Effects of personal and school characteristics on estimates of the return to education

Joseph G. Altonji

Introduction and summary

Hundreds of studies have shown that more educated workers receive higher wages and earnings than less educated workers.[1] This earnings gap has varied over time but has always been substantial. Recent research by Murphy and Welch (1992) shows that the difference in the average wages of college graduates and high school graduates increased substantially during the 1980s. Rosenbaum (1997) reports an earnings gap of more than 60 percent in the 1990s. However, there is much disagreement on the extent to which the earnings difference is due to the education difference. Does college make people better workers, or are better workers simply more likely to attend college? The wisdom of expanding the higher education system hinges in part on the relative importance of these two explanations of the college / high school wage differential.

There are two main channels through which a spurious correlation between education and wages might arise. First, family background, primary and secondary school quality, and ability might affect both postsecondary schooling and the wage level independent of postsecondary schooling. Second, family background, ability, and primary and secondary school characteristics may affect the rate at which students learn. Students who are more able, from better family backgrounds, or from better schools may choose more postsecondary education than the less advantaged because they receive a larger payoff to a year in college. In this case, the difference in earnings between high school graduates and college graduates will exceed the gain in earnings that a typical high school graduate would receive if he or she had chosen college. See Siebert (1985), Willis (1987), and Griliches (1977) for discussions of these issues.

The empirical evidence on whether controlling for family background and ability reduces estimates of the financial return to education is inconclusive. Much of this literature uses a statistical technique called ordinary least squares (OLS) regression to hold constant other factors while comparing the earnings of people with different levels of education. Many studies show a reduction in the estimated return, but some that have paid attention to the fact that mismeasurement of education becomes a more serious problem when one controls for ability or family background find somewhat smaller levels of bias and, in some cases, obtain higher estimates of the return to education. (See Griliches, 1979, and Siebert, 1985, for surveys.) Ashenfelter and Krueger (1991) and Angrist and Krueger (1991) find that conventional OLS regression estimates, if anything, understate the return to education.[2] These papers and other related recent work have led some to argue that failure to control for ability and background may lead to a substantial *underestimate* of the return to education. As Lang (1993) notes, if well-educated parents push their children to obtain education beyond the point of diminishing returns, then regression estimates of the return to education could be understated.

Joseph G. Altonji is a professor of economics and acting director of the Institute for Policy Research at Northwestern University, a research associate at the National Bureau of Economic Research, and a consultant to the Federal Reserve Bank of Chicago. *The author gratefully acknowledges research support from the National Center on Education and Employment, the Teachers College at Columbia University, under a research contract from the U.S. Department of Education, the Spencer Foundation, and the Institute for Policy Research at Northwestern University. The author thanks Thomas Dunn, Alan Gustman, Bruce Meyer, James Spletzer, and Barbara Wolfe for comments on an earlier draft.*

In contrast to the extensive literature on family background and ability measures, there has been little work on whether failure to control for school quality, secondary school curriculum, and community characteristics leads to bias in estimates of the return to postsecondary education. Most of the data sets that have been used to study the returns to education contain relatively little information about school curriculum and the community. Furthermore, it is hard to envision a data set that would contain measures of all of the relevant school and community characteristics. There are substantial differences across schools in parental and school characteristics that I do observe. (See appendix table 1). One naturally suspects that there are unobserved differences among high schools and communities that influence both education and wages.[3]

Data from the National Longitudinal Survey of the High School Class of 1972 (NLS72) and a matching postsecondary transcript survey (PETS) provide an opportunity to make some progress on this issue. Because the NLS72 contains several students from a large number of high schools, it is possible to statistically control for *all* observed and unobserved characteristics common to students from the same high school. One may also control for characteristics common to students in the same program (that is, academic or nonacademic track) within a given high school. In addition, the data set contains information on parental background, high school curriculum, and test scores. Consequently, I am able to control for a much richer set of factors than previous studies. At the same time, I am able to deal with potential downward bias in estimates of the return to education that would be induced by misreporting of college attendance. I do this by using information on education from PETS along with the sample members' reports of education.

My main conclusion is that controlling for family background leads to a substantial reduction in estimates of the rate of return to postsecondary education, which is defined as the percentage increase in wages that results from a year of college. The OLS estimate of the return to post-secondary academic education falls from 8.2 percent when one does not control for family background to 6.5 percent when one does. The results using the PETS data indicate that measurement error is not responsible for the reduction. Similar reductions are found among the samples of students in high school academic programs and those in nonacademic programs. I conclude that OLS estimates without detailed controls for family background and ability are overstated by about one fourth. It is important to point out, however, that the earnings gap between high school and college graduates has risen since the NLS72 data were collected. Even if the earnings gap between high school graduates and college graduates substantially overstates the return to going to college, that gap has grown so large in recent years that my results imply that college is currently a good financial investment for most people.

My other conclusions are as follows. First, estimates of the rate of return to postsecondary academic education for academic and nonacademic track high school students are remarkably similar. This is true despite the fact that students from academic programs earn substantially more than those from nonacademic programs, even after controlling for observed family background characteristics and achievement and aptitude measures. Second, controlling for high school curriculum does not have much effect on the education coefficients. Third, controlling for the specific high school the student attended has only a modest effect on the rate of return to education. For the combined sample, controlling for these factors reduces estimates of the percentage increase in earnings from a year of college by about 0.5 percentage points (for example, from 6.0 to 5.5). This suggests that failure to control for differences in high school variables does not lead to serious biases in studies of education and wages. This is good news because few data sets permit one to control for these factors.

Below, I present the wage equation that underlies most of the econometric analysis and the econometric methodology used to estimate it. Next, I discuss the data and present estimates of the return to education.

Econometric framework

The empirical analysis is based on a regression model that says that the natural logarithm of the real wage of an individual is determined by years of education, a set of other factors that I observe and can control for statistically, and a set of other factors for which I do not have data. The model takes the form

$$1)\quad W = S\rho + \textit{Effects of Control Variables} + \textit{Error Term},$$

where S is a measure of postsecondary education, such as years of schooling, obtained by a particular individual, ρ is a regression coefficient, and W is the natural logarithm of the real average hourly wage rate in a particular year of a particular person who attended a particular high school. The error term captures the influence of a potentially large number of factors that affect the wage that I do not know about. These factors include characteristics of the high school and the community that are the same for all persons who attended the same high school. Box 1 provides more detail about the form of control variables and the error term of the model.

