The Effect of High School Courses on Earnings

Heather Rose

Public Policy Institute of California

Julian R. Betts

Public Policy Institute of California
and
University of California, San Diego

This paper has been published as Heather Rose and Julian R. Betts, "The Effect of High School Courses on Earnings", *Review of Economics and Statistics*, (May 2004), (**86**:2), pp. 497-513.

The authors would like to thank Jeff Owings and Robert Atanda at NCES for providing us with the math and science course classifications. Participants at the annual meetings of the American Economics Association, the Association for Public Policy Analysis and Management, and at numerous Economics Department seminars, as well as

Dan Hamermesh and two referees provided useful insights. We would also like to thank the Public Policy Institute of California for research support for this project.

Abstract:

We estimate the effect that six types of high school math courses have on students' earnings nearly a decade after graduation. We use High School and Beyond transcript data to differentiate courses at a more detailed level than in previous research. This enables us to show that more-advanced courses have larger effects than less-academic courses. We also provide evidence that math courses can help close the earnings gap between students from low-income and middle-income families. Finally, by incorporating other academic subjects, we demonstrate how specific course combinations can explain the earnings premium related to an additional year of school.

JEL: J310 and J240.

1. Introduction

Education has been at the forefront of the nation's concerns for decades. Falling test scores throughout the 1960s and 1970s prompted government officials to prescribe a new curriculum. In 1983, the National Commission on Excellence in Education advised that all high school students should follow a more rigorous curriculum.[1] Since then, state policymakers invoked new graduation requirements and curriculum standards to satisfy the commission's recommendations.

Although many observers believe that an enhanced curriculum is the vehicle to improved educational outcomes, little research has been done to understand its long-term effects. Altonji (1995) marks one of the primary attempts by an economist to systematically establish a direct link between curriculum and wages. His work, which examines high school graduates from 1972, produces the puzzling result that curriculum has an extremely weak effect on wages. Studying graduates from the late 1970s and early 1980s, Levine and Zimmerman (1995) find somewhat stronger results from some of their model specifications, but conclude that any potential effects of math curriculum are restricted to certain sub-groups of the population (men with low education levels and highly educated women). Because the notion that curriculum does not matter raises serious questions about the effectiveness of the American public school system, it is essential to investigate further.

There are also more general reasons why it is important to understand the effects of high school curriculum. First, if high school curriculum has little influence over student outcomes,

---

[1] It dubbed this curriculum the "New Basics" and consisted of four years of English, three years of math, three years of science, three years of social studies, two years of foreign language (for college-bound students), and six months of computer science.

then intervention may be necessary at an earlier stage. Second, with the recent elimination of affirmative action in some states, minority access to higher education may suffer. As the returns to a college education continue to rise, such limited access would aggravate income equality among ethnic groups.

Carrying forward the literature that Altonji initiated, we estimate the effect that specific high school math courses (vocational math, pre-algebra, algebra/geometry, intermediate algebra, advanced algebra, and calculus) have on earnings nearly ten years after graduation for a cohort of students who were high sophomores in 1980.[2] We also determine whether a varied math curriculum can explain the earnings gap between students of different ethnicities, socioeconomic statuses, and genders. Our study differs dramatically from previous studies in that we use very detailed transcript data to analyze the effects of specific math courses rather than just the total number of math courses. Our principal data source is High School and Beyond (HSB).

The paper proceeds as follows. Section 2 presents a theoretical discussion of the link between mathematics curriculum and wages and reviews the existing empirical literature. Section 3 describes the econometric model that we use to estimate the effects of curriculum on

---

[2] Research by Murnane, Willett, and Levy (1995) and by Grogger and Eide (1995) shows that between the 1970s and the 1980s the relative importance of math test scores in determining earnings grew substantially and that math achievement is a better predictor of adult earnings than are other types of test scores commonly available. We therefore focus mainly on math curriculum, although we broaden this analysis to include curriculum from other fields. Another practical reason to focus on math is that the content of math courses is much more comparable across schools than is the content of courses in other subjects (see Porter et. al, 1993).

earnings and provides an in-depth description of our data. Section 4 presents the results from our earnings models as well as several robustness checks. In Section 5, we investigate whether curriculum can explain ethnic, socioeconomic, and gender-based earnings gaps. Section 6 concludes.

2. How Might Curriculum Affect Earnings? A Review of Theory and Evidence

A. The Human Capital and Signaling Models

Human capital theory implies that curriculum has value because it imparts skills that make students more productive and better rewarded in the labor market. This mechanism can work in several ways. Students who take more-advanced math classes learn skills that may apply *directly* to certain jobs. They may also learn logic and reasoning skills that *indirectly* make them more productive. In addition, advanced math may also teach students *how* to learn. Finally, even if a job only requires basic math skills, a student who has taken advanced math has had an additional chance to master those skills.[3]

In contrast to the human capital model, the signaling model (Spence, 1973) suggests that math courses do not *cause* the student to be more productive. Rather, the innately more productive, i.e., "more able", students choose to obtain the specific levels of education that provide signals to employers.

In the case of curriculum, the signaling model is applicable at several levels. Students who take a more rigorous curriculum provide a signal of ability to colleges. College attendance in

[3] Gamoran (1998) mentions these paths and cites other corroborating studies as well.

turn provides a signal of ability to employers.[4] It is less clear how taking more high school courses could act as a signal for students who do not attend college. Perhaps employers of high school graduates look at high school transcripts – an assumption not generally supported by the research of Bishop (1989). Nonetheless, more able students may signal that they have taken certain courses during job interviews.

The signaling/human capital debate matters for policy. If a student gains no productivity by taking a specific math course but merely *buys* a signal of ability, then requiring all students to take that course would not raise labor market productivity or aggregate wages. Further, such a policy change could lead to inefficiencies in the labor market because a standardized curriculum would make it more difficult for employers to identify the most productive students. In contrast, human capital theory contends that additional math courses could perhaps make all workers more productive, so there is a causal relation between curriculum and wages. Given the stark difference between the implications of the two theories, our analyses involve numerous robustness checks.

B. Previous Research

[4] It is possible that employers do not use educational background as a signal of ability, but that the student possesses some characteristic, unobservable to the researcher, that causes him or her to take a more-advanced curriculum and to earn higher wages. Such a pattern would lead to endogeneity bias. This is closely related to the signaling model, because it recognizes the possibility that differences in returns from different courses could be caused by selection effects that are the result of underlying ability. We thank Deborah Reed and Kim Rueben for this insight.

The economics literature has been slow to incorporate high school curriculum into wage models, with the exceptions of Altonji (1995) and Levine and Zimmerman (1995).[5] Altonji asks: Does an extra year of education serve merely as a screening device or do the courses that make up that year possess some intrinsic value? Using the National Longitudinal Survey of the High School Class of 1972 (NLS72), he models the log-wage of each person as a function of credits completed in eight different subjects during grades 10-12, standard background variables, and years of postsecondary education.[6] Because the eight curriculum variables are highly correlated, he conducts the same analysis using combinations of courses by subject instead of entering the individual courses separately.

He uses three methods to estimate the effects of the curriculum: OLS, OLS with high school fixed effects, and a model in which he uses a school's average number of credits earned per student within each subject as an instrument for each student's number of credits earned in that subject. All three approaches lead to similar results. Altonji's overall conclusion is that "the

[5] Gamoran (1998) provides an excellent review of other studies that have undertaken similar goals. Most of these studies are quite dated and not in the economics literature. Many focus on the effects of tracking rather than specific high school courses. Others that do look at courses restrict their samples to students who obtain no postsecondary education. In unpublished work, Ackerman (2000) also addresses the issue of math curriculum but does not divide up courses in as detailed a manner as we do.

[6] The eight subjects are science, math, English, social studies, foreign language, industrial arts, commercial courses, and fine arts. One credit refers to an additional year's worth of the course.

effect of a year equivalent of courses is much smaller than the value of one year in high school." In other words, the whole is greater than the sum of its parts. Even before controlling for background characteristics, the IV estimates suggest that each additional year of science, math, English, social studies, and foreign language combined leads to a miniscule 0.3 percent increase in wages. He finds stronger curriculum effects if he excludes the negative effects of English and social studies. An additional year of math, science, and foreign language increases earnings by 3.1 percent.[7] Because an additional year of school is estimated to increase wages by 7 percent, Altonji's results lend support to the view that high school serves as a screening device rather than as a mechanism for human capital formation.

On the other hand, Altonji's results could be not so much a refutation of human capital theory as a sign that schooling increases human capital in ways quite distinct from curriculum. For instance, schooling may improve student's critical thinking and punctuality. See Bowles, Gintis and Osborne (2001) for a discussion of these ideas.

Levine and Zimmerman focus on the effect that math and science courses have on wages. They use data from two main sources: the National Longitudinal Survey of Youth (NLSY) and HSB's 1980 senior cohort. Levine and Zimmerman focus on students who graduated in the late 1970s through the early 1980s and estimate separate models for men and women. Like Altonji, they use the number of credits earned in math and science courses (separately) as their curriculum measures in the HSB data.

---

[7] OLS estimates are slightly larger, and OLS with high school fixed effects are substantially larger. OLS estimates without fixed effects predict that the effect of an additional year of mathematics on earnings is 1.8 percent, but that disappears once ability controls are added.

Levine and Zimmerman find that the number of science classes has very little effect on wages for either males or females. However, their OLS results indicate that an additional semester of math increases male wages on average by 3 percent and female wages on average by about 2 percent. Further, they find that the math effects are limited to men who have only a high school degree and to women who have completed some college or have earned a college degree. For female college graduates, an additional semester of math during high school leads to a 5.4 percent increase in log wages. The effect for men with only a high school diploma is 3.1 percent. When Levine and Zimmerman use instrumental variables with Altonji's instrument, the math effects disappear. The variation in results across subgroups may be the result of the small sample sizes within each group. Furthermore, they are looking at wages only about six years after high school graduation, so males may not have settled into careers indicative of their curriculum and educational attainment.

C. Contributions of this Paper

A key factor that distinguishes our work from the two earlier contributions is our detailed analysis of the *types* of math courses taken. A second distinguishing factor is our focus on the role, if any, that high school curriculum plays in creating the well-known wage gaps between workers of different races, ethnicities, and genders.

