Education, Income, and Ability

Zvi Griliches

Harvard University

William M. Mason

Duke University

I. Introduction

Current estimates of the contribution of education to economic growth have been questioned because they ignore the interaction of education with ability. Whether the neglect of ability differences in the analyses of the income-education relationship results in estimates that are too high was considered in an earlier paper by one of the authors (Griliches 1970), and a negative answer was conjectured. In this paper, we pursue this question a bit further, using a new and larger body of data. Unfortunately, a definitive answer to this question is hampered both by the vagueness and elasticity of "education" and "ability" as analytical concepts and by the lack of data on early (preschooling) intelligence.

The data examined in this paper are based on a 1964 sample of U.S. military veterans. The variables measured include scores on a mental ability test, indicators of parental status, region of residence while growing up, school years completed before service, and school years completed during or after service. These have allowed us to inquire into the separate effects of parental background, intelligence, and schooling.

The basic problem and analytical framework can be set out very simply. Let income be a linear function of education and ability, or, $Y = \alpha + \beta_1 E + \beta_2 G + u$, where Y is income, E is education, G is ability,

This work has been supported by National Science Foundation grant no. GS2762X. This paper was originally prepared for the Social Science Research Council Conference on Structural Equations Models, Madison, Wisconsin, November 1970. We are indebted to Paul Ryan for research assistance and to Edward F. Denison, O. D. Duncan, A. S. Goldberger, A. C. Kerckhoff, and K. O. Mason for comments on earlier drafts.

and u represents other factors affecting income, which are assumed to be random and uncorrelated with E and G. The relation is presumed to hold true for cross-sectional data. If education and ability are positively associated, then a measure of the contribution of education to income that ignores the ability variable (most commonly, the simple least-squares coefficient of Y on E) will be biased upward by the amount $\beta_2 b_{GE}$, where b_{GE} is the regression coefficient of ability on education in the particular sample. The first substantive section of this paper (Section III) investigates the magnitude of this bias via the estimation of income-generating equations containing measures both of education and ability.[1]

In our data the output of the educational process is measured by the number of school grades completed in the formal education system, while ability is measured by the performance on a test at an age when most of the schooling has already been completed. Both of these measures are far from ideal for our purposes. Consider the education variable: What we would like to have is a measure of education achieved (E); what we have is years of schooling completed (S) without reference to the conditions under which individuals obtained their formal schooling and the kinds of schooling pursued. Let us call the discrepancy between these two variables "quality" (Q, where $E = S + Q$) and assume that it is uncorrelated with the quantity of schooling (S).[2] At the same time, the quality of schooling is likely to be correlated with ability because (1) there is some correlation between socioeconomic status and ability, (2) more able students are more likely to get into better schools, and (3) performance on intelligence tests taken at age 18 or so also reflects in part differences in both the quantity and quality of education.

Allowing for differences in the quality of education makes the assessment of the bias in the estimated education coefficient somewhat more complicated. The true income-generating equation becomes $Y = a + \beta_1 E + \beta_2 G + u = a + \beta_1 S + \beta_1 Q + \beta_2 G + u$.

In this framework, ignoring not only G but also Q leads to the same result as before since b_{QS} (the regression coefficient of quality on quantity of schooling) is zero by assumption. But when a measure of ability is included in the estimating equation, the estimated education coefficient becomes $b_{YS.G} = \beta_1 + \beta_1 b_{QS.G}$, where $b_{QS.G}$ is the partial regression co-

[1] Concern with the accuracy of the education estimate due to the omission of ability may, of course, be readily extended to other factors associated with educational attainment and known also to contribute to the determination of socioeconomic outcomes. Denison (1964), for instance, notes the salience of race, inherited wealth, family position, and diligence, and the list can easily be lengthened. In the present analysis we control for these factors to a considerable degree.