I wish to estimate the coefficient ρ, where ρ is the effect of an extra year of school on the wage for a randomly selected person. Because wages are measured in natural logarithms, the percentage increase in wages induced by a unit increase in education is approximately equal to 100 *ρ when ρ is smaller than 0.1. The standard approach to estimating ρ is to estimate the effect of an additional year of schooling by OLS regression. OLS estimates of ρ will be biased if the unobserved factors that influence the wage also influence S. S is likely to be positively related to variables that increase the productivity of higher education, lower the direct costs to the student or lower the discount rate, or raise the nonmonetary benefits of education. Consequently, one would expect family background, ability and achievement, course of study in high school, and other high school and community factors to affect not only wage rates but also postsecondary schooling. The evidence for the NLS72 is that they affect both the wage and schooling. (See Altonji, 1988). If one does not adjust for these factors by including them in the set of control variables in equation 1, then S will “get credit” for them when one uses OLS to estimate the

BOX 1

The wage regression

The log wage rate is determined by

$$2)\quad W_{iht} = X_{ih} B_1 + C_{ih} B_2 + S_{ih}\rho + Z_h G + \omega_{iht} + S_{ih}\rho_i + S_{ih}\rho_h ,$$

where I have suppressed controls for labor market experience and the year.

In equation 2, W_{iht} is the log of the real average hourly wage rate of person i from high school h in year t. The vector X_{ih} contains controls for whether the individual is female, black and / or Hispanic, a set of family background characteristics, location, and a set of aptitude and achievement measures. The elements of C_{ih} are measures of the high school curriculum taken by person i. S_{ih} is a measure of postsecondary education, such as years of schooling, and Z_h is a vector of observed high school and community characteristics. The vectors B_1, B_2, and G and the variable ρ are regression coefficients.

The composite error component ω_{iht} is

$$3)\quad \omega_{iht} = \upsilon_i + \upsilon_h + m_h + \varepsilon_{iht};\qquad \upsilon'_i = \upsilon_i + \upsilon_h ,$$

where υ'_i is an index of student and family specific factors that affect W_{iht} independently of the high school and community environment, υ_h is the mean in the high school of υ'_i, υ_i is the difference between υ'_i and υ_h for student i, m_h is an index of high school and community factors that affect W_{iht}, and ε_{iht} is a transitory error component that is assumed to be uncorrelated with all explanatory variables in the wage equation and with the other error components. The component υ_i is uncorrelated with υ_h and m_h by construction.

There are two additional error components in the wage equation. The rate of return to education $\rho + \rho_i + \rho_h$ varies across individuals and depends on an individual-specific component ρ_i and a high school component ρ_h, where ρ_i and ρ_h are uncorrelated by construction and have means of 0. The unobserved term $S_{ih}(\rho_i + \rho_h)$ is treated as part of the wage equation error in estimation. Below, I allow ρ to depend on whether a student is in an academic or nonacademic program by estimating separate equations for these groups. My econometric methods assume that variation in ρ_i and ρ_h is unrelated to S_{ih}.

effect of a change in *S* on the wage, and the estimate of ρ will be too large. In fact, many studies of the return to education have few controls for ability, family background, curriculum in high school, and other characteristics of the high school and community. Even when one uses a rich data set such as the NLS72, the fact both education and wages are influenced by *observed* measures of family background, student achievement, and the high school environment suggests that unobserved determinants of education are correlated with the wage error term. This is because the observed measures are likely to be incomplete or unreliable.

In the empirical work below, I systematically add controls for family background, curriculum in high school, aptitude and achievement, and observed high school characteristics to the wage equation and examine the sensitivity of estimates of the return to education to choice of control variables. I also use a statistical procedure called ordinary least squares-fixed effects (OLS-fixed effects) to control for the influence of unobserved factors that are common to students who attended the same high school. Specifically, I add a set of indicator variables ("dummy" variables) to the set of control variables, one for each high school in the sample. The indicator variable for a particular high school takes on the value 1 if the individual attended that high school and 0 otherwise. The indicator variables will absorb the effects of all factors that are common to students who attended the same high school.[4] Essentially, the OLS-fixed effects procedure estimates the effect of education on wages by relating differences in wages to differences in education across individuals who attended the *same* high school. I present separate estimates for students who were in the academic track and for students who were in the nonacademic track in high school, as well as for the combined sample. The OLS-fixed effects estimates for a specific track relate differences in wages among students who were in the same track in the same high school to differences in their postsecondary education.

Unfortunately, the use of high school fixed effects does not eliminate all of the factors that could lead to biased estimates of ρ. Even after one controls for observed measures of family background and aptitude and achievement, unobserved ability differences among students from the same high school may affect both *S* and the log wage. Furthermore, the quality of instruction and peer group experiences of students probably varies substantially even within a track in a given high school, so the fixed effect analysis does not control for all high school characteristics that influence particular students.[5] However, this study goes further than previous studies by controlling for high school and high school track-specific observed and unobserved variables and for high school curriculum.

The fact that people sometimes misreport years of schooling poses an additional problem.

BOX 2

The instrumental variables estimator

The mechanics of the IV estimator are as follows. First, I regress the person's report of *S* on the PETS measures of education and the control variables in the wage model. When high school indicators are included in the wage equation, I include them in the first stage regression for *S* along with the transcript measures. Then I use the predicted values from this first stage regression as the measure of *S* when I estimate the wage model. I use the transcript information as instrumental variables rather than as direct measures of education because PETS was not successful in obtaining transcripts for all students who claimed to have attended postsecondary schools, in some cases due to lack of cooperation from the schools. Consequently, the PETS measure of postsecondary education will also differ from actual schooling. If the measurement errors in the PETS data are uncorrelated with the information on years of schooling and degree attainment provided by the student, then the use of the predicted measure of *S* will eliminate the bias from measurement error.[1]

[1]Students were asked during each follow-up survey to identify any schools that they were attending or had attended. Correlated measurement errors could arise if a student attended college but said that he or she did not. In this case the student would not provide the name of the postsecondary school attended and no transcript would be found. I assume that people do not hide the fact that they attended college if they attended college for a significant period of time.

