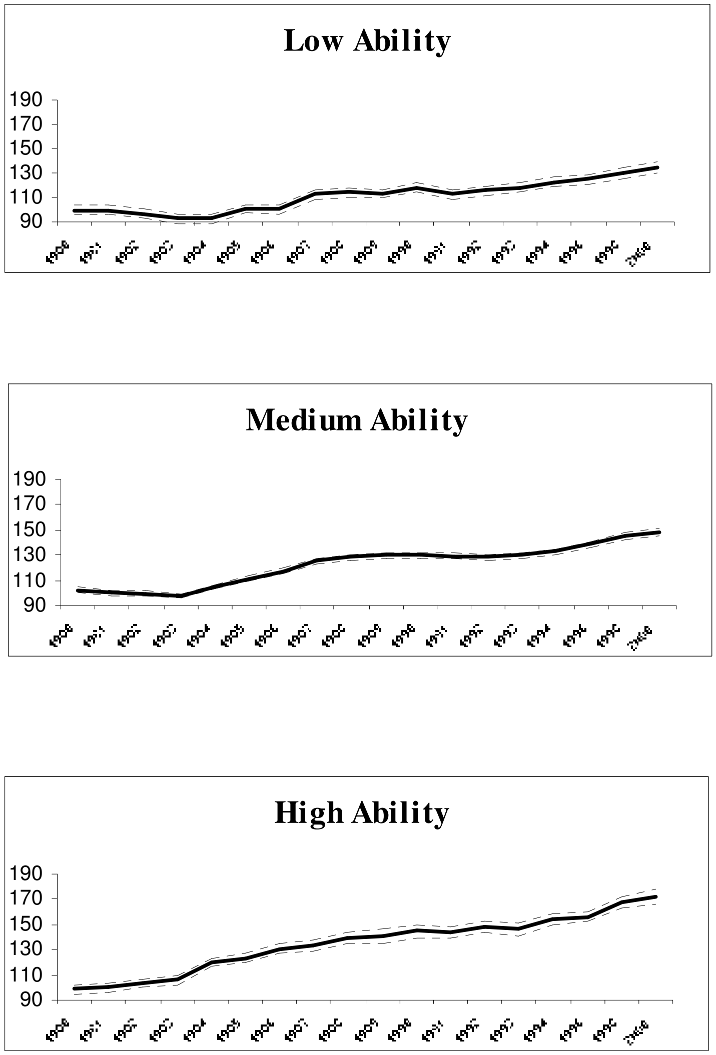




Source: NLSY79

**Figure 4.b) Log Wages: High School Graduates by Ability**

White Males. 1981 = 100

Dashed lines represent the 95% confidence levels

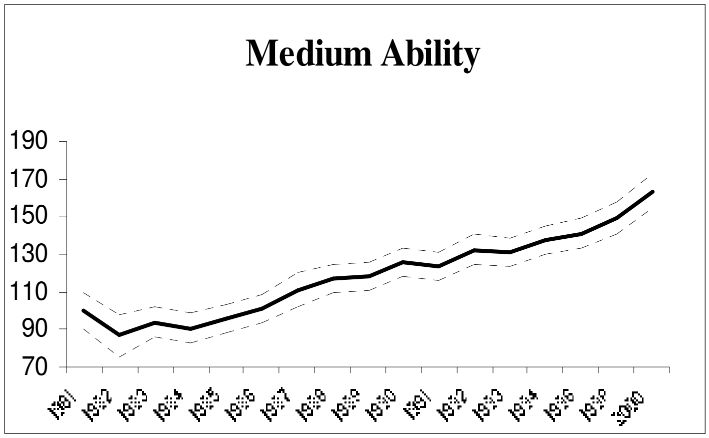


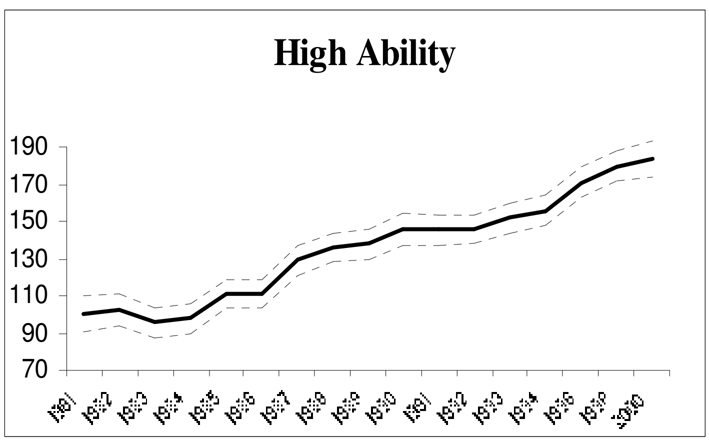