[2] This is not too unreasonable an assumption since there is a wide variation in quality of education at all levels of schooling. It is possible, however, that children going to better schools also are more likely to accumulate more years of schooling. If that is the case, we define Q to be that part of the "quality" distribution which is uncorrelated with "quantity." The rest follows in a similar manner.

efficient of quality on quantity of schooling, holding ability constant.[3] Given our assumptions, it can be shown (see the Appendix) that $b_{QS.G} = -b_{QG} \cdot b_{GS}/(1 - r^2_{GS})$, where r^2_{GS} is the square of the correlation coefficient between the quantity of schooling and ability. Since we expect both b_{QG} (the regression coefficient of educational quality on individual ability) and b_{GS} (the regression coefficient of individual ability on quantity of schooling) to be positive, $b_{QS.G}$ will be negative. Substituting this expression for $b_{QS.G}$ back into the expression for $b_{YS.G}$ gives $b_{YS.G} = \beta_1 - \beta_1 b_{QG} \cdot b_{GS}/(1 - r^2_{GS})$. Since $b_{YS} = \beta_1 + \beta_2 b_{GS}$, it is clear that by going from b_{YS} to $b_{YS.G}$ we reduce the coefficient of schooling for two reasons. First, we eliminate the upward bias due to the earlier omission of ability. Second, however, we *introduce* another bias due to the correlation of ability with the left-out quality variable. This new bias is partly a function of the magnitude of the correlation between quantity of schooling and ability. We solve the problem of this second bias by concentrating our attention on that part of schooling occurring during or after military service (SI—schooling increment), which turns out to be almost entirely uncorrelated with our measure of ability and hence is not subject to this type of bias.

The availability of the schooling-increment variable also helps us to solve another vexing problem—how to disentangle the question of causality when the available measure of ability may itself be in part the result of schooling. Since the intelligence test available in these data is administered pior to entering service, performance on it cannot be affected by the schooling increment. Thus, because our measure of ability is causally prior to SI, and because using SI reduces the bias problem in estimating the effects of education on income, we shall be putting most of the stress on the results for only a *part* of schooling (SI) in the subsequent sections.

We have already noted that our ability measure is not ideal because it is obtained after most of the formal schooling has been completed. What we would like is a measure of ability obtained before the major effects of the school system have been felt. Although it is possible using data such as ours to construct models which incorporate estimates of the effects of *early* ability (see Duncan 1968; Bowles 1970), we have chosen to work exclusively with our measure of *late* ability. Given this decision, the ability variable we work with still is not ideal for our purposes. For it is possible that our measure of ability, *taken as a measure of late ability*, has errors in it. These errors may have a number of sources, and some may be random, others nonrandom. To the extent the errors are random, we know that a direct application of least squares in their presence may understate the effect of ability on income and simultaneously bias the estimated education

[3] These formulae hold as computational identities between least-squares coefficients. They also can be interpreted as expectations of computed least-squares coefficients from random samples from a population satisfying our assumptions.

coefficient upward. To circumvent this effect of random errors we devise, in Section IV, a model of income determination that contains an unobserved achievement variable in place of measured ability. Manipulation of this model leads to equations estimable by means of a two-stage or instrumental-variables approach and secures a reading of the effect of ability freed of random errors.[4]

In Section V our results are summarized and compared with previous work in this field. Unlike other studies, we can focus on a relatively independent part of total schooling—that gained during or after military service. This gives us a less-biased estimate of the effect of a change in schooling than was possible before.

II. The Sample and the Variables

Our analysis is based on a sample of post–World War II veterans of the U.S. military, contacted by the Bureau of the Census in a 1964 Current Population Survey (CPS). The population consists of men who were then in the age range of 16–34 years, primarily the ages of draft eligibility. The sample includes about 3,000 veterans for whom supplementary information from individual military records was collated with the CPS questionnaire responses.[5] Of special interest to us is that a substantial proportion of the veterans' military records contain individual scores on the Armed Forces Qualification Test (AFQT), which we use here in lieu of standard civilian mental ability (IQ) tests.

The men who serve in the U.S. military do not represent any recent cohort of draft-age men, since those at either extreme of the ability and socioeconomic distributions are less likely to serve than those in the middle.[6] Thus, conclusions based on our analysis of these data apply only to the veterans' population. But, since this population is sizable, the data are of interest despite their obvious limitation. Moreover, this is one of the few relatively large sets of data combining information on income, education, demographic characteristics, mental test scores, and family socioeconomic background. The latter three are important as controls in estimating the income-education relationship.

[4] Ideally we would like to correct all of our variables for random errors. But although it is possible to adjust some others besides ability for random errors (Siegel and Hodge 1968), we do not have enough information to adjust them all. Since our major interest is with changes in the education coefficient due to the inclusion of the ability measure, the errors in the latter are most crucial to our analysis.

[5] See Rivera (1965) and Klassen (1966) for a description of the sample. Duncan (1968) and Mason (1968, 1970), among others, used these data.