Measurement error in *S* will bias the estimate of ρ toward 0. This is because the "noise" in *S* will reduce the sample correlation between wages and *S*. The inclusion of controls for the high school, curriculum, family background, and test scores may exacerbate downward bias in the education coefficient arising from measurement error in education, because much of the true variation in schooling will be correlated with these controls while the measurement error will not. I address the measurement error issue by using the independent information about educational attainment in the PETS that accompanied the NLS72 to create "instruments" for the education measures and then estimate the wage model by the method of instrumental variables (IV) instead of OLS, as described in box 2.

Data: National Longitudinal Survey of the High School Class of 1972

The NLS72 is a Department of Education survey of individuals who were high school seniors during the spring of 1972. Thus, high school dropouts are excluded. The individuals were resurveyed in 1973, 1974, 1976, and 1979. A subsample was resurveyed in 1986.[6]

The key variables used in the study are listed in table 1. Note that the indicator variables VOC79, SOC1479, SOC1579, COLL79, and ADV79 are mutually exclusive. I also construct a set of education measures from PETS to use as instruments.[7]

The control variables for region and city size, family background, aptitude and achievement measures, high school curriculum (semester hours in each of eight subjects), and high school characteristics are listed in the footnotes to the tables. Descriptive statistics and variable definitions are provided in appendix table 1. Only the education coefficients are shown in tables 2 and 3.

TABLE 1

Definitions of key variables

Variable	Definition
W	Natural logarithm of the real hourly wage rate.
YRSACD79	Years of postsecondary academic education completed by 1979.
YRSVOC79	Years of postsecondary vocational education by 1979.
VOC79	Indicator variable that equals 1 if a person attended a postsecondary vocational education program and did not attend a postsecondary academic program.
SOC1479	Indicator variable that equals 1 if a person has less than two years of college (regardless of whether the person also attended vocational school) and 0 otherwise.
SOC1579	Indicator variable that equals 1 if a person attended college for two or more years but did not receive a four-year degree and 0 otherwise.
COLL79	Indicator variable that equals 1 if a person received a four-year degree but did not receive an advanced degree and 0 otherwise.
ADV79	Indicator variable that equals 1 if a person received a graduate degree and 0 otherwise.

Estimates of the return to education

Table 2 presents OLS estimates of the effects of YRSACD79 and YRSVOC79. Columns 1–4 do not include dummy variables for each high school, while columns 5–7 do. All equations contain controls for race, sex, experience, and the year the wage data refer to.[8] The column headings indicate whether controls for region and city size (region), family background and achievement and aptitude measures (family / achievement), and high school curriculum and high school characteristics (high school) are included. The high school indicators absorb the effect of any variables that are constant within the high school, and so region and city size and fixed high school characteristics are implicitly controlled for in columns 5–7.

The returns to academic education

The coefficients in the table for YRSACD79 are estimates of the average amount that the log wage rises in response to an extra year of academic postsecondary education. For example, the coefficient on YRSACD79 is .0817 when only the basic controls are included (column 1). This coefficient implies that spending an extra year in college raises the log wage by .0817. This translates into an increase in the wage of about 8 percent. This is typical of estimates from other data sets for the year 1980, which is in the middle of the time period that the wage data are drawn from. The coefficient falls to .0653 when family background and ability and aptitude

TABLE 2

Effect of education on wages: OLS estimates (Dependent variable: log wage)

	OLS				OLS-fixed effects (Constants for each high school)		
	Basic controls	Region	Family/ achievement, region	High school, family/ achievement, region	Region	Family/ achievement, region	High school, family/ achievement, region
				Combined sample			
YRSACD79	.0817	.0790	.0653	.0644	.0749	.0605	.0598
	(.0028)	(.0028)	(.0035)	(.0036)	(.0029)	(.0036)	(.0036)
YRSVOC79	.0145	.0150	.0133	.0135	.0173	.0163	.0154
	(.0053)	(.0052)	(.0052)	(.0052)	(.0053)	(.0052)	(.0052)
				Students in academic programs			
YRSACD79	.0731	.0734	.0636	.0637	.0663	.0568	.0567
	(.0043)	(.0042)	(.0047)	(.0047)	(.0046)	(.0050)	(.0051)
YRSVOC79	.0133	.0152	.0165	.0166	.0183	.0201	.0180
	(.0076)	(.0074)	(.0075)	(.0075)	(.0081)	(.0081)	(.0080)
				Students in nonacademic programs			
YRSACD79	.0689	.0670	.0572	.0563	.0651	.0547	.0550
	(.0046)	(.0046)	(.0056)	(.0057)	(.0053)	(.0065)	(.0065)
YRSVOC79	.0196	.0192	.0162	.0178	.0138	.0121	.0116
	(.0074)	(.0073)	(.0073)	(.0074)	(.0078)	(.0078)	(.0078)

Notes: Region = NO.CENTRAL, SOUTH, WEST, SMLTOWN, MED.CITY, BIGCITY, HUGECITY, MED.SUBURB, BIGSUBURB, HUGESURB, COLL-PROX.

Family/achievement = FATHER-ED, MOTHER-ED, LOWSES, ED-MONEY, MOTHER-WORK, BLUECOLF, ENGLISH, FATH-COLL, MOTH-COLL, DISC-PLANS, PAR-INTEREST, PAR-INFL, IMPTAVER, COLLEGE-ABILITY, TEACHER-ASSESSMENT, VOCABULARY, PICTURE.NUMB, READING, LETTER.GROUP, MATH, MOSAIC.COMP, HOMEWORK, and dummy variables for whether data were missing for FATH-COLL, MOTH-COLL, or BLUECOLF.

High school characteristics include controls for the level and square of the fraction of the student body who are black, the student/teacher ratio, whether the school is private or parochial, the number of grades in the high school, the daily attendance rate, the dropout rate, the teacher turnover rate, the fraction of teachers with master's or Ph.D. degrees, the availability of advanced science courses, the number of students in the school, and the means across students of the number of courses taken between tenth and twelfth grade in science, foreign language, social studies, English, mathematics, industrial arts, commercial arts, and fine arts.