Like Levine and Zimmerman, we use the HSB dataset. Unlike Levine and Zimmerman, we use the sophomore cohort of the HSB dataset, most of whose members graduated from high school in 1982, rather than the senior cohort that graduated in 1980. This alternative sample provides several advantages. First, our data reflect earnings ten years after graduation rather than only six years, as is the case with Levine and Zimmerman's data. The effects of curriculum on earnings could look quite different for workers in their late twenties than for a sample of 24-yearolds who may not yet have settled into careers. Even students who obtained ample postsecondary education have relevant earnings data in our sample. Second, the transcript data for the 1982 HSB seniors are much more detailed than is the course information for the 1980 HSB seniors. Third, both Levine and Zimmerman and Altonji study a cohort of seniors, thus excluding high school dropouts. Because we begin with a tenth grade cohort, we are able to include some dropouts in our models. Fourth, whereas Altonji examines earnings in the 1970s and in 1986, and Levine and Zimmerman examine earnings in 1986, we follow the students into the early 1990s. This update may be important given the dramatic increase in the returns to education in the United States between the late 1970s and the mid-1990s.

3. Earnings Model and Data

A. Estimating the Effects of Curriculum on Earnings

We construct the following linear model of the log of 1991 annual earnings for student *i* at school *s:*

$$\ln earn_{is} = \alpha + \beta_0 Curric_{is} + \beta_1 Demo_{is} + \beta_2 Fam_{is} + \beta_3 Sch_{is} + \beta_4 HiDeg_{is} + \varepsilon_{is} \qquad (1)$$

where $Curric_{is}$ denotes curriculum (a vector of the credits earned in each of six math course categories); $Demo_{is}$ refers to demographic information; Fam_{is} and Sch_{is} are family and school characteristics, respectively; $HiDeg_{is}$ represents a series of dummy variables indicating the highest degree the student has earned by 1992; and ε_{is} is an i.i.d. error term.[8] The specific

[8] In our initial analyses, we assume the error term is independent across students. However, because some shocks may affect all students at a particular school in the same way, we also estimated random effects models in which the error term also contains a school specific

regressors included are listed in Table 1. Each element in the vector of coefficients, β_0, describes the effect of an additional credit in the corresponding math course on the log of earnings. Because we include the school variables upon which the HSB survey was stratified, we do not weight the regressions.

Insofar as educational attainment could itself be an endogenous function of high-school curriculum, we also estimate reduced form models that exclude educational attainment. We elaborate on the specific models we estimate after describing the data. In Section 4, we discuss the issue of omitted ability extensively and present several alternative specifications, including an instrumental variables approach similar to that used by Altonji (1995) and a model that includes school fixed effects.

B. Data Description

The principal source of data for this study is the High School and Beyond (HSB) Sophomore Cohort: 1980-92 data. This longitudinal study surveyed over 30,000 high school sophomores in 1980 and followed up on approximately 15,000 of them in 1982, 1984, 1986, and 1992. This is an excellent source of data for several reasons. It provides extremely detailed high school transcript information, including every course taken by the student, the term it was taken, the

---

component. The estimated curriculum coefficients and standard errors were nearly identical to those estimated by OLS (the coefficients differed by 0.0002 at most and the standard errors by even less, depending on the specific model specification), so we report the OLS estimates for simplicity.

grade received, and the number of credits earned. It also provides a wealth of personal and family characteristics and includes high school dropouts in the transcript and follow-up surveys.[9]

We use the log of annual earnings in 1991 as our primary dependent variable. There are two primary shortcomings of these data. The earnings data are missing for approximately 20 percent of the public school sample, thus reducing the number of usable observations. These data are missing primarily because of the lack of participation in the final follow-up. Secondly, the earnings data measure annual earnings rather than an hourly wage, which is a much better measure of actual productivity.[10] Thus any apparent curriculum effect may operate through two channels: an effect on wages and an effect on employment status and hours worked. It is impossible to disentangle these two effects entirely. We therefore restrict the range of earnings that we model, eliminating those who earned less than $2,000 (to exclude those most likely working part-time) and those few who earned more than $75,000.[11]

We constructed data on mathematics curriculum from the restricted high school transcript data. In this dataset, every high school math course a student took is classified into one of 42 categories using the standard Classification of Secondary School Courses (CSSC). We

---

[9] Respondents who missed a year were still included in subsequent follow-ups if possible. Even students who were selected into the base-year survey but missed it were included in the follow-ups if possible.

[10] The survey did gather extremely detailed wage data until 1986 but stopped after that.

[11] Grogger (1996) and Grogger and Eide (1995) make similar data restrictions. In a subsequent section, we discuss how the results change when we relax this income restriction and when we use a version of monthly earnings as the dependent variable.

aggregated these 42 classes into 6 broader categories based on a classification system provided by the National Center for Education Statistics (NCES). In increasing level of rigor these are: vocational math, pre-algebra, algebra/geometry, intermediate algebra, advanced algebra, and calculus. Table A.1 describes the specific math courses included in each category.[12] The number of credits a student earned in each class is also available, where a typical one-year course is assigned one credit and a half-year course is assigned 0.5 credits (technically, these credits are Carnegie units). Combining these two variables yields our primary measure of curriculum: the number of credits earned by student *i* in each of the 6 math course categories.[13]

Students for whom earnings data or curriculum data are missing are excluded from this analysis. We also restrict our sample to students who attended public schools and exclude students who transferred schools during high school. In addition, we exclude students who were enrolled in postsecondary education at any time during the 1991 year or for whom enrollment

---

[12] Although we describe math here for simplicity, we do analyze other academic subjects. Table A.2 shows the classification system that we use for science courses.

[13] In the unrestricted version of the data, only the total number of math classes that a student took is available. Thus, no measure of course difficulty is available. Also problematic is that the pre-calculated course counts in the transcript data give each course taken a count of one. So, if one student takes a one-year algebra course and another student takes two one-semester algebra courses, they will have course counts of 1 and 2, respectively. In essence, the two students have taken the same course, but their tally is misleading. This could lead to measurement error bias in models that use either the unrestricted version of the data (or the pre-calculated course counts in the restricted version), which would bias the estimated effect of math toward zero.

data are missing, because their earnings may not truly reflect their human capital formation, or their final signal.[14] Tables A.3 and A.4 list the number of missing observations for the primary variables used in the analysis.

We treated questionable data values as missing (for example, student-teacher ratios of 0.17 and of 2000). For variables other than the dependent variable and the curriculum measures, we set missing values equal to zero and included a dummy variable indicating whether the variable was missing. In approximately 25 cases, we imputed values for earnings or considered them to missing when the reported data seemed implausible. For example, we assumed incomes that jumped from $20,000 in 1990 to $200,000 in 1991 and then back to $20,000 in 1992 were data entry errors and we corrected them appropriately (in the above case, set to $20,000). We did not change large jumps in earnings that coincided with the completion of a bachelor's degree.[15]

---

[14] School attendance data are nonexistent for August of 1991, so in practice the restriction applies to those enrolled, or missing enrollment data, during the remaining eleven months of the year.

[15] Although the last follow-up took place in the spring of 1992, we did not use the annual earnings data from that year because they seem inaccurate. Whereas the average annual earnings steadily increase from 1982 through 1991 in an expected fashion, they fall to about half of their expected value in 1992, as if some respondents gave year-to-date earnings information. Even after discussions with the HSB personnel from the Department of Education, we could not find a clear cause. Although the 1991 earnings reports are self-reported and retrospective in nature, they are likely to be accurate because the data are gathered in 1992 (and near tax time for 1991 income).

C. Descriptive Statistics

Because HSB used a stratified national probability sample of schools in which schools with a high percentage of Hispanic students were oversampled, the summary statistics must be weighted to make meaningful projections to the population as a whole.[16] In Table A.5, we present both weighted and unweighted means and standard deviations of the primary variables used in our analysis. We present these descriptive statistics for the 11,724 students in public schools and, because some crucial data are missing, for the subsample of observations used to estimate the earnings models (the regression sample). The means and standard deviations in the regression sample are strikingly similar to those obtained when using the full set of potential public school observations.[17] This similarity offers some assurance that sample attrition and missing values have not distorted our sample.

---

[16] Although that type of school was oversampled, within the school 36 students were randomly selected. The ethnic composition of the oversampled schools still leads to a higher than nationally representative proportion of Hispanics in the sample. Our regressions do not use weights, opting instead to include controls for the variables used to stratify schools in the HSB sample.

[17] We examined mean differences for *all* regressors, including those not shown in Table A.5. The biggest mean difference occurs in the percentage of sample members who are male. This percentage is 4.5 points higher in the usable regression sample, indicating that we lose a disproportionate number of females. This is not surprising because, on average, more females will be out of the labor force and therefore missing earnings data in the appropriate range. As an additional test of whether the correlations between the curriculum variables and other regressors

4. Earnings Model Results: Does Math Curriculum Affect Earnings?

A. Basic Log-Earnings Models

Table 2 presents the coefficients and standard errors from the model in equation (1) as well as from two more parsimonious models. The coefficients can be interpreted as the percentage change in earnings associated with an increase of one credit, i.e., one year, for each of the specific math courses.[18] The predicted effects of taking high school math vary across models, but the final conclusion appears robust: Math matters. [19]

---

in our regression subsample are different than in the excluded data, we ran regressions of each curriculum variable on the other conditioning variables and tested for variations in coefficients between the full sample and the regression sample. We found little evidence of any changes in covariations between curriculum and other conditioning variables.

[18] These are approximate percentage changes. The regression coefficients represent a first order approximation to the proportional increase in earnings from a one-unit increase in a regressor. The exact percentage change is given by $(e^{\beta}-1)$*100%, where β is the regression coefficient.

[19] These models are unweighted regressions. Although the sample is stratified, DuMouchel and Duncan (1983) argue that the preferred estimation technique is not to weight but rather to include controls for all the variables upon which the sample was stratified, which is the course we take here. We replicated the main models in this paper using the weights we used to report sample means in Table A.5, and found similar results. One substantial change was the calculus coefficient in column 3 becomes insignificant at the 5 percent level.

All of the main models in this table disaggregate math courses by type. However, under the six disaggregated math courses, we provide a row showing the results of otherwise identical models that control only for the total number of math courses taken. While these 'average' effects are often statistically significant, they mask some fairly large variations in the impact of different types of math courses. For this reason, our discussion of Table 2 will focus on the more detailed models that disaggregate math courses by type.

Column 1 contains the results from the simplified version of the model in equation (1) that does not control for any student characteristics, summarizing the variation in mean earnings among workers with different numbers of math courses. The math coefficients are quite large and vary by the level of the course. An additional year of calculus is predicted to increase earnings by approximately 19.5 percent, whereas an additional year of algebra/geometry is predicted to increase earnings by about 8.0 percent. Vocational math courses, however, seem to have almost no effect on earnings.