[6] Educational deferments have channeled substantial numbers of young men into entirely civilian careers, and a low score on the AFQT reduces the probability of being drafted. For a general discussion of this aspect of the Selective Service System see U.S. President's Task Force on Manpower Conservation (1964). Davis and Dolbeare (1968) give an overview of Selective Service.

Within the veterans' sample, the individuals on whom we base our conclusions are 1,454 men who were employed full time when contacted by the CPS; who were between the ages of 21 and 34 and not then enrolled in school; who were either white or black; who provided complete information about their current occupation, income, education, family background; and for whom AFQT scores were available.[7]

The major characteristics of our sample and the variables we used are summarized in table 1. The definition and measurement of most of the

TABLE 1

MEANS AND STANDARD DEVIATIONS OF VARIABLES: VETERANS AGE 21–34 IN 1964 CPS SUBSAMPLE

Variable	Mean or Fraction in Sample	SD	Symbol in Subsequent Tables	Group Name
Personal background:				
Age (years)	29.0	3.5	Age	
Color (white)	0.96	*	C	
Schooling before service (years)	11.5	2.3	SB	
Total schooling (years)	12.3	2.5	ST	
Schooling increment (years)	0.8	1.4	SI	
AFQT (percentile)	54.6	24.8	AFQT	
Length of active military service (months)	30.7	16.9	AMS	
Father's schooling (years)	8.7	3.2	FS	Fa. stat.
Father's occupational SES	29.0	20.6	FO	Fa. stat.
Grew up in South	0.29	*	ROS	Reg. bef.
Grew up in large city	0.22	*	POC	Reg. bef.
Grew up in suburb of large city	0.05	*	POS	Reg. bef.
Current location:				
Now living in the South	0.27	*	RNS	Reg. now
Now living in the West	0.15	*	RNW	Reg. now
Now living in an SMSA	0.68	*	SMSA	Reg. now
Current achievement:				
Length of time in current job (months)	54.3	42.8	LCJ	Curr. exp.
Never married	0.14	*	NM	Curr. exp.
Current occupational SES	39.2	22.7	...	
Log current occupational SES	3.47	0.68	LOSES	
Actual income (weekly, dollars)	122.5	52.4	...	
Log actual income	4.73	0.40	LINC	

NOTE.—$N = 1,454$, for this and subsequent tables based on the 1964 CPS. Fa. stat. = father's status; reg. bef. = region before; reg. now = region now; curr. exp. = current experience.
* The standard deviation for a dummy variable is equal to $\sqrt{f(1-f)}$, where f is the fraction in the sample having the requisite characteristic. Thus, it is computable from the numbers given in the first column.

[7] The variables noted above account for the greatest reduction in sample size, but

variables is standard, and we shall comment here only on a few of the more important ones.

Income is gross weekly earnings in dollars. It is an answer to the request: "Give your usual earnings on this job before taxes and other deductions." The data provide also another concept of income, "earnings expected from all jobs in 1964." We experimented at some length with both concepts of income, getting somewhat better (more stable) results for the first (actual) income measure. Since the major results were similar for both measures of income, we shall report here only those for the first (actual) income measure. We also experimented a bit with functional form before settling on the semilog form for the "income-generating" function leading to the use of the logarithm of income (LINC) as our main dependent variable.

Education is measured in years of school (highest grade) completed and is recorded at two points in time: before entry into military service and at the time of the survey. By taking the difference between total grades of school completed (ST) and grades of school completed before military service (SB) we get a measure of the increment in schooling (SI) acquired during or after military service.[8] The minimum value of this variable is zero (no increment in schooling), and the maximum is six grades. As noted above, this incremental measure of education is central to our analysis both because it occurs after the time at which ability was measured and because it is so little correlated with our measure of ability.

Performance on the AFQT is scaled as a percentile score estimated from eight grouped categories.[9] This test includes questions on vocabulary, arithmetic, and spatial relations, but also contains a section on tool knowledge. The AFQT has been treated by other investigators (including Duncan 1968; and Jensen 1969) as an intelligence test, so that we are following in the footsteps of others in this regard. We are unaware, though, that the comparability of the AFQT with civilian intelligence tests has ever been documented.[10]

the data file used also contains a number of other variables of interest and is consequently slightly smaller than it would be solely on the basis of the above-mentioned variables.