The coefficients in the table are estimates of the effect of additional years of education on the log wage. The combined sample contains 38,595 person-year observations on 9,239 students from 897 high schools. The academic sample contains 18,653 person-year observations on students from the academic programs in 858 high schools. The nonacademic sample contains 19,942 person-year observations on students from the vocational or general programs in 864 high schools. Summary statistics and variable definitions are given in appendix table 1.

All equations include BLACK, HISP, CSEX, a quadratic in years of work experience, and a quadratic in the calendar year that the wage measure refers to.

Variables that do not vary across high schools, such as the region variables and the high school variables noted above, are implicitly controlled for in the equations with high school dummies.

"White" standard errors in parentheses account for arbitrary forms of heteroscedasticity and correlation across observations on students from a given high school.

Source: Author's calculations based on data from the *National Longitudinal Survey of the High School Class of 1972* (U.S. Department of Education, 1972–86).

measures are added, a decline of .0164. This reduction is consistent with the findings of most other studies that have used detailed controls for family background and ability or made use of sibling pairs.[9] On the other hand, adding controls for the student's courses and a set of high school characteristics lowers the YRSACD79 coefficient by only .0009 to .0644.

Does the fact that almost all studies of the economic value of college fail to control for unobserved high school and community characteristics matter? The answer is that there is only a small upward bias without these controls. For example, when one adds a separate constant term (or fixed effect) for each high school to the specification in column 2, which does not

TABLE 3

Effect of education on wages: Instrumental variables estimates (Dependent variable: log wage)

	Instrumental variables estimator				Instrumental variables estimator with fixed effects (Constants for each high school)		
	Basic controls	Region	Family/ achievement, region	High school, family/ achievement, region	Region	Family/ achievement, region	High school, family/ achievement, region
				Combined sample			
YRSACD79	.0817	.0793	.0582	.0572	.0770	.0560	.0551
	(.0031)	(.0031)	(.0037)	(.0037)	(.0033)	(.0039)	(.0040)
YRSVOC79	.0254	.0080	–.0127	–.0151	.0184	–.0104	–.0133
	(.0170)	(.0167)	(.0169)	(.0169)	(.0171)	(.0173)	(.0175)
				Students in academic programs			
YRSACD79	.0765	.0765	.0618	.0615	.0699	.0570	.0568
	(.0058)	(.0056)	(.0062)	(.0062)	(.0065)	(.0071)	(.0072)
YRSVOC79	.0550	.0379	.0338	.0327	.0283	.0336	.0314
	(.0340)	(.0331)	(.0333)	(.0329)	(.0379)	(.0379)	(.0378)
				Students in nonacademic programs			
YRSACD79	.0724	.0722	.0589	.0578	.0758	.0636	.0641
	(.0054)	(.0055)	(.0061)	(.0062)	(.0064)	(.0073)	(.0073)
YRSVOC79	.0180	.0033	–.0074	–.0095	–.0041	–.0143	–.0173
	(.0212)	(.0213)	(.0212)	(.0213)	(.0225)	(.0225)	(.0226)

Notes and source: See table 2. In addition, for columns 1–4 the instruments consist of dummies for whether the individual had a postsecondary transcript, a transcript from a vocational school, a transcript from a two-year public college, a four-year public college, a private college, dummies for whether the individual's highest degree was a license or certificate, an associate degree, a bachelor's degree, or an advanced degree, and a count of the number of transcripts for the individual. The instrumental variables estimator with fixed effects includes dummy variables for each high school in both the instruments and the wage equation.

contain controls for family background, aptitude and achievement, or courses taken, the coefficient on YRSACD79 falls from .0790 to .0749 (see column 5). That is, the estimate of the percentage change in wages induced by an extra year of education falls from 7.9 percent to 7.49 percent. When one controls for background and achievement, the comparable coefficients without and with high school dummies are .0653 and .0605, respectively. When one controls for curriculum and observed high school characteristics, adding the high school constants reduces the coefficient on YRSACD79 from .0644 to .0598. Thus, failure to control for high school differences leads to an upward bias of .005 in the education coefficient, which (multiplying by 100) is an upward bias of 0.5 percentage points in the rate of return to education.

Similar results are obtained for students from academic and nonacademic programs. The coefficients for the two subgroups are remarkably similar. They are also a bit below the coefficients for the combined sample. This reflects the fact that both the wage level and YRSACD79 are positively correlated with whether one is in an academic high school program, even after controlling for background, aptitude and achievement, and semester hours by subject area.

Appendix table 2 reports OLS estimates of the effects of academic education when the dummy variables VOC79, SOC1479, SOC1579, COLL79, and ADV79 are used to parameterize the model. The coefficients on the education variables are all relative to a high school graduate. The results are qualitatively consistent with those based upon the linear specification in table 2.

Instrumental variables estimates

For the combined sample, the use of the transcript measures of education as instruments for the person's report of education has no effect (to four digits) on the estimated return to YRSACD79 when one does not control for family background and test scores. It leads to a slight *reduction* (relative to OLS) in estimates of the return to academic education when one controls for family background and test scores. This implies that the reduction in the education slope from about .079 with only regional controls to .058 when family background and test scores are added is not an artifact of measurement error in the education variable. There is only a small drop in the IV estimate (from .058 to .056) when high school fixed effects are added to the equation with family background and test scores (table 3). The IV results confirm the earlier OLS finding that failure to control for high school and community variables leads to only a small bias in estimates of the return to education.

The use of IV in place of OLS does not significantly change the conclusions for the academic and nonacademic groups.[10] The IV estimates of models that use five dummy variables for education outcomes indicate that controlling for high school makes almost no difference for academic education and, if anything, leads to an increase in the estimated return to vocational education. (See appendix table 3.)