Source: NLSY79

**Figure 4.c) Log Wages: College Graduates by Ability**

White Males. 1981 = 100

Dashed lines represent the 95% confidence levels




Source: NLSY79

**Figure 5. Impact of Ability on Log Wages**

Dashed lines represent the 95% confidence levels

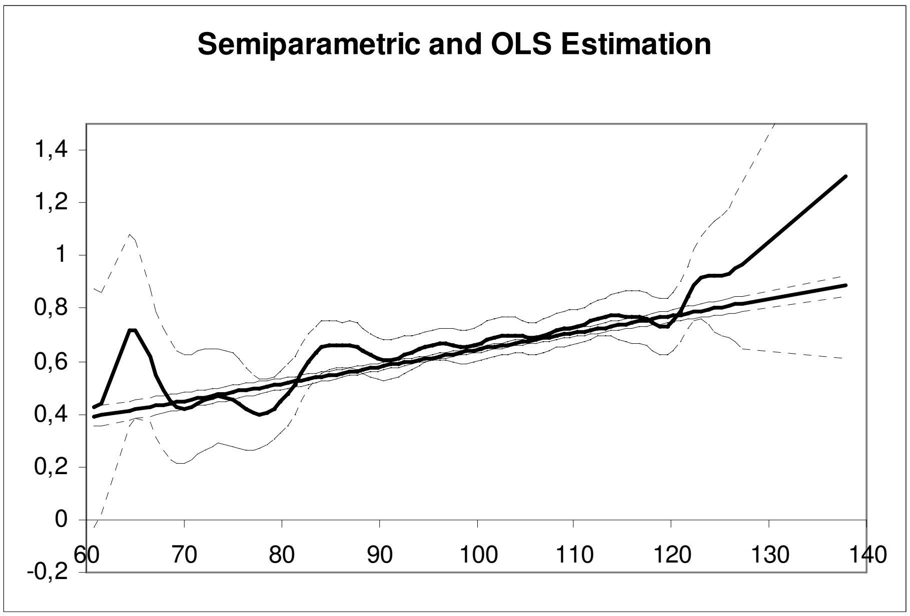
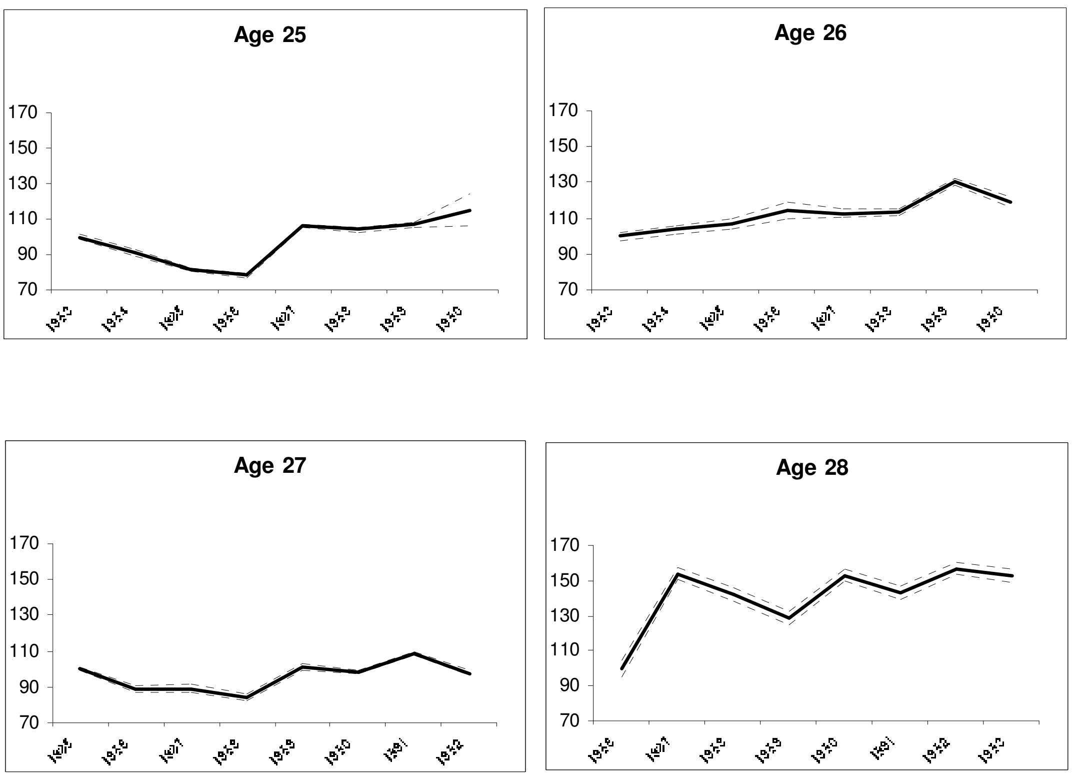
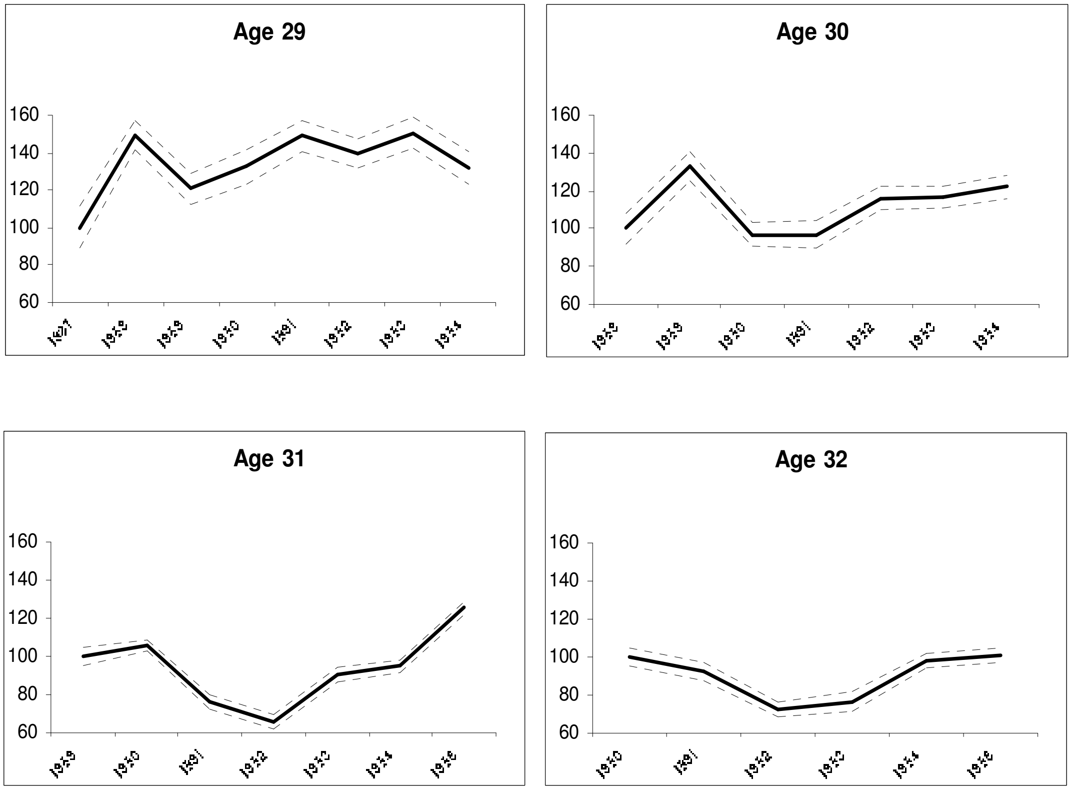


**Figure 6. Log Wages: College-High School Wage Gap**
**Highest Ability. By Age.**
White Males. 1981 = 100
Dashed lines represent the 95% confidence levels






Source: NLSY79