[8] Each of the education measures is based on eight categories of school years completed and is scored as follows: Less than 8 years = 4; 8 years = 8; 9–11 years but not high school graduate = 10; high school graduate = 12; some college but less than 2 years = 13.5; 2 or more years of college but no degree = 15; B.A. = 16; and graduate study beyond the B.A. = 18. As a matter of convenience we will hereafter refer to SI as *post*service schooling, ignoring the possibility that some of the increment may have occurred while the man was in service.

[9] The percentile scores are the midpoints of each of the eight categories provided in the data. For a number of individuals in the sample there were records of results for mental tests other than the AFQT. Prior to our acquisition of the data these scores were converted to AFQT-equivalents following instructions provided by the Department of Defense. Despite use of the AFQT to select individuals into the armed forces, all levels of performance on the AFQT are represented in our sample.

[10] We would welcome information on this point. Our own review turned up nothing about the reliability of the AFQT or about correlations between it and civilian IQ

It is clear from even this brief discussion of the AFQT that *some* error may arise from using the AFQT as an intelligence test in addition to the kinds of errors which could be present in using one of the standard civilian IQ tests.[11] Another difficulty with the use of the AFQT in our analysis, a difficulty which is inherent in the use of *any* global IQ test for purposes such as ours, is that IQ by definition is an aggregation of several different traits (for example, verbal and mathematical ability) sampled from some larger population of traits. The weights used in combining these traits to obtain a global IQ score are not necessarily those which would maximize the contribution of each trait to some other variable (such as income). Therefore, the use of AFQT instead of the separate traits which comprise it, and the use of only those traits, may lead to attenuation in our estimate of the effect of ability on income. This explains our interest in the errors-in-variables approach to be taken up in Section IV.

The long list of other variables considered can be divided, somewhat imperfectly, into personal background and current location and success variables. In the first group, we have the usual variables for age (in years), color dummy (white = 1, black = 0), and region and place of *origin* dummies (these are in terms of places "you lived most until age 15") that record growing up in the South, in a large city (over 100,000 in population), or in a suburb of such a city. In addition to these, we also have two measures of parental status: father's schooling (in years of school completed—FS) and father's occupation (FO, coded according to Duncan's 1961 SES scale).[12]

The age variable is usually included in such studies because older men (within the range of our data) are likely to have had more training on the job and more opportunity to find the better jobs that are appropriate to their training. This, however, is probably measured better not by calendar time but by the actual time spent in the civilian labor force accumulating

---

tests. Karpinos (1966, 1967), the only articles we found discussing the AFQT, focused on characteristics of those failing the test, not the test itself. We have seen fragmentary evidence about the AGCT, predecessor of the AFQT, but to extrapolate from experiences with the former to the latter would be merely to speculate.

[11] If the AFQT is not virtually interchangeable with the standard civilian IQ tests, then Jensen (1969) could well be wrong in assuming that the heritability of the AFQT is the same as for the standard civilian tests. Griliches (1970, pp. 92–104) suggests that the heritability of the AFQT may be lower than Jensen supposes, and pursues related issues.

[12] These are, of course, only incomplete measures of the family's socioeconomic status and are subject moreover to the possibility of recall error and misperception by respondents (sons) from whom this information was elicited. Blau and Duncan (1967, appendices D and E) take up the issue of recall error for these two variables in their occupational changes in a generation (OCG) sample. Conclusions drawn from their discussion should apply here, since the OCG sample is comparable with ours in the same age group. For evidence on this see Duncan (1968), who reports virtually identical correlations between father's education and occupation for the OCG and the CPS sample from which we draw.

work "experience."[13] We can estimate this roughly by defining: potential experience = age — 18 — (education before service — 12) — education after service — (total months in service)/12. Since this measure is a linear function of variables that we include anyway (age and schooling), there is no need to compute it explicitly. It does provide, however, an interpretation for the role of time spent in military service (AMS), when the latter variable is introduced separately.[14]

The "current location and success" variables are represented by a regional dummy variable classification of current location as south, northeast-northcentral and west (RNS and RNW); a dummy variable for current residence in a Standard Metropolitan Statistical Area (SMSA); a measure of the length of time on current job (LCJ, in months); a dummy variable for never married (NM) as opposed to other possibilities; and a measure of the socioeconomic status of the individual's current occupation (LOSES, the logarithm of Duncan's occupational SES scale). Each of these factors intervenes between education and income and helps to explain the relationship between these two variables. For example, more education may lead to greater interpersonal competence and other socially desirable characteristics which in turn may lead to a greater likelihood of being married. Individuals in this status may be expected to have the incentive of responsibility for others, and this may in turn lead to higher income.