The returns to vocational education

Tables 2 and 3 report OLS and IV estimates of the effect of years of vocational education (YRSVOC79) on wages for the combined sample and the academic and nonacademic subgroups. The mean of YRSVOC79 is .5110 for the combined sample and .5031 and .5183 for the academic and nonacademic subsamples, respectively, which says that the average high school graduate from the class of 1972 obtained about a half year of postsecondary vocational education. For the combined sample, the OLS results for the linear specification indicate a much lower return for vocational education than for academic education, with a coefficient of .0145 in the absence of controls (table 2, column 1), and .0154 when one controls for background, aptitude and achievement, high school curriculum, and the high school (table 2, column 7). These estimates imply that the financial return to spending a year in postsecondary vocational education is only about 1.5 percent. The estimates are similar for students who took an academic program in high school and students who took a nonacademic program. The fact that these estimates *rise* when one adds more detailed control variables is consistent with abundant evidence that less advantaged individuals tend to pursue vocational education.

However, the low estimates of the return to a year of vocational education should be treated cautiously for two reasons. First, vocational education is a very heterogenous category and programs lasting just a few months may be coded as lasting a year. (See Grubb, 1993.) This would lead to downward bias. Second, it is possible that the value of vocational education is lower if one has also obtained academic postsecondary education. This would make sense if the skills acquired in vocational education are not used by students who later pursue academic education. The wage models in appendix tables 2 and 3 that use the indicator variables VOC79, SOC1479, SOC1579, COLL79, and ADV79 as the education measures shed some light on this issue. This is because the vocational education variable, VOC79, excludes individuals who obtained both academic and vocational postsecondary education. It is 1 if the person obtained some vocational education and did not obtain any academic education and 0 otherwise. As a result, the mean of VOC79 is much lower for the academic high school track sample than for the nonacademic track sample, despite the fact that the mean of YRSVOC79 is similar for the two groups. For the combined sample, the OLS coefficient on VOC79 implies that vocational education raises wages by 4.8 percent to 6.5 percent, depending upon what one controls for. I suspect there are differences in the content of postsecondary vocational education for academic track versus nonacademic track students, and these differences may underlie the larger coefficient on VOC79 for the academic sample.

The IV estimates for YRSVOC79 and VOC79 follow the same general pattern as the OLS estimates, but are imprecise, particularly for the academic sample. Some of the point estimates for YRSVOC79 are negative but not statistically significant. However, for the combined sample the coefficient on VOC79 is quite substantial (.1179) when one controls for the high school, family background, curriculum, and test scores, although the standard error is .064 (appendix table 3). A possible explanation (other than sampling error) is that the returns to vocational programs that are sufficiently well established to lead to a transcript and / or a license or certificate are larger than the returns to other programs. The IV estimates give more weight to such programs than the OLS estimates do. Grubb's (1993) analysis of NLS72 suggests substantial heterogeneity in vocational programs. A key policy issue is how to enhance the labor market skills of persons who are not well suited for or interested in academic postsecondary education. The results suggest that some vocational training programs have substantial labor market value for students who specialize in vocational education after high school.

The impact of controlling for high school and community characteristics and for family background and achievement measures on estimates of the return to vocational education is sensitive to whether one uses OLS or IV, to the form of the education variables, and to whether the student was in an academic or nonacademic program in high school. I will not discuss the detailed results in the tables. Part of the problem is that the IV coefficient estimates for VOC79 are very imprecise, particularly for the academic sample.

Conclusion

The OLS and IV estimates with high school fixed effects indicate that only modest biases result from the failure of previous studies to control for differences in high schools and for differences in primary school and community characteristics common to students from the same high school. This is good news for researchers, because few data sets permit one to study clusters of students from the same high school. On the other hand, in contrast to several recent studies, I find that failure to control for family background and aptitude and achievement measures leads one to overestimate the rate of return to college education by about one fourth.

NOTES

[1]Siebert (1985) and Willis (1986) provide surveys of the link between education and earnings.

[2]Ashenfelter and Krueger obtain a 16 percent return to education when they contrast wages of identical twins with different schooling levels and use an instrumental variables scheme based on a twin's estimate of his / her sibling's schooling to deal with measurement error. However, Ashenfelter and Rouse (1997) use a larger sample of twins and obtain estimates closer to those obtained here.

[3]The evidence from Akin and Garfinkel (1977), Morgan and Sirageldin (1968), and Johnson and Stafford (1973) collectively suggests a positive link between school quality proxies and labor market outcomes. Card and Krueger (1992) find that school quality proxies that are related to educational attainment are also related to education slopes.

[4]The standard errors for both the OLS and instrumental variables regressions with and without high school fixed effects allow for arbitrary high school-specific forms of heteroscedasticity, serial correlation, and correlation across students from the same high school.

[5]There is information on tracking in the NLS72, and in future work it would be interesting to use a fixed effect to control for observed and unobserved characteristics that are common to students from the same track in high school. In terms of the model in box 1, the use of fixed effects controls for the high school error component $\upsilon_h + m_h$. It does not eliminate potential bias from the correlation between S_{ih} and the individual error component υ_i or between S_{ih} and the component ρ_i and ρ_h of the rate of return to education.

[6]I restrict the sample to the 16,683 individuals from the schools that participated in the base year survey. The sample is reduced to 15,680 by eliminating observations with missing high school test information and to 12,980 by eliminating individuals who did not respond to all of the first four follow-ups. Information from the 1986 follow-up was then added for persons who were in the earlier sample of 12,980. The yearly wage observations are created using information on earnings divided by hours for 1977, 1978, and 1979, and information on the wage at the beginning and end of each job held between 1980 and 1986 up to a maximum of the four most recent jobs. An observation for 1977 is included if 1) the individual was not a full-time student in October 1976 or October 1977, 2) the number of hours worked in 1977 was greater than 1,040, *and* 3) the log of the 1977 real wage was between $.50 and $75 in 1967 dollars. Observations for 1978 and 1979 were included if they met the corresponding three criteria for 1978 and 1979, respectively. Data for beginning and ending job dates (1980–86) were included if 1) the number of hours worked in the appropriate year was greater than 1,040, and 2) the log of the real wage was between $.50 and $75 in 1967 dollars. Restriction of the sample to cases with complete data on the variables used in the wage analysis reduced the sample size to 38,595 observations on 9,239 individuals from 897 high schools. The subsample of students in academic programs contains 18,653 person–year observations from 858 high schools. The corresponding figures for the nonacademic (general and vocation tracks) subsample are 19,942 and 864.