Obviously, this first model is simplistic because it does not take account of many other observable variables that are known to affect wages. As column 2 shows, adding demographic, family, and school characteristics causes the effect of math courses at or above the algebra/geometry level to drop by 24 to 38 percent; the lower level math effects drop by even more. This pattern suggests that a portion of the curriculum effects from the previous model should be attributed to these other factors.[20] Nonetheless, all the curriculum coefficients, except that of pre-algebra, are still quite large and statistically significant at the 5 percent level. An additional credit in algebra/geometry is predicted to increase earnings by 6.1 percent, but

[20] See Table 1 for a list of the specific demographic, family, and school variables used.

advanced algebra is associated with an 8.8 percent gain. In this model, vocational math has a significant negative coefficient, indicating that taking additional vocational math courses leads to lower earnings. This result might seem to suggest that signaling, rather than human capital formation, is at work. While this may be, the result is consistent with either signaling or human capital theories. There are virtually no students in the sample who take no math at all. Because the model contains a constant term, each coefficient is identified by variations in the corresponding variable from the sample average. Thus, in our model, the negative sign does not imply that taking an extra vocational math course actually lowers earnings relative to a student who takes no math courses. Rather, the negative coefficient means that taking an additional vocational math course lowers earnings relative to the student who has taken the average curriculum, which includes only 0.69 of a year of vocational math courses. This could indicate that vocational math is a negative ability signal. It could also indicate that the minority of students who take one or more years of vocational math have paid the opportunity cost of taking more advanced math courses that would have helped them develop human capital needed in the labor market.

To further illuminate the path through which these curriculum effects work, we control for the ultimate educational attainment of the student. We add to the previous model a series of dummy variables indicating the student's highest educational degree attained by 1992 and present the results in column 3. With these controls, the math curriculum coefficients drop by about one-half. The signaling interpretation of this drop is that about one-half of the overall effect of high school math reflects the way in which math courses enable more productive students to attend college and therefore signal their ability to their employers. The human capital interpretation is that high school math courses increase a student's efficiency, thus increasing his or her chances of attending college, and in this way increase the student's productivity further. In the human capital interpretation, column 2 continues to show the overall effects of curriculum, whereas the results in column 3 show the effect that works directly rather than indirectly through education. The striking curriculum effects that remain in column 3 after controlling for educational attainment suggest that there may be a direct effect of math curriculum on labor market productivity that works independently of the final degree attained. In this model, the vocational math coefficient is still negative and significant, the coefficient on pre-algebra credits is no longer significant, but the high-level math coefficients remain significant.[21] A course in

[21] To determine the extent to which the negative sign on vocational math is being driven by students who take only vocational math courses and nothing higher, we re-estimated the model in column 3 but included a dummy variable indicating whether the student had taken only vocational math. The coefficient on this indicator is –0.065 and is significant at 5 percent. The magnitude of the vocational math credits coefficient becomes slightly less negative at –0.019 but is now only significant at 10 percent. This indicates that, on average, students who take only vocational math earn less than those who take vocational math and some higher math. It also indicates that even those students who take some vocational math but also take some higher math (approximately 35 percent of the students who take one vocational math course fall into this category) still earn less than the average student who does not take any vocational math. Because the average student does not take an entire credit of vocational math, students who do take one credit are taking it at the expense of a more advanced course. Thus, there is some opportunity cost to taking vocational math.

algebra/geometry is estimated to increase earnings by 3.1 percent and a calculus course appears to increase earnings by 6.5 percent.[22] [23] [24]

---

[22] The calculus coefficient is very large relative to most others in the model. One effect that is almost as large is coming from a family with an income greater than $25,000 rather than coming from a family with a mid-level income ($20,000 to $25,000). Furthermore, the calculus effect is almost great enough to offset the negative effect of coming from a family with extremely low income (less than $7,000) relative to coming from a family with mid-level income. The calculus effect also counterbalances the effect of having a mother with less than a high school degree rather than having a mother with a high school degree. In contrast, standard measures of school quality such as the student teacher ratio are not statistically significant. Not even reducing the percentage of disadvantaged students at a school by 25 percentage points outweighs the effect of taking a calculus course.

[23] To allay concerns that excluding observations with earnings not between $2,000 and $75,000 was driving our results, we estimated the column 3 model but set all nonmissing earnings values less than $2,000 equal to $2,000 (about 80 percent of these earnings were $0). This imputed earnings value can be thought of as earnings that individuals forgo by remaining at home, or it may represent a low-value of the household work for those not working. In this case, the results are also quite similar to those in of column 3 in Table 2. As a second solution to elicit productivity rather than variations related to labor force attachment, we modeled the log of monthly earnings (calculated as annual earnings in 1991 divided by the number of months that the respondent was employed during that year). At each stage, the resulting math coefficients were very close to those from the original model using annual earnings. Because using the

Another mechanism through which curriculum could affect earnings is by channeling students into majors, or by keeping the door open to majors and occupations that are more highly rewarded in the labor market. We estimated, but do not show, models that also controls for the student's college major and occupation (with 9 and 13 categories, respectively). The purpose of these models is to understand the channels through which high school curriculum increases wages. As expected, in models that control for students' college major, the curriculum coefficients all fall, but not enough to indicate that the bulk of the remaining curriculum effect operate though the college-major channel. Even with these additional controls, many of the math effects are still statistically significant.[25] Separate models that controlled for students'

---

monthly earnings measure meant losing 150 more observations resulting from missing employment data, we chose to use annual earnings throughout this paper.

[24] Another important issue is whether the returns to taking math courses are non-linear. We re-estimated model 3 after adding the squares of the number of credits taken in each math subject. None of the squared terms were close to being significant. Although we suspect that diminishing returns should set in at some point, the relatively small number of courses taken in each subject, as shown in Appendix Table A.5, prevents us from determining at what point such non-linearities begin.

[25] The algebra/geometry coefficient is .027, whereas that of advanced algebra is larger at .034. Calculus still has the largest effect at 5.7 percent, but it is only statistically significant at the 10 percent level. We thought that there might be important interactions between educational attainment and major. However, the curriculum effects did not change substantially when we added these interactions terms. Indeed, none of the interactions of curriculum and highest degree

occupation were quite similar. Vocational math, algebra/geometry and advanced algebra remain significant at 5 percent while calculus and intermediate algebra become insignificant. [26]

To understand better the degree to which educational attainment, college major and occupation can account for the observed impact of math courses on earnings, we calculated the predicted impact of taking an average number of courses in each category (from the final column of Appendix Table A.5) before and after accounting for educational attainment (Table 2, models 2 and 3, respectively), and in a version of model 3 that also added controls for college major and occupation. We found that 64 percent of the impact of math courses appears to work through educational attainment, while an additional 18 percent works through students' subsequent occupation and college major. This leaves 18 percent of the curriculum effect that apparently works through other mechanisms.

In sum, it appears that the effects of curriculum operate much more through educational attainment than through the choice of major or occupation. The results from this section indicate that, even after accounting for a multitude of factors, a more rigorous curriculum is associated

---

dummies were statistically significant, suggesting that the impact of taking additional courses is widespread.

[26] The coefficients on the math variables are also slightly smaller once we control for occupation, at -0.02 (vocational math), 0.005 (pre-algebra), 0.025 (algebra/geometry), 0.026 (intermediate algebra), 0.038 (advanced algebra) and 0.050 (calculus). We also tried models that simultaneously controlled for occupation and college major. The general pattern was a further slight reduction in the math coefficients, with vocational math, algebra/geometry and advanced algebra remaining significant at 5 percent.

with higher earnings, indicating that math curriculum may directly affect labor market productivity.

B. Omitted Ability

Two factors that are impossible to include fully in any school-earnings study are ability and motivation. Theory suggests that these traits are positively related both to students' level of education and to their subsequent wages.[27] Thus, if these characteristics are omitted from an earnings model, the coefficients on the schooling variables (in our case, the curriculum variables) will be biased upward to the extent that these characteristics are positively correlated with curriculum and earnings. This could be a particularly large problem when we count the number of courses of a specific type, rather than when we aggregate all math courses into one category.

We adopt two main strategies to deal with this issue. Our main strategy is to add ability and motivation controls in the form of the student's mathematics grade point average (GPA). We also use an instrumental variables approach similar to that used by Altonji (1995) to eliminate the part of curriculum that is related to the student's own ability and motivation.

i) Controlling for Ability and Motivation

The student's math grade point average provides a potentially good measure of ability and motivation because it represents how well students understand and apply themselves given a particular curriculum. But GPA may in part measure factors quite distinct from human capital, such as punctuality, neatness, and the level of involvement by parents in the student's

[27] In theory, a negative relation could arise between ability and education if more able students found it optimal to leave school earlier because of the high opportunity costs of schooling in the form of forgone earnings. See Griliches (1977).

schoolwork. Considering the difficulty in distinguishing one effect from the other, we refer to the ensemble of ability and motivation as "ability."

The model with math GPA appears in column 4 of Table 2. The algebra/geometry coefficient is still significant at the 5 percent level and approximately as large as before, with a predicted effect of 2.9 percent. Similarly, the advanced algebra coefficient remains significant at the 5 percent level, but does drop somewhat in magnitude. The calculus coefficient is no longer significant with math GPA in the model. This latter result is certainly consistent with the idea that some of the most advanced courses may reflect ability in part rather than the creation of human capital.[28]

[28] We experimented with other potential ability controls and the results were fairly consistent regardless of the control we used. We hoped to include a pre-high school test score to control for pre-high school math ability (and more generally prior achievement). Unfortunately, the earliest test score data available in HSB is from a series of tests administered during spring semester of the student's sophomore year. Because this score is likely to be affected by the courses students take during grades 9 and 10, it does not provide an adequate measure of pre-high school math aptitude. It may actually "overcontrol" for ability, causing a downward bias on the curriculum coefficients. Nonetheless, we did estimate a model that included the student's mathematics test score and the results are comparable to the case with GPA controls. In this model, the algebra/geometry coefficient is 0.025 and significant at the 5 percent level; the advanced algebra coefficient is slightly smaller at 0.027 and significant at the 10 percent level. In contrast to the model with GPA, controlling for test score yields a calculus coefficient that is significant at the 10 percent level and slightly larger in magnitude at 0.055. The math coefficients may change

We also re-estimated model 4 after adding controls for college major and occupation, both separately and together. The results mirror our discussion of these extensions to model 3: there is evidence that college major and occupation explain some but not all of the impact of high school math courses on earnings. The most striking changes came when we added controls for both major and occupation, in which case vocational math remained significant at 5 percent, algebra/geometry was significant at almost 5 percent and intermediate algebra was significant at about the 10 percent level. In a model not shown, where we repeat model 4 but without educational attainment, we find that the estimated impact of math courses rises substantially, similarly to what we show in model 2.