[7]The variables constructed from the PETS survey include the number of transcripts found for each student and nine indicator variables for whether the student had the following transcript combinations: 1) at least one transcript; 2) a transcript from a nonacademic institution; 3) a transcript from a two-year public academic institution; 4) a transcript from a four-year public ac-

ademic institution; 5) a transcript from a private academic institution; 6) a license or certificate but no academic degree; 7) an associate degree but no bachelor's or advanced degree; 8) a college degree but no advanced degree; and 9) an advanced degree. The PETS survey contains at least one transcript for 83 percent of the sample members who reported some postsecondary education by 1979, 74.8 percent of those who reported vocational education or some college but no degree, and 96.16 percent of those who reported a college or advanced degree. Transcript evidence of a college or advanced degree was found for 82.29 percent of the sample members who reported a college or advanced degree. Transcript evidence of a college or advanced degree was found for 3.16 percent of the sample who did not report a college or advanced degree by 1979. Also, transcript evidence of an advanced degree was found for 8.13 percent of the persons who reported college as their highest degree in 1979, which may in part be due to completion of their advanced degrees after 1979.

[8]I include a quadratic in years of work experience and a quadratic in the calendar year that the wage measure corresponds to.

[9]See Griliches (1979) and Olneck (1979) for discussions of alternative estimates of the return to education based on sibling data.

[10]In a study conducted after the initial drafts of this article were completed, Kain and Rouse (1995) use the NLS72 and PETS and also find that controlling for family background and ability measures leads to a substantial reduction in OLS estimates of the returns to two- and four-year colleges. However, they obtain higher estimates of the return when they use distance from the college and tuition as instrumental variables for college attendance.

APPENDIX

APPENDIX: TABLE 1

Means and standard deviations of wage and education variables

	Combined sample				Academic		Nonacademic	
Explanatory variables	Mean	Standard deviation	Standard deviation in high school	Fraction of variance across high schools	Mean	Standard deviation	Mean	Standard deviation
Wages								
LOGWAGE, log of real average hourly wage, 1967 dollars	.9196	.4635	.4402	.0980	.9916	.4746	.8523	.4425
Education								
YRSACD79, years of postsecondary academic education by 1979	1.988	1.843	1.666	.1829	2.936	1.753	1.101	1.439
YRSVOC79, years of postsecondary vocational education by 1979	.5309	.7641	.7183	.1163	.5228	0.794	0.5385	.7347
VOC79, 1 if some vocational, no college	.0851	–	–	–	.0381	–	.1290	–
SOC1479, 1 if less than 2 years college	.1803	–	–	–	.1413	–	.2168	–
SOC1579, 1 if more than 2 years college, no degree	.1738	–	–	–	.1995	–	.1497	–
COLL79, 1 if college degree, no advanced degree	.3022	–	–	–	.4928	–	.1240	–
ADV79, 1 if advanced degree	.0285	–	–	–	.0534	–	.0053	–
Gender and race/ethnicity								
BLACK	.0892	–	–	–	.0663	–	.1106	–
HISP	.0366	–	–	–	.0213	–	.0509	–
FEMALE	.4916	–	–	–	.4780	–	.5043	–
Family background								
FATHER-ED, father's education	12.75	2.545	2.171	.2723	13.49	2.632	12.06	2.250
MOTHER-ED, mother's education	12.43	.098	1.854	.2191	12.96	2.158	11.93	1.190
LOWSES, 1 if low SES	.2340	–	–	–	.1343	.3410	.3273	.4692
ED-MONEY, 1 if worry over money interfered with high school education	.2891	–	–	–	.2261	.4183	.3480	.4764
MOTHER-WORK, 1 if mother worked while in elementary school	.4021	–	–	–	.3804	.4855	.4223	.4939
BLUECOLF, 1 if father blue collar	.3216	–	–	–	.2925	.4549	.3488	.4766
ENGLISH, 1 if English spoken at home	.9207	–	–	–	.9183	.2739	.9230	.2666
FATH-COLL, 1 if father wants college or grad school	.5787	–	–	–	.7814	.4133	.3890	.4875
MOTH-COLL, 1 if mother wants college or grad school	.6140	–	–	–	.8183	.3856	.4229	.4940
DISC-PLANS, 1 if often discussed plans with parents	.7902	–	–	–	.8510	.3561	.7332	.4423
PAR-INTEREST, 1 if uninterested parents interfered with high school	.2019	–	–	–	.1272	.3332	.2718	.4449
PAR-INFL, 1 if parents influenced post high school plans a great deal	.4397	–	–	–	.5037	.5000	.3799	.4854
Geographic variables								
SMLTOWN	.2953	–	–	–	.3027	–	.2883	–
MED.CITY	.0832	–	–	–	.0863	–	.0802	–
MED.SUBURB	.0487	–	–	–	.0596	–	.0386	–
BIGCITY	.1020	–	–	–	.0955	–	.1080	–
BIGSUBURB	.1046	–	–	–	.1171	–	.0930	–
HUGECITY	.0785	–	–	–	.0890	–	.0686	–
HUGESURB	.0950	–	–	–	.1127	–	.0784	–
NO.CENTRAL	.2916	–	–	–	.2750	–	.3072	–
SOUTH	.3176	–	–	–	.2737	–	.3585	–
WEST	.1678	–	–	–	.1461	–	.1880	–
COLL-PROX	1.785	–	–	–	1.701	–	1.863	–
Aptitude and achievement measures								
IMPTAVER, grades	15.64	7.586	3.469	.7909	14.44	7.437	16.76	7.553
COLLEGE-ABILITY, 1 if definitely college material; 5 if definitely not	1.843	.9659	.8984	.1349	1.477	.6991	2.186	1.052

(Cont. on following page)

APPENDIX: TABLE 1 (cont.)

Means and standard deviations of wage and education variables

	Combined sample				Academic		Nonacademic	
Explanatory variables	Mean	Standard deviation	Standard deviation in high school	Fraction of variance across high schools	Mean	Standard deviation	Mean	Standard deviation
TEACHER-ASSESSMENT, 1 if teacher expectation high; 5 if low	2.085	.8701	.8227	.1060	1.816	.7948	2.337	.8621
VOCABULARY	52.31	9.896	8.640	.2377	56.34	9.328	48.54	8.872
PICTURE.NUMB, associative memory	51.57	9.680	8.938	.1474	53.75	9.138	49.54	9.729
READING	52.31	9.424	8.458	.1945	56.07	8.321	48.80	9.033
LETTER.GROUP, inductive reasoning	52.33	8.878	8.034	.1811	55.33	7.082	49.54	9.456
MATH, quantitative comparisons (basic competence in math)	52.50	9.539	8.486	.2086	57.06	7.939	48.25	8.924
MOSAIC.COMP, perceptual speed and accuracy	51.46	9.187	7.407	.3499	53.11	8.681	49.92	9.377
HOMEWORK, hours on homework per week	4.467	3.278	3.018	.1523	5.315	3.442	3.674	2.899

Notes: Means and standard deviations of variables used in the wage equations for the full sample, the academic sample, and the nonacademic sample. The combined wage sample contains 38,595 observations on 9,239 individuals from 897 high schools. The academic (nonacademic) sample contains 18,653 (19,942) observations on 4,292 (4,947) individuals from 858 (865) high schools. The table also reports the standard deviation of each variable within a high school, and the fraction of the sample variance that is across high schools. The standard deviations and the variance decomposition in the table refer to the cross section–time series sample, to which individuals contribute different numbers of observations. Consequently, they provide only a rough indication of relative importance of variation within the high school and variation across high schools in wages, education, and background characteristics. (See Altonji, 1988, for a more thorough treatment of this issue.) However, the results indicate that there is substantial variation across high schools in background characteristics, aptitude and achievement measures, and curriculum. Note also that there are substantial differences in the means for the academic and nonacademic samples.

Source: See text table 2.

APPENDIX: TABLE 2

Effect of postsecondary education on wages: OLS estimates (Dependent variable: log wage)

	OLS estimator				OLS-fixed effects (Constants for each high school)		
	Basic controls	Region	Family/ achievement, region	High school, family/ achievement, region	Region	Family/ achievement, region	High school, family/ achievement, region
				Combined sample			
VOC79	.0634	.0615	.0479	.0479	.0658	.0528	.0521
	(.0137)	(.0134)	(.0135)	(.0134)	(.0138)	(.0137)	(.0137)
SOC1479	.0773	.0574	.0241	.0251	.0479	.0149	.0146
	(.0113)	(.0111)	(.0119)	(.0120)	(.0118)	(.0125)	(.0127)
SOC1579	.1880	.1682	.1191	.1160	.1521	.1017	.0486
	(.0120)	(.0117)	(.0133)	(.0134)	(.0125)	(.0141)	(.0143)
COLL79	.3437	.3285	.2571	.2542	.3130	.2390	.2354
	(.0126)	(.0126)	(.0155)	(.0158)	(.0132)	(.0159)	(.0162)
ADV79	.5057	.4908	.4101	.4040	.4658	.3781	.3736
	(.0291)	(.0289)	(.0310)	(.0313)	(.0299)	(.0317)	(.0319)
				Students in academic programs			
VOC79	.1525	.1507	.1342	.1297	.1840	.1629	.1568
	(.0310)	(.0314)	(.0330)	(.0325)	(.0329)	(.0338)	(.0337)
SOC1479	.0876	.0701	.0462	.0477	.0711	.0492	.0440
	(.0244)	(.0245)	(.0258)	(.0257)	(.0269)	(.0270)	(.0274)
SOC1579	.1933	.1757	.1427	.1403	.1560	.1254	.1185
	(.0237)	(.0236)	(.0259)	(.0260)	(.0271)	(.0294)	(.0294)
COLL79	.3319	.3221	.2677	.2667	.2975	.2450	.2398
	(.0237)	(.0238)	(.0271)	(.0272)	(.0265)	(.0294)	(.0296)
ADV79	.4734	.4658	.3995	.3997	.4378	.3711	.3665
	(.0347)	(.0341)	(.0369)	(.0366)	(.0383)	(.0407)	(.0480)
				Students in nonacademic programs			
VOC79	.0388	.0379	.0293	.0306	.0302	.0243	.0254
	(.0149)	(.0146)	(.0144)	(.0144)	(.0159)	(.0157)	(.0156)
SOC1479	.0681	.0495	.0294	.0310	.0414	.0210	.0235
	(.0135)	(.0133)	(.0139)	(.0139)	(.0146)	(.0155)	(.0156)
SOC1579	.1610	.1473	.1187	.1165	.1310	.1004	.0998
	(.0159)	(.0156)	(.0167)	(.0169)	(.0172)	(.0184)	(.0186)
COLL79	.2889	.2844	.2407	.2388	.2836	.2379	.2400
	(.0194)	(.0194)	(.0219)	(.0221)	(.0222)	(.0254)	(.0253)
ADV79	.4882	.4809	.4253	.4223	.3848	.3318	.3327
	(.1436)	(.1438)	(.1456)	(.1479)	(.1461)	(.1477)	(.0475)

Notes and source: See text table 2. In addition, the indicator variable VOC79 is 1 if an individual never attended college but did attend a postsecondary vocational school and 0 otherwise. SOC1479 is 1 if a person has less than two years of college (regardless of whether the student also attended vocational school) and 0 otherwise. SOC1579 is 1 if a person attended college for two or more years but did not receive a four-year degree and 0 otherwise. COLL79 is 1 if a person received a four-year degree but did not receive an advanced degree and 0 otherwise. ADV79 is 1 if a person received a graduate degree and 0 otherwise. The coefficients are estimates of difference in the log wage of a high school graduate and a person whose highest education level is in the particular category.