We re-calculated our earlier decomposition of the sources of the math effects, but this time using the models that condition on math GPA. The results suggest that 68 percent of the effect works through students' educational attainment and an additional 19 percent works through

---

across model specifications in part because of the falling sample sizes that result from missing data in some of the controls (13 percent of the regression sample is missing test score data). As an additional way of accounting for student motivation and parental influence, we controlled for a set of attitudinal variables indicating the academic inclination of students and parents (such as whether the parents closely monitor the student's schoolwork, whether the parents know where their children are at all times, whether the student intends to go to college, the amount of television the student watches and how much time the student spends reading outside of class). Adding these variables, rather than GPA, causes even smaller changes in the math coefficients from the case of no ability controls.

students' college major and occupation. The unexplained portion, about 13 percent, appears to work through other mechanisms.

ii) Instrumental Variables Estimation

In another attempt to curb omitted ability bias, we follow Altonji's lead and use a school's average math credits earned in each of the six math categories as instruments for the student's own math credits earned in those categories. The intuition is that we want to purge the portion of the curriculum effect that is related to ability. We use the *school's average curriculum* to predict the student's actual curriculum, and any deviation of the student's actual curriculum from the predicted level is assumed to be caused by variations in ability, thus leaving the predicted value independent of ability. Therefore, if we use this predicted level of curriculum in our model instead of the actual level, we will be estimating the effect of pure curriculum rather than the effect of a mix of curriculum, ability, and motivation. However, note that if parents of highly motivated students all flock to the same school, the instrument will not fully eliminate ability bias. (We control for this possibility, but rather imperfectly, by including dummies for census region and for suburban and rural schools.) We return to this issue after discussing the IV results.

We estimate our earnings model using two stage least squares and present the results in columns 5 and 6 of Table 2.[29] Column 5 shows the IV results with no additional ability controls

---

[29] We exclude the student's own curriculum when calculating the school average curriculum for that student. In our second stage regression, we exclude students who come from schools where the school average curriculum was calculated using fewer than four observations, causing us to exclude about 4 percent of the students. The results change only minimally if we change

and column 6 includes GPA. The results are strikingly similar in both IV specifications.[30] Vocational math credits have a statistically significant negative effect on earnings of around 8 percent. Credits earned in the algebra/geometry category are significant at the 5 percent level in both specifications and are of similar magnitude at approximately 8 to 9 percent. This is a rather large increase from the OLS estimates. It appears that the effect of higher-level math courses has been condensed into the algebra/geometry category. It is important to stress that although the higher-level math coefficients now have negative signs, they are not significantly different from zero.[31]

iii) High School Fixed Effects

As another robustness check, we use OLS to estimate an earnings model with high school fixed effects both with and without GPA (see columns 7 and 8). Whereas the IV estimates should net out ability effects within each school, the fixed effects estimates should control for

---

the requirement for the number of observations for computing the average from each school. The number of students per school ranges from 1 to 36, with a median of about 12. We tried an alternative specification where we estimated the first-stage regressions using the largest possible sample of students, including those students whom we excluded from the second-stage regression because of missing earnings data. However, the results changed only minimally.

[30] In fact, the following results also hold in other IV model specifications that include controls for math test score and models that include both math GPA and math test score.

[31] Given that few students took the high-level math courses, the quality of the instrument may be reduced for this level of course, therefore explaining the lack of precision in estimating these coefficients.

variations in abilities across schools. Without GPA in the model, half of the curriculum coefficients are higher than the comparable model in column 3. The pre-algebra and algebra/geometry coefficients remain practically unchanged and the coefficient on vocational math is no longer significant. The results with math GPA added in addition to the fixed effects are similar to, but somewhat stronger than, the corresponding model in column 4 with GPA but without high school fixed effects. [32]

C. Additional High School Subjects

Although our research focuses on the effects of mathematics curriculum, we also incorporate courses in English, science, and foreign language into our model using detailed curriculum categories.[33] We classified the number of English credits earned into four levels: below grade

---

[32] F-tests that the high school dummies are jointly equal to zero produced p-values of 0.0001 without GPA in the model and 0.00014 with GPA in the model. However, our other models that condition upon high school characteristics appear to capture much of this variation, with p-values for the F-tests of 0.02 regardless of whether we include GPA. The school characteristics that were significant at the 5 percent level were many of the dummies indicating the stratifications used to sample HSB schools, and region dummies. None of the measures of school resources were significant, while the percentage of students who were disadvantaged was the demographic characteristic closest to being significant at the 5 percent level.

[33] These curriculum measures display collinearity with math. The number of math credits earned has a correlation of 0.54, 0.36, and 0.38 with science, English, and foreign language credits, respectively. We also estimated models that included social science credits, but the main results were not altered. We decided to leave it out to reduce collinearity problems.

level, average grade level, above grade level, and English literature courses. We measured science curriculum as the number of credits earned in six science course categories (see Table A.2 for the classification system). We classified foreign language curriculum into two variables. One designates whether the student took one or two courses whereas the other signals that the student took three or four courses.

We estimate models containing all four subjects using both OLS and IV methods and present the results in Table 3. For comparative purposes, we include models that do and do not condition on math GPA. However, we focus on the models in columns 2 and 4 that condition on math GPA. We note that the other subjects may also serve as ability/motivation controls. All models include controls for demographic, family, and school characteristics as well as the highest educational degree attained by the student.

Including the credits earned in other subjects (column 2) causes the math coefficients to drop by approximately 30 to 40 percent from the base level case, depending on the math course (except for the case of vocational math). It appears that taking an average or above-level English course is associated with increases in earnings roughly similar to the predicted gains from taking a math class at the algebra/geometry level or higher. The above-level English credits are predicted to have a larger effect on earnings than are average-level English credits. None of the science coefficients are statistically significant except for primary physics, which is predicted to have a negative effect on earnings (most likely for the same reasons that vocational math does). Taking three or four foreign language courses also has a significant positive effect. At 5.6 percent, its coefficient seems relatively large compared to those of other subjects. However it represents the effect of *three to four* credits whereas the predicted effects of the other subjects represent the effect of *one* additional credit.

We repeat the analysis using the IV technique. The results from adding the additional curriculum measures are remarkably similar to the IV results in Table 2 column 6, in that the sole math variable that is significant at 5 percent is algebra/geometry. The IV models are less supportive of the notion that foreign language and science are associated with higher earnings than are the OLS models.

Our results indicate that mathematics courses have a large effect on earnings, regardless of whether we also control for other types of courses taken. In fact, the IV estimates imply that the returns to taking a one-unit algebra/geometry course are statistically significant and large in magnitude – over 9 percent. This is higher than the average returns to an additional year of schooling (often cited as 7 percent). Other math courses, as in the simpler IV model, become insignificant.

In contrast to the OLS estimated model that contains all four curriculum measures, the effect of the above-average English credits in the IV estimated model is tripled. Perhaps accumulating credits in foreign language is a sign of ability or motivation, which the IV method eliminates. Finally, in the IV estimated model, the low-level science courses are still predicted to have negative, and now even larger, weakly significant effects.

Unlike some of the earlier literature, we find that the sum of the parts (i.e., the effect of high school courses) can be as large as the whole (i.e., the effect of an additional year of high school, often cited as a 7 percent increase in earnings) contingent upon the student taking the *right* courses. In the case of students who take the most demanding courses, curriculum predicts earnings gains of more than 7 percent per year. This can be seen best by considering some simple thought experiments.

To compare our estimated effects of a year's worth of curriculum to the 7 percent effect of a year's worth of schooling, Table 4 shows OLS estimates of potential combinations of high school courses. The first row shows that students who drop out in tenth grade experience a 12 percent earnings deficit compared to those who stay in school (calculated as the effect of having less than a high school degree in the Table 2, column 4 model). This resonates with the estimated 7 percent benefit per year that school provides. The remaining rows in the table compare the earnings gain for a student who takes the stated set of courses, relative to an otherwise identical student who does not take the courses. In each case, the set of courses is meant to approximate what a typical student might take by staying in school one year longer. For students who do not drop out, we present three hypothetical course loads (low, medium, and highly academic combinations) that they could take during their eleventh and twelfth grades. The returns to curriculum depend critically on the type of courses taken. A low-level curriculum taken in either year has a predicted effect on earnings of about 2 percent, compared to a 5-6 percent return for a medium-level curriculum. An additional year of a high-level curriculum carries a predicted earnings premium closer to 7 to 9 percent.[34]

[34] Throughout this exercise, we assume that all other background characteristics are held constant and that the only difference between the students we are comparing is their curriculum. In other words, the effects within each academic year (e.g., twelfth grade) are measured relative to a hypothetical student who stays in school that academic year but does not take any math, English, science, or foreign language courses. The more interesting comparisons are between students who take a high-level curriculum rather than a low-level curriculum.

We conclude that on average, the returns to one year of high school can be largely explained by the courses, especially in math and English, that students take.

D. Comparison to Previous Research

So far, our unique contribution to this branch of literature is to classify the number of math courses that students take based on the academic level of the course. To highlight the difference that classifying courses by their type makes, we re-estimate the previous set of models using the aggregate credits earned in a particular subject rather than the detailed credit counts.[35] We also present a specification using aggregate credits that matches Altonji's (1995) specification as closely as possible. We present the outcomes from the aggregate models in Table 5.

Notably, the results from our models with aggregate course counts in the four subjects closely approximate the results in Altonji (1995).[36] Thus our method of classifying the

[35] Rather than relying on the pre-calculated course counts provided in HSB, we calculated the total number of credits earned directly from the transcript data.

[36] This applies to models estimated by both OLS and IV methods. To further approximate Altonji's results, we estimated these models but excluded high-school dropouts. The results changed minimally, with the predicted effects differing by 0.002 at the most. We had hypothesized that our results differed from Altonji's because we used a more recent cohort and included high school dropouts in our analysis. Now it seems clear that the differences stem mostly from our more detailed classification of curriculum.

curriculum by level seems to explain some of the "curriculum puzzle."[37] We also find that some of our other hypotheses concerning Altonji's results are incorrect or only partially correct.