APPENDIX: TABLE 3

Estimates of the return to education: Instrumental variables (Dependent variable: log wage)

	Instrumental variables estimator				Instrumental variables estimator with fixed effects (Constants for each high school)		
	Basic controls	Region	Family/ achievement, region	High school, family/ achievement, region	Region	Family/ achievement, region	High school, family/ achievement, region
				Combined sample			
VOC79	.0404	.0611	.0573	.0625	.1195	.1202	.1179
	(.0617)	(.0613)	(.0614)	(.0605)	(.0641)	(.0645)	(.0642)
SOC1479	.0640	.0352	.0171	–.0177	.0353	–.0136	– .0145
	(.0293)	(.0287)	(.0289)	(.0290)	(.0296)	(.0296)	(.0297)
SOC1579	.1968	.1828	.1112	.1032	.1871	.1142	.1072
	(.0275)	(.0271)	(.0278)	(.0280)	(.0074)	(.0288)	(.0292)
COLL79	.3134	.3024	.2056	.2029	.3049	.2097	.2052
	(.0209)	(.0205)	(.0216)	(.0217)	(.0215)	(.0277)	(.0227)
ADV79	.6264	.6160	.4904	.4779	.6177	.4883	.4837
	(.0643)	(.0643)	(.0646)	(.0642)	(.0659)	(.0665)	(.0663)
				Students in academic programs			
VOC79	.0858	.1572	.0505	.0411	.2551	.2225	.2179
	(.1590)	(.1564)	(.1589)	(.1569)	(.1725)	(.1727)	(.1721)
SOC1479	.00925	–.0247	–.0731	–.0804	–.0283	–.0659	–.0779
	(.0743)	(.0716)	(.0713)	(.0714)	(.0761)	(.0755)	(.0754)
SOC1579	.1931	.1669	.1142	.1060	.1668	.1187	.1090
	(.0567)	(.0562)	(.0577)	(.0576)	(.0625)	(.0645)	(.0644)
COLL79	.2722	.2534	.1773	.1726	.2607	.1893	.1815
	(.0547)	(.0536)	(.0552)	(.0551)	(.0595)	(.0610)	(.0609)
ADV79	.5433	.5333	.4162	.3983	.5443	.4291	.4198
	(.0811)	(.0797)	(.0816)	(.0810)	(.0889)	(.0921)	(.0918)
				Students in nonacademic programs			
VOC79	.0272	.0666	.0616	.0677	.0281	.0286	.0276
	(.0644)	(.0641)	(.0634)	(.0639)	(.0698)	(.0704)	(.0701)
SOC1479	.0923	.0609	.0321	.0312	.0690	.0409	.0414
	(.0304)	(.0298)	(.0303)	(.0304)	(.0332)	(.0342)	(.0342)
SOC1579	.1491	.1490	.1102	.1027	.1373	.1006	.0966
	(.0345)	(.0345)	(.0348)	(.0355)	(.0375)	(.0384)	(.0386)
COLL79	.2877	.2909	.2324	.2283	.2921	.2377	.2383
	(.0299)	(.0300)	(.0310)	(.0313)	(.0340)	(.0364)	(.0362)
ADV79	.7301	.7546	.6786	.7012	.8399	.7895	.8112
	(.3376)	(.3425)	(.3429)	(.3485)	(.3740)	(.3731)	(.3749)

Notes and source: See text tables 2 and 3 and appendix table 2.

REFERENCES

Akin, John S., and Irwin Garfinkel, 1977, "School expenditures and the economic returns to schooling," *Journal of Human Resources*, Vol. 12, Winter, pp. 460–477.

Altonji, Joseph G., 1988, "The effects of family background and school characteristics on education and labor market outcomes," Northwestern University, unpublished paper, December.

Angrist, J., and A. Krueger, 1991, "Does compulsory schooling attendance affect schooling and earnings?," *Quarterly Journal of Economics*, Vol. 106, No. 4, November, pp. 979–1014.

Ashenfelter, Orley, and Alan Krueger, 1991, "Estimates of the economic return to schooling from a new sample of twins," National Bureau of Economic Research, working paper, No. 4143.

Ashenfelter, Orley, and Cecilia Rouse, 1997, "Income, schooling, and ability: Evidence from a new sample of identical twins," National Bureau of Economic Research, working paper, No. 6106, July 1.

Card, David, and Krueger, Alan, 1992, "Does school quality matter? Returns to education and the characteristics of public schools in the United States," *Journal of Political Economy*, Vol. 100, No 1, February, pp. 1–40.

Griliches, Zvi, 1979, "Sibling models and data in economics: Beginnings of a survey," *Journal of Political Economy*, Vol. 87, October, pp. S37–64.

__________, 1977, "Estimating the returns to schooling: Some econometric problems," *Econometrica*, Vol. 45, January, pp. 1–22.

Grubb, W. Norton, 1993, "The varied economic returns to postsecondary education," *Journal of Human Resources*, Vol. 20, No 2, Spring, pp. 363–382.

Hanushek, Eric, 1986, "The economics of schooling," *Journal of Economic Literature*, Vol. 24, No. 3, September, pp. 1141–1177.

Johnson, George, and Frank Stafford, 1973, "Social returns to quantity and quality of schooling," *Journal of Human Resources*, Vol. 8, Winter, pp. 139–155.

Kain, Thomas, and Cecilia Rouse, 1995, "Labor market returns to two- and four-year colleges," *American Economic Review*, Vol. 85, No. 3, June, pp. 600–615.

Lang, Kevin, 1993, "Ability bias, discount rate bias, and the return to education," Boston University, Department of Economics, unpublished paper, May.

Morgan, James, and Ismail Sirageldin, 1968, "A note on the quality dimension in education," *Journal of Political Economy*, September / October, pp. 1069–1077.

Murphy, K., and F. Welsh, 1992, "The structure of wages," *Quarterly Journal of Economics*, Vol. 7, February, pp. 284–326.

Olneck, Michael R., 1979, "The effects of education," in *Who Gets Ahead?: The Determinants of Economic Success in America*, C. Jencks (ed.), New York: Basic Books.

Rosenbaum, Dan T., 1997, "Ability, schooling ranks, and labor market trends: Controlling for changes in the education distribution when making comparisons between groups over time," Northwestern University, unpublished paper, November.

Siebert, W. Stanley, 1985, "Developments in the economics of human capital," in *Labour Economics*, London and New York: Longman.

U.S. Department of Education, National Center for Education Statistics, 1972–86, *National Longitudinal Survey of the High School Class of 1972.*

Willis, Robert J., 1986, "Wage determinants: A survey and reinterpretation of human capital earnings functions," in *Handbook of Labor Economics*, O. Ashenfelter and R. Layard (eds.), Amsterdam: North-Holland, pp. 525–602.