Columns 1 and 2 of Table 5 repeat the models from columns 1 and 2 of Table 3 but use aggregate credits instead of detailed course counts. Both models control for the demographic, family, school, and educational attainment characteristics that have been used throughout this paper. In model 1, the math coefficient is only marginally significant, whereas in this model's Table 3 counterpart, several of the disaggregated math coefficients are significant at the 5 percent level. Model 2 in Table 5 adds GPA to model 1. The aggregate math coefficient is no longer significant. Total English credits become marginally significant.

Column 3 takes the first step towards approximating Altonji's Table 2 model 7. The main changes are that we remove the school resources that are not included in Altonji's model, and we also exclude high school dropouts as Altonji does.[38] Neither of these differences between Altonji's model and our own seem to matter much, whether these changes are made jointly, as shown in the table, or individually.

At this point, the major difference between Altonji's model and our replication is that we use math GPA as a control for ability, whereas Altonji uses test scores. Column 4 of Table 5 takes the final step towards replicating Altonji's model. It makes the same restrictions as the column 3 model, but it includes senior year test scores instead of GPA. In this replication, aggregate math

---

[37] This term was coined by Altonji when describing the small estimated effects of curriculum.

[38] See the notes to Table 5 for more details of how columns 3 and 4 differ from columns 1 and 2 in a bid to mimic the specification used by Altonji as closely as possible.

credits are not significant. English and foreign language credits are significant but the effects are modest. The main difference between these findings and those of Altonji is that English credits in our model are significant. Altonji reports that one year of math, science and foreign language is associated with a gain in earnings of about 1.5 percent. Our Table 5 column 4 gives an estimate of 1.9 percent. But when we add in a year of English our estimate jumps to 3.6 percent compared to 0.6 percent in Altonji. Still, overall the results seem highly similar to Altonji's in that we do not come close to "explaining" a 7 percent increase in earnings from attending one more year of school.

Disaggregation of courses taken is the most important innovation we make, but the use of GPA rather than test scores is a secondary reason for why our results are somewhat more optimistic than Altonji's. Although the math variable is still not significant in either models 3 or 4, it does fall by half when we include test scores rather than GPA. Overall, we interpret this as a sign that our decision not to use high school test scores as a control, because they are endogenous, makes a substantive difference.

It is also notable that the timeframes differ significantly between our samples and Altonji's -- our earnings observations are from 1991 compared with 1977 through 1986 in Altonji (1995). Because the returns to education increased markedly between the late 1970's and the early 1990's, we had speculated that perhaps math curriculum matters more today than it did in the past. Our results do not provide strong evidence that this is the case, except perhaps for our finding that in our specification closest to Altonji's, English courses matter much more than he found.

The remaining columns in Table 5 repeat the same series of specifications but use the IV estimator. Perhaps the most notable finding here is that the coefficient on math courses taken is never significant and is always negative. Our earlier IV models that disaggregate suggest that in fact some math courses do indeed matter.

Overall, we conclude that the main reason for the divergence between our results and those of Altonji is our disaggregation of courses by type; a secondary reason appears to be our decision not to condition on high school test scores because they are an endogenous function of courses taken.

The finding that course type matters has many implications for curriculum reform. In particular, merely increasing the number of math courses required of students may not achieve the desired effect. It will be important to focus on the type of courses students are required to take as well.[39] In particular, our results suggest that algebra and geometry courses should be a fundamental part of any curriculum reform.

5. An Analysis of Earnings Gaps among Ethnic, Socioeconomic and Gender Groups

This section examines whether differences in high school curriculum contribute to the well-known gaps in earnings among workers of different races and ethnicities. Similarly, we test whether gaps in earnings related to a person's parental background of the person's gender in part reflect variations in high school courses.

In the early 1980s, math course completion rates varied considerably by ethnicity. Nearly 9 percent of Hispanic students and 10 percent of black students completed math credits in

---

[39] When graduation requirements are increased, there is the risk of more students dropping out. See Costrell (1994) and Betts (1998) for a theoretical analysis and Lillard (1998) for an empirical analysis.

advanced algebra or calculus compared to rates of 22 percent and 43 percent for white and Asian students, respectively.[40] The same disparities emerge when we examine the number of credits earned rather than the highest course completed. Hispanic, black, and Native American students tended to earn more credits in vocational math and fewer credits at or above the algebra/geometry level than Asian and white students.

Similarly, students from the lowest-income families (those with parents who earned less than $7,000 annually) were concentrated in the lower-level math courses, with 46 percent failing to progress beyond vocational math. For students from middle-income families (those earning $20,000 to $25,000), only 19 percent failed to advance beyond that level. Whereas 24 percent of middle-income students took courses at or above the advanced algebra level, only 8 percent of the lowest-income students did.

How much of the earnings gap between members of different ethnic groups or parental income groups can be attributed to these variations in mathematics course-taking behavior? We first estimate a model of earnings that only includes ethnicity variables as explanatory factors, with white students as the omitted group. The results appear in column 1 of Table 6. Consistent with common perception, this simple model shows that Hispanics and blacks earn significantly less than whites on average - about 5.2 percent and 10 percent less, respectively.[41]

---

[40] These figures are based on every public school observation for which we have ethnicity and math data. The same course-taking trends are evident for the weighted regression sample as well.

[41] These earnings deficits are smaller than those reported in other literature because of the HSB sampling scheme. The HSB dataset includes only those people who are still in school in

To establish a baseline earnings gap for students from different parental income levels, we carry out a similar procedure – estimating an earnings model that only contains parental income levels as explanatory factors. We measure the differences in average earnings relative to students whose parents were in the middle-income category, i.e., families that earn between \$20,000 and \$25,000 a year in 1980.[42] In this model (shown in column 2 of Table 6), students in the lowest-parental-income category (less than \$7,000) earn about 30 percent less than students from middle-income families, whereas those in the highest-parental-income group earn 10 to 11 percent more.

Next, we estimate a model that conditions on both parental income and ethnicity while taking account of demographic, family and school characteristics. The resulting earnings gaps are presented in column 3 of Table 6. The non-curriculum factors that we add to this third model can explain almost all of the ethnic earnings differentials and a large portion of the socioeconomic earnings gaps. In this more realistic model of earnings, the Hispanic and black earnings gaps disappear entirely. Asian students still experience an earnings premium relative to white students, but the effect is statistically weak. Native American students still experience an earnings deficit. Which factors are responsible for the closure in the Hispanic and black earnings gaps? Entered into the model on their own, either parental income or parental education can explain nearly all of the Hispanic gap and about half of the black gap. Together, the two

---

the second half of their sophomore year in high school. Thus, it excludes students who drop out of school at an early age, as well as immigrants who had no U.S. education.

[42] The income categories in this section are defined in 1980 dollars.

measures of parental background can account for the entire earnings gap between whites and either of these minority groups.

As column 3 of Table 6 also demonstrates, the earnings gaps related to parental- income groups diminish substantially after controlling for the remaining demographic, family, and school factors, yet they are still present. Students from the lowest-income families earn about 16 percent less than those from middle-income families and students from the highest-income families now earn about 6 percent more.

Column 3 also suggests that there is a gender gap in earnings, with males earning 27 percent more than females.

Next, we add math curriculum to the model in column 3 to see whether it can explain the ethnic, SES and gender differences. The results appear in column 4. Adding the six distinct math curriculum controls changes some of the ethnic effects. Curriculum appears to entirely explain the remaining Native American earnings deficit and the Asian earnings premium. In fact, controlling for curriculum provides weak evidence (significant at the 10 percent level) that Hispanics are predicted to earn 3.6 percent more than whites given similar curriculum and background characteristics.

Math curriculum appears to be responsible for around 27 percent of the earnings gap experienced by students from lowest-income families relative to middle-income families.[43] The

---

[43] This is computed as the percentage change in the lowest parental-income (less than $7,000) effect from the column 3 to column 4 of Table 6. Including detailed measures of English, science and foreign language curriculum in the model that controls for math narrows the gap by another 7 percent. Adding the aggregate number of math credits earned in lieu of the six

gap becomes insignificant for the second-lowest family income group ($7,000 to $15,000) after adding curriculum to the model. However, students from the two highest parental-income categories still experience the same earnings premium that they do in the model without curriculum. Thus, curriculum explains a large portion of the earnings gap between students from low-income and medium-income families, but it does not explain the gap between students from high- and medium-income families.[44] The overall picture does not change if we also control for GPA. These results are shown in columns 5 and 6. The parental income gap between the lowest and the middle-income students drops 23 percent drop rather than 27 percent. Once again, the gap between the highest parental-income and middle parental-income groups does not change from the model without curriculum to the model with curriculum.

Finally, we note that math courses taken explain none of the observed earnings gap between men and women, regardless of whether we control for each student's GPA. Some other mechanism must be at work here.

---

curriculum measures does not induce as much of a change in ethnic or parental income effect, further indicating the contribution that these new detailed measures of curriculum make to the literature.

[44] If we additionally control for the student's highest educational attainment in the column 3 and column 4 models, the earnings gap between the lowest parental-income and middle parental-income groups changes from –0.123 to –0.107 once we control for curriculum representing a 13 percent change, yet the earnings gap between middle and high-income students does not close. We chose to exclude highest degree from the displayed results to show the overall curriculum effect.

The results on parental income may carry important policy implications. Whereas many other factors help to determine the labor market success of students whose parents have average and high incomes, students of low-income families could significantly improve their earnings prospects with a better curriculum. Policies aimed at encouraging and preparing low-income students to take a more rigorous curriculum could yield significant benefits. Although we do not explicitly find that standardizing the curriculum can help narrow the ethnic earnings gap, the ethnic composition of students in the lowest parental income group was such that narrowing this gap would be a step toward narrowing the ethnic earnings gap as well: 30 percent of the students who came from the lowest-income families were Hispanic, 20 percent were black, and 34 percent were white.

6. Conclusion

The main message of this study is that math matters. The math courses that students take in high school are strongly related to students' earnings around ten years later, even after taking account of demographic, family, and school characteristics, as well as the student's highest educational degree attained, college major and occupation. Another important message is that not all math courses are equal. More advanced math courses have a larger effect on earnings than less academic courses. Our results suggest that a curriculum that includes algebra and geometry is systematically related to higher earnings for graduates a decade after graduation. As we have stressed throughout the paper, it is not entirely clear whether the link between math and earnings is causal or merely reflects unobserved variations in student ability or motivation. But it is noteworthy that math continues to matter even after controlling for proxies of ability and motivation, such as math GPA and math test scores, and after using instrumental variable methods. The IV estimates suggest that algebra/geometry credits have the largest effect on earnings, whereas the remaining courses have insignificant effects. Indeed, throughout this battery of robustness tests, credits earned in the algebra/geometry category most reliably remain a statistically significant predictor of earnings. Furthermore, both the OLS and IV results persist when we include English, science, and foreign language curriculum measures in the model, although some of these other types of courses (for example, above-level English) appear to influence earnings as well.

Can gaps in curriculum explain gaps in earnings among racial and socioeconomic groups or the gap related to gender? In answering this question, it is important to account first for variations in family background. Virtually all of the earnings gap between whites and most minority groups can be accounted for by differences in family background and, to a far lesser extent, school characteristics, without any need to control for curriculum. In contrast, the earnings gap related to variations in parental income levels or in gender cannot be fully explained by family and school characteristics. We found no evidence that math courses taken can explain the gender pay gap. But math curriculum can explain nearly one-quarter of the gap between students with parental income in the lowest and middle groups. This latter finding is important because it suggests a tool – namely the math curriculum - for increasing the degree of equity in students' earnings opportunities later in life.

Of course, it is difficult to know how in practice to implement this deceptively simple policy prescription of enriching the math curriculum. Why do low-income students fail to take more high school math if it is true that this effort would substantially boost their earnings later in life? We suspect that part of the answer is that young students have rather imperfect knowledge of the labor market returns to education, let alone to specific courses. For instance, Betts (1996) surveys undergraduates at an elite public university and finds that while on average undergraduates have a reasonably good knowledge of the returns to education in various fields, the beliefs of individual students are often massively wrong. It stands to reason that high school students from low-income families, who lack the same windows into the higher education system, have a less accurate understanding of the returns to education overall or to taking courses in individual subjects. A second reason why students from low-income families may not want to enroll in the most demanding high school courses is that they have a rather high discount rate. Indeed, the idea that young people discount their future income too highly can explain why we have compulsory attendance laws. The excessive discounting of future income is likely to be particularly common among the less affluent. For evidence that people living in poverty have higher discount rates than average, see Lawrance (1991). A third reason why students living in poverty may not take “enough” high school math courses could well be that by the time they reach high school, their accumulated educational deficit makes it difficult for them to take further courses at grade level.

Better information about the realities of the labor market, combined with higher standards for all students, might do much to solve the first two problems. The standards movement that has swept the United States during the 1990’s has already led to much clearer, and in general higher, requirements for high school graduation in most states, through the tightening of course requirements, and in many cases, the creation of high school exit exams. The third issue, that students living in poverty may be far behind grade level by the start of high school, is far more problematic. It would suggest that setting higher course requirements in high school could backfire unless at the same time education policymakers closely examined the deficiencies in student performance at lower grades and intervene early enough in students’ careers to minimize these deficiencies. Clearly, such interventions could prove costly, and we cannot speculate on the cost-benefit ratio of such interventions. Still, our results do provide some concrete evidence that the benefits of a richer curriculum are real and meaningful in size; it is the cost side of the equation that is less clear.

Our results emanate from the detailed manner in which we measure curriculum, and this is a major contribution of this study. If we simply analyze the effect of aggregate math credits earned, we find that it is an average of specific math course effects. When aggregate credits in other subjects are also included, the aggregate math effect is only marginally significant and the effects of English and science subjects disappear entirely. Furthermore, the aggregate number of math credits a student earns cannot explain as much of the earnings gap as the credits in individual courses can.

A perennial problem facing researchers working on the economics of education concerns the human capital/signaling debate. Does education, or in our case curriculum, increase earnings by increasing human capital (i.e., skills) or simply by signaling pre-existing student ability? This issue is of pivotal importance for policymakers. If the only role that math courses play is to provide more able students with an opportunity to signal this fact to the labor market by surpassing other students, then a policy requiring a more enriched curriculum for all students will not make students intrinsically more productive.

Although our data cannot provide unequivocal results on how curriculum affects earnings, our preferred earnings models that condition on math GPA suggest that 68 percent of the effect works through students' educational attainment and an additional 19 percent works through students' college major and occupation. The unexplained portion, about 13 percent, appears to work through other mechanisms.

Furthermore, our controls for student ability (math GPA and test score) suggest that signaling of ability and motivation is unlikely to be the only way in which curriculum matters. Unlike some earlier work, we find evidence that courses taken during high school can indeed explain most of the economic returns to a year of schooling. In this sense, as well, our results suggest that education does more than signal ability. As public attention continues to focus on such issues as school spending and class size, this study shows that it is crucial to remain focused on the heart of the matter: what students actually learn in school.

Appendix A: Data

Math Test Score. The math test score that we included was based on a math test in which students had 21 minutes to answer 38 questions. The questions required them to compare two quantities and determine whether one was greater, whether they were equal, or whether the relationship was indeterminable based on the given data. We used the HSB-computed Item Response Theory (IRT) scores from this test as our measure of math test score.[45]

Grade Point Average and Number of Credits Earned. We computed the student's math GPA on a scale of 0 to 4.3. We took a weighted average of the student's grade points for each course, where the weights were the number of credits that the student earned for the class. These credits were either 0.25, 0.33, 0.5, or 1, depending on whether the course length was a quarter, trimester, semester, or, more commonly, a year-long course. We converted letter grades to grade points. An "A" received 4 points, a "B" received 3 points, a "C" received 2 points, and a "D" received 1 point. We added 0.3 points for a "plus" and deducted 0.3 points for a "minus." For example, we counted a "B+" as 3.3 points.

---

[45] IRT is a "method of estimating achievement level by considering the pattern of right, wrong, and omitted responses on all items administered to an individual student. Rather than merely counting right and wrong responses, the IRT procedure also considers characteristics of each of the test items, such as their difficulty and the likelihood that they could be guessed correctly by low-ability individuals. IRT scores are less likely than simple number-right formula scores to be distorted by correct guesses on difficult items if a student's response vector also contains incorrect answers to easier questions." See Ingels et al. (1995, p M-4).

If a student failed a course, but subsequently retook it and passed it, we included the letter grade and credit information from the successful completion of the course in the student's GPA, but we ignored the information from the failed attempt. We believed that it is the *final* level of success in a given course that is likely to have the greater effect on a student's subsequent educational attainment and earnings. If, however, the student failed a course and did not repeat it successfully, we included the credit information in the student's GPA calculation but assigned 0 grade points to the course. In other words, we gave the student credit for having taken the course, but lowered the GPA accordingly.

The credit values that we used to compute the GPA also served as our measure of math curriculum. So, for example, a student who took a one-year calculus course had a tally of 1 credit for that course category. A student who took three semesters worth of a geometry course had a tally of 1.5 credits for that category. In a few cases, students received a grade of "pass" in a course. We did not include these grades in the GPA calculations but we did add the number of credits earned to the tally of credits earned by the student.

We estimated an alternative specification where we included only the courses that the student passed in the GPA calculation and in the tally of credits that the student earned. The results from our main models did not change when we used this alternative measure.

Table 1 – Regressors Included in Main Models

Math Curriculum	=	Math credits earned in each of the following six math categories: vocational math, pre-algebra, algebra/geometry, intermediate algebra, advanced algebra, and calculus.
Demographic Information	=	Ethnicity, gender, age in 1991, and marital status in 1991.
Family Characteristics	=	Parental income, parental education, parental nativity, and the number of siblings.
School Characteristics	=	Student-teacher ratio, books per pupil, length of the school year, school enrollment, percentage of disadvantaged students, percentage of teachers with a master's degree, district's average spending per pupil, teacher salary, whether teachers are unionized, and the public school type (regular, alternative, Cuban Hispanic, or other Hispanic), geographic region (nine U.S. regions), and urbanicity (rural, urban, or suburban).
Highest Degree	=	High school dropout, high school diploma, some postsecondary education (but no degree), a certificate, an associate's degree, and a bachelor's degree or higher.

Note: Specific categories and means of the above variables are available from the authors on request.

Table 2
The Effects of Specific Math Courses on Log-Earnings:
OLS, IV and Fixed-Effects Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Vocational	0.001	-0.024 **	-0.027 **	-0.029 **	-0.084 **	-0.086 **	-0.019	-0.023 *
	(0.011)	(0.011)	(0.010)	(0.011)	(0.030)	(0.030)	(0.012)	(0.013)
Pre-Algebra	0.067 **	0.023	0.007	0.005	0.015	0.013	0.006	0.004
	(0.014)	(0.014)	(0.014)	(0.014)	(0.034)	(0.034)	(0.017)	(0.017)
Algebra/Geometry	0.080 **	0.061 **	0.031 **	0.029 **	0.090 **	0.083 **	0.029 **	0.027 **
	(0.010)	(0.010)	(0.010)	(0.010)	(0.035)	(0.035)	(0.012)	(0.012)
Intermediate Algebra	0.109 **	0.078 **	0.032 **	0.022	-0.107	-0.100	0.054 **	0.042 **
	(0.017)	(0.016)	(0.016)	(0.017)	(0.068)	(0.067)	(0.019)	(0.019)
Advanced Algebra	0.134 **	0.088 **	0.042 **	0.029 **	-0.077	-0.082	0.054 **	0.039 **
	(0.014)	(0.014)	(0.014)	(0.015)	(0.050)	(0.050)	(0.017)	(0.017)
Calculus	0.195 **	0.120 **	0.065 **	0.047	-0.132	-0.140	0.077 **	0.058
	(0.033)	(0.032)	(0.032)	(0.032)	(0.167)	(0.167)	(0.036)	(0.036)
Total Math Credits	0.106 **	0.069 **	0.027 **	0.019 **	-0.009	-0.010	0.036 **	0.027 **
	(0.006)	(0.007)	0.007	(0.007)	(0.024)	(0.024)	(0.008)	(0.008)
Math GPA				0.036 **		0.063 **		0.039 **
				(0.009)		(0.015)		(0.009)
Other Controls								
Demographic Information	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family Characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Characteristics	No	Yes	Yes	Yes	Yes	Yes	No	No
Highest Educational Degree	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Estimation Method	OLS	OLS	OLS	OLS	IV	IV	FE	FE
R-squared	0.069	0.172	0.199	0.201	0.187	0.192	0.316	0.319
Number of Obs	5,919	5,919	5,919	5,896	5,864	5,841	5,919	5,896

Notes: ** Significant at the 5 percent level; * significant at the 10 percent level. Standard errors are in parentheses. All models include an intercept. Adding dummy variables for college major changes the coefficients only minimally. In the first-stage regressions of each IV-estimated model, the p-values for the F-test of the hypothesis that the coefficients on the six school-average instruments are equal to zero are 0.0001.

Table 3
The Effects of Specific Math, English, Science, and Foreign Language Courses on Log-Earnings: OLS and IV Estimates

	OLS (1)		OLS (2)		IV (3)		IV (4)	
Vocational	-0.030	**	-0.032	**	-0.075	*	-0.075	*
	(0.011)		(0.011)		(0.039)		(0.040)	
Pre-Algebra	0.004		0.002		0.030		0.029	
	(0.014)		(0.014)		(0.045)		(0.045)	
Algebra/Geometry	0.020	*	0.019	*	0.097	**	0.093	**
	(0.011)		(0.011)		(0.044)		(0.045)	
Intermediate Algebra	0.019		0.010		-0.129		-0.134	
	(0.017)		(0.017)		(0.082)		(0.082)	
Advanced Algebra	0.030	**	0.018		-0.093		-0.101	*
	(0.015)		(0.016)		(0.058)		(0.059)	
Calculus	0.043		0.028		-0.162		-0.187	
	(0.033)		(0.033)		(0.171)		(0.173)	
Below Level English	0.004		0.005		0.019		0.020	
	(0.013)		(0.013)		(0.033)		(0.033)	
Average English	0.015	**	0.015	**	0.025		0.024	
	(0.008)		(0.008)		(0.024)		(0.024)	
English Literature Courses	0.015	*	0.015	*	0.034		0.035	
	(0.009)		(0.009)		(0.029)		(0.029)	
Above Level English	0.026	**	0.025	*	0.071	**	0.076	**
	(0.013)		(0.013)		(0.036)		(0.036)	
Basic Biology	-0.008		-0.009		-0.080	*	-0.084	*
	(0.019)		(0.019)		(0.046)		(0.046)	
General Biology	-0.015		-0.014		-0.069	*	-0.070	*
	(0.013)		(0.013)		(0.040)		(0.040)	
Primary Physics	-0.023	**	-0.024	**	-0.060	*	-0.059	*
	(0.012)		(0.012)		(0.031)		(0.031)	
Secondary Physics	0.005		0.005		0.143		0.149	
	(0.034)		(0.034)		(0.100)		(0.100)	
Chemistry 1, Physics 1	0.020		0.015		-0.004		-0.012	
	(0.014)		(0.014)		(0.060)		(0.060)	
Chemistry 2, Physics 2,	0.020		0.020		0.028		0.045	
AP Biology	(0.020)		(0.020)		(0.073)		(0.071)	
Foreign Language (1-2)	0.025		0.028		-0.024		0.000	
	(0.017)		(0.017)		(0.104)		(0.103)	
Foreign Language (3-4)	0.054	**	0.056	**	0.129		0.138	
	(0.026)		(0.026)		(0.143)		(0.142)	
GPA - Math			0.036	**			0.067	**
			(0.009)				(0.020)	
R-squared	0.200		0.203		0.186		0.191	
Number of Observations	5,735		5,718		5,681		5,664	

Notes: ** Significant at the 5 percent level; *significant at the 10 percent level. Standard errors are in parentheses. All models control for demographic, family, school, and highest-degree characteristics. See Table 1 for a complete list. Each model contains an intercept. For the first-stage regressions in each IV model, the p-values for the F-test of the hypothesis that the coefficients on the instruments are equal to zero are 0.0001

Table 4
The Predicted Earnings Effect of Hypothetical Course Combinations During Grades 11 and 12

School Year & Level	Hypothetical Curriculum	Predicted Effect
10 Dropout	No more subjects	-0.122** (0.036)
11 Low	No math, average English, secondary physics, no foreign language	0.021 (0.035)
11 Med	Intermediate algebra, average English, chemistry 1, foreign language (third year)	0.054** (0.022)
11 High	Advanced algebra, advanced English, chemistry 1, foreign language (third year)	0.072** (0.021)
12 Low	Same as grade 11 low	0.021 (0.035)
12 Med	Advanced algebra, English literature, physics 1, foreign language (fourth year)	0.062** (0.019)
12 High	Calculus, advanced English, chemistry 2, foreign language (fourth year)	0.086** (0.037)

Notes: ** Significant at the 5 percent level. Standard errors are in parentheses. Except for the effect of dropping out (which is simply the coefficient on the dropout dummy variable in column 4 model of Table 2), the predicted effects are computed by summing up the individual effects of the hypothetical class list from column 2 of Table 3. In other words, the effects within each academic year (e.g., grade 12) are measured relative to a hypothetical student who stays in school that academic year but does not take any math, English, science, or foreign language courses. The standard errors are computed by taking the square root of the variance of the sum of the coefficients from the hypothetical class list. To compute the effects more easily, we estimate a slightly different specification of the column 2 model in Table 3 in which we enter the total number of foreign language credits rather than the two dummy variables. Coefficients on the other subjects are practically unchanged. That of foreign language credits becomes 0.014.

Table 5
The Effects of Aggregate Math, English, Science, and Foreign Language Courses on Log-Earnings: OLS and IV Estimates

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math	0.014 *	0.008	0.009	0.005	-0.029	-0.034	-0.040	-0.025
	(0.008)	(0.008)	(0.008)	(0.009)	(0.033)	(0.034)	(0.034)	(0.034)
English	0.011	0.013 *	0.016 **	0.017 **	0.030	-0.030	0.030	0.042 *
	(0.007)	(0.007)	(0.008)	(0.008)	(0.023)	(0.023)	(0.022)	(0.023)
Science	0.005	0.003	-0.001	-0.008	-0.025	-0.026	-0.030	-0.045 *
	(0.008)	(0.008)	(0.008)	(0.008)	(0.026)	(0.026)	(0.025)	(0.027)
Foreign Language	0.025 **	0.024 **	0.025 **	0.021 **	-0.034	0.036	0.053 **	0.010
	(0.007)	(0.007)	(0.007)	(0.008)	(0.028)	(0.028)	(0.026)	(0.030)
Controls								
Demographic Information	yes	yes	yes	yes	yes	yes	yes	yes
Parental SES	yes	yes	yes	yes	yes	yes	yes	yes
School Region/Type	yes	yes	yes	yes	yes	yes	yes	yes
School Resources	yes	yes	no	no	yes	yes	no	no
Highest Educational Degree	yes	yes	yes	yes	yes	yes	yes	yes
Test Scores (1982)	no	no	no	yes	no	no	no	yes
GPA	no	yes	yes	no	yes	yes	yes	no
Dropouts included in model	yes	yes	no	no	yes	yes	no	no
R-squared	0.193	0.197	0.182	0.182	0.189	0.193	0.178	0.179
Number of Obs	5,738	5,718	5,404	4,834	5681	5664	5,356	4,790

Notes: ** Significant at the 5 percent level; * significant at the 10 percent level. Standard errors are in parentheses. For the first-stage regressions in each IV model, the p-values for the F-test of the hypothesis that the coefficients on the instruments are equal to zero are 0.0001. Models 1 and 2 use the same demographic, family, and school characteristics found in Table 2. The difference between models 1 and 2 is that model 2 includes GPA. Models 3 and 4 use our data to mimic Altonji's (1995) Table 2, model 7. These models exclude high school dropouts. In these models, the demographic data include ethnicity, gender, and age. Age is our best proxy for Altonji's control for the number of months in the labor force. The parental SES category includes parental education, parental income, and whether the parent is a U.S. native. We do not have a comparable variable to Altjoni's control for parental involvement. We also include school region and type, as Altonji does, but we do not control for city size and college proximity. These replications exclude the respondent's marital status, the number of siblings, and a host of high school characteristics, all of which are included in models 1 and 2. The difference between models 3 and 4 is that model 3 controls for ability using GPA (our preferred specification), where as model 4 controls for ability using the student's senior year test scores in math, reading, and vocabulary (our closest approximation to Altonji's specification). The progression of models in this table highlights how our results can be compared to Altonji's when we use the total number of credits earned in the different subjects. Models 5 through 8 repeat this progression using IV estimates instead of OLS estimates. Comparing results from this table to those in Table 3 also highlights how using the total number of credits is not as informative as using detailed credit counts. Models 1 and 2 in this table corresponds to models 1 and 2 in Table 3. Whereas the non-curriculum controls are the same in both models, this table uses the total number of credits in a particular subject whereas Table 3 uses detailed credit counts in each subject. Similarly, models 5 and 6 in this table corresponds to models 3 and 4 in Table 3, respectively.

Table 6
Earnings Gaps Based on Ethnicity and Parental Income
(Before and After Curriculum is Added to the Model)

	(1)	(2)	(3)	(4)	(5)	(6)
Hispanic	-0.052 **		0.010	0.036 *	0.026	0.043 **
	(0.019)		(0.021)	(0.021)	0.021	(0.021)
Black	-0.100 **		-0.010	-0.0002	0.004	0.008
	(0.024)		(0.025)	(0.0249)	(0.025)	(0.025)
Asian	0.092 **		0.072 *	0.023	0.033	0.010
	(0.043)		(0.044)	(0.043)	(0.043)	(0.043)
Native American	-0.229 **		-0.102 **	-0.068	-0.088 *	-0.064
	(0.054)		(0.052)	(0.051)	(0.052)	(0.051)
< $7K		-0.293 **	-0.171 **	-0.124 **	-0.159 **	-0.122 **
		(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
$7K - $15K		-0.098 **	-0.039 *	-0.025	-0.041 *	-0.029
		(0.023)	(0.022)	(0.022)	(0.022)	(0.022)
$15K - $20K		-0.026	0.002	0.001	0.001	-0.0001
		(0.025)	(0.024)	(0.023)	(0.024)	(0.0233)
$25K - $38K		0.096 **	0.070 **	0.070 **	0.068 **	0.068 **
		(0.026)	(0.025)	(0.025)	(0.025)	(0.025)
$38K and higher		0.112 **	0.058 **	0.059 **	0.061 **	0.060 **
		(0.030)	(0.029)	(0.029)	(0.029)	(0.029)
Male			0.266 **	0.272 **	0.285 **	0.286 **
			(0.014)	(0.014)	(0.014)	(0.014)
Curriculum	No	No	No	Yes	No	Yes
GPA	No	No	No	No	Yes	Yes
R-squared	0.007	0.035	0.137	0.172	0.156	0.178
Number of Obs	5,919	5,919	5,919	5,919	5,896	5,896

Notes: ** Significant at the 5 percent level; * significant at the 10 percent level. Standard errors are in parentheses. The effects of ethnicity are measured relative to whites. The effects of parental income are measured relative to students from families with incomes between $20,000 and $25,000. Each column represents one model. Column 1 contains only ethnicity controls. Column 2 contains only parental income controls. The remaining models control for all of the demographic, family, and school characteristics listed in Table 1 but exclude the student's highest degree. The change in the coefficient from column 3 to column 4 represents the portion of the earnings gap that curriculum can explain without controlling for GPA. Changes from column 5 to column 6 account for GPA. The percentage of students in the income categories on the above table are 8, 28, 18, 15, and 9, respectively. The omitted category contains 17 percent of the students. We do not display results for the group of students who had missing data about their family income (this explains why the percentages do not sum to 100).

Table A.1
Mathematics Course Classification

Course Label	Included Courses
Non-Academic (Vocational)	General (1 and 2), Basic (1, 2, and 3), Consumer, Technical, Vocational, Review
Low-Academic (Pre-Algebra)	Pre-Algebra, Algebra 1 (Part 1), Algebra 1 (Part 2), Geometry Informal
Middle Academic I (Algebra/Geometry)	Algebra 1, Geometry (Plane and Solid), Unified 1, Unified 2
Middle Academic II (Intermediate Algebra)	Algebra 2, Unified 3
Advanced I and II (Advanced Algebra)	Algebra 3, Algebra-Trigonometry, Analytic Geometry, Linear Algebra, Probability, Statistics, Pre-Calculus
Advanced III (Calculus)	Advanced Placement Calculus, Calculus-Analytic Geometry, Calculus

Notes: The math classification system used by NCES included a separate category for pre-calculus (called Advanced II), but after some initial analysis, we decided to combine it with the Advanced I level.

Table A.2
Science Course Classification

Course Label	Included Courses
Basic Biology	Basic Biology
General Biology, Secondary Life Science	General Biology 1, Secondary Life Sciences (ecology, marine biology, zoology, human physiology)
Primary Physics	Primary Physical Sciences (applied physical science, earth science, college prep earth science, unified science), General Science
Secondary Physics	Secondary Physical Sciences (astronomy, environmental science, geology, oceanography, general physics, consumer chemistry, introductory chemistry)
Chemistry 1 and Physics 1	Chemistry 1, Physics 1
Chemistry 2, Physics 2, AP Biology	Chemistry 2, Physics 2, Advanced Placement Biology

Notes: This classification system is a synthesis of the three different science classification systems provided by NCES.

Table A.3
Number of Additional Missing Observations
As More Restrictions Are Enacted

Remaining Observations Tally			Lost Observations Relative to Most Recent Tally
Total Observations in HSB Data	14,825		
		-3,101	Lose 21 percent of total sample because some students attend private schools.
Public School Observations	11,724		
		-2,608	Lose 22 percent of public school sample because of missing earnings data.
With Non-Missing Earnings Data	9,116		
		-1,208	Lose 13 percent of non-missing earnings data because earnings are less than $2,000.
		-28	Lose 0.3 percent of non-missing earnings data because earnings are greater than $75,000.
Between $2,000 and $75,000	7,880		
		-896	Lose 11 percent of valid earnings data because of missing curriculum data.
With Valid Curriculum Data	6,984		
		-418	Lose 6 percent of those with valid earnings and curriculum data because they transferred schools.
		-219	Lose 3 percent of those with valid earnings and curriculum data because they are enrolled in 1991.
		-460	Lose 7 percent of those with valid earnings data because they have missing 1991 enrollment data.
		36	Must add back 36 observations so that we do not double count those who transferred schools *and* are either enrolled or missing enrollment data.
		-4	Lose 4 observations because of missing age data.
Observations Remaining for Analysis	5,919		

Notes: This table documents how the 14,825 observations in HSB are reduced to 5,919 usable regression observations. The right-hand column records the number of omitted observations for each given reason and the left-hand column keeps a running tally of the remaining observations after each loss. The number and percentage of observations lost are calculated relative to the most recent tally.

Table A.4
Missing Observations out of 11,724 Total Public School Observations

Type of Missing Observation	Number of Missing Observations	Missing Observations out of 11,724 (Percent)
Missing Earnings Data	2,608	22.2
Out of Range Earnings Data	1,236	10.5
Missing Curriculum Data	1,591	13.6
Transfer Students	949	8.1
Students Enrolled in Postsecondary Education	304	2.6
Missing Enrollment Data	2,426	20.7
Transfers *and* Invalid Enrollment (enrolled or missing)	255	2.2

Notes: This table documents the number and percentage of missing values for some key variables in the analysis. Unlike the previous table, the number of missing values in this table is always calculated relative to the total number (11,724) of public school observations.

Table A.5
Summary Statistics of Key Variables

	HSB Public School		Regression Sample	
	Unweighted	Weighted	Unweighted	Weighted
Annual 1991 Earnings ($)	19,168	19,092	22,288	22,077
	(13,532)	(13,534)	(10,954)	(10,929)
Log of 1991 Earnings	9.755	9.757	9.872	9.860
	(0.790)	(0.782)	(0.574)	(0.580)
Math Curriculum Measures				
Vocational Math Credits	0.758	0.757	0.685	0.708
	(0.912)	(0.932)	(0.887)	(0.904)
Pre-Algebra Credits	0.258	0.262	0.261	0.262
	(0.541)	(0.552)	(0.543)	(0.546)
Algebra / Geometry Credits	0.908	0.917	0.988	0.945
	(0.874)	(0.877)	(0.881)	(0.865)
Intermediate Algebra Credits	0.265	0.266	0.294	0.279
	(0.461)	(0.470)	(0.476)	(0.469)
Advanced Algebra Credits	0.223	0.212	0.250	0.228
	(0.524)	(0.514)	(0.543)	(0.522)
Calculus Credits	0.042	0.039	0.045	0.039
	(0.220)	(0.212)	(0.225)	(0.209)
Educational Attainment				
Higher Than Bachelor's Degree	0.029	0.031	0.032	0.030
	(0.168)	(0.173)	(0.175)	(0.170)
Bachelor's Degree	0.166	0.177	0.222	0.210
	(0.372)	(0.382)	(0.416)	(0.408)
Associate's Degree	0.070	0.078	0.085	0.084
	(0.255)	(0.268)	(0.279)	(0.277)
Certificate	0.090	0.110	0.102	0.109
	(0.287)	(0.313)	(0.303)	(0.312)
High School Plus	0.187	0.190	0.202	0.191
	(0.390)	(0.392)	(0.402)	(0.393)
High School Diploma	0.261	0.337	0.297	0.320
	(0.439)	(0.473)	(0.457)	(0.466)
Less Than High School	0.060	0.061	0.051	0.049
	(0.238)	(0.239)	(0.219)	(0.215)
Degree Missing	0.137	0.015	0.009	0.008
	(0.344)	(0.122)	(0.092)	(0.088)
Additional Controls				
Math GPA	2.075	2.096	2.164	2.139
	(0.962)	(0.933)	(0.934)	(0.919)
Math IRT Test Score	11.778	12.249	13.184	12.893
	(9.823)	(9.736)	(9.726)	(9.679)
Number of Observations	11,724	11,724	5,919	5,919

Note: Similar statistics for student and family demographics and school characteristics are excluded to save space, but they are available upon request.

References

Ackerman, Deena (2000). "Do The Math: High School Mathematics Classes and the Lifetime Earnings of Men," manuscript, Department of Economics, University of Wisconsin – Madison and University of Vermont.

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

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

Betts, Julian R. (1998). "The Impact of Educational Standards on the Level and Distribution of Earnings," *American Economic Review*, 88(1):266-275.

Bishop, Hohn H. (1989). "Why the Apathy in American High Schools?" *Educational Researcher*, 18:6-10, 42.

Bowles, Samuel, Herbert Gintis and Melissa Osborne (2001). "The Determinants of Earnings: A Behavioural Approach," *Journal of Economic Literature*, 39(4):1137-1176.

Costrell, Robert M. (1994). "A Simple Model of Educational Standards," *American Economic Review*, 84(4): 956-971.

DuMouchel, William H. and Greg J. Duncan (1983). "Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples", *Journal of the American Statistical Association*, 78(383):535-542.

Gamoran, Adam (1998). "The Impact of Academic Course Work on Labor Market Outcomes for Youth Who Do Not Attend College: A Research Review," in Adam Gamoran and Harold Himmelfarb, eds., *The Quality of Vocational Education: Background Paper from the 1994 National Assessment of Vocational Education*, National Institute on Postsecondary Education, Libraries, and Lifelong Learning Office of Educational Research and Improvement U.S. Department of Education, Washington, D.C.

Griliches, Zvi (1977). "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45(1):1-22.

Grogger, Jeff and Eric Eide (1995). "Changes in College Skills and the Rise in the College Wage Premium," *Journal of Human Resources*, 30(2):280-310.

Grogger, Jeff (1996). "School Expenditures and Post-Schooling Earnings: Evidence from High School and Beyond," *The Review of Economics and Statistics*, 78(4):628-637.

Ingels, Steven, et al. (1995). *National Longitudinal Study of 1988 – Second Follow-Up: Transcript Component Data File User's Manual*, NCES 94-377, Washington, D.C.: National Center for Education Statistics.

Lawrance, Emily C. (1991). "Poverty and the Rate of Time Preference: Evidence from Panel Data," *Journal of Political Economy* 99 (February), 54-77.

Levine, Phillip B. and David J. Zimmerman (1995). "The Benefit of Additional High-School Math and Science Classes for Young Men and Women," *Journal of Business and Economic Statistics,* 13(2):137-149.

Lillard, Dean R. (1998). "Accounting for Substitution between Credentials in Estimates of the Returns to GED and High School Diplomas," unpublished manuscript, Cornell University, Ithica, New York.

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

Porter, A. C., M. W. Kirst, E. J. Osthoff, J. S. Smithson, and S. A. Schneider (1993). "*Reform Up Close: An analysis of high school mathematics and science classrooms*," Consortium for Policy Research in Education Final Report, Madison, Wisconsin: University of Wisconsin-Madison, Wisconsin Center for Education Research.

Spence, Michael (1973). "Job Market Signalling," *Quarterly Journal of Economics*, 87(3):355-74.