Although we present some results that take into account factors intervening between education and income, they are not of central interest to us. We shall, therefore, not emphasize them in our discussion but concentrate instead on the contribution of the education and ability estimates in the presence of background factors alone.

Table 1 presents means and standard deviations for the variables to be used. Note that this group of veterans is young and hence will not exhibit differentials in income by education as large as those occurring in later, peak-earnings years. Also, because the number of blacks is quite small, white-black income differences will be characterized only by the multiplicative coefficient for the *color* dummy variable (since we are using the logarithm of income as our dependent variable). Although there are "interactions" between the color dummy variable and some of the other variables in the income-generating equation (Duncan 1969), there are too few blacks to estimate reliably the coefficients of the interaction terms. Observe, finally, that the average increment in schooling for this group of men is nearly one complete grade (0.8). Actually, 68 percent of the group did not return to school after service, so that those with additional schooling must have completed on average more than one additional grade. Since the

13 The use of such a measure was suggested to us by Jacob Mincer.

14 There is scant reason (Mason 1970) to believe that military service conveys a subsequent advantage in the civilian labor force. Thus we expect the AMS variable to have a negative coefficient in the income-generating equation.

grades completed range from a high school grade to a graduate school grade, it appears that the incremental-schooling variable may justifiably stand alone in the income-estimating equations.

In table 2 we list the simple correlation coefficients between the major variables of our sample. Note that there is very little correlation between the increment in schooling (SI) and various personal background variables such as color, father's schooling and occupation, and the respondent's AFQT score. None of these accounts for more than 1 percent of the variance of the schooling-increment variable. We have in this variable something as close to a well-designed experimental situation as we are likely to get in social science statistics.

III. Direct Results

A major objection to the usual estimates of the contribution of education to economic growth is their dependence on cross-sectional income-schooling relationships. The latter are likely to overestimate the "true" effect of schooling because of its intercorrelation with the omitted measures of social status and mental ability. Our sample provides two ways of meeting this objection. First, we do have measures of ability and parental status and can thus attempt to control for these biases directly. But more importantly, we can break down our schooling variable into two, the second part of which, the schooling increment (SI), is much less related to such other factors and hence also much less subject to such bias.

The causal model we use to guide our assessment of the relationships between income, education, ability, and other variables at our disposal can be stated as follows (using the variable labels given in table 1): (1) SB $= F$(fa. stat., reg. bef., C); (2) AFQT $= G$(fa. stat., reg. bef., C, SB); (3) AMS $= H$(fa. stat., reg. bef., C, age, SB, AFQT); (4) SI $= J$(fa. stat., reg. bef., C, age, SB, AFQT, AMS); (5) LINC $= K$(fa. stat., reg. bef., C, age, SB, AFQT, AMS, SI); where each of these functional relationships indicates a (linear) structural equation. Figure 1 provides a slightly more globally stated graphic equivalent to (1)–(5). As it stands, the model is given by a set of recursive equations. Including other functional relationships linking current achievement and location variables to income and other factors would lead to some simultaneous relationships, and in any case would complicate the model unnecessarily for our purposes. Thus, since we are primarily interested in the *total* effects of schooling and ability *net* of potential labor-force experience and background factors, we will not report on all the structural equations that inclusion of occupational SES, marital status, and other variables would entail.[15] For

[15] At one point (see table 5), though, we do use some of these additional variables to expand the list of regressors in order to determine the maximum ability of our data to predict income.

TABLE 2

CORRELATIONS (r) BETWEEN SELECTED VARIABLES IN THE 1964 CPS SUBSAMPLE

Variables	Variables (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Age	1.000	—.055	—.010	.109	.052	—.056	.216	.120	—.093	—.004
(2) Color	...	1.000	.011	—.028	—.006	.174	.116	.031	.004	.089
(3) Schooling before AMS	...	...	1.000	—.170	.832	.469	.264	.397	.283	.307
(4) Schooling increment	...	...	...	1.000	.405	.098	.149	.216	.103	.085
(5) Total schooling	...	...	...	...	1.000	.490	.329	.490	.321	.333
(6) AFQT	...	...	...	...	...	1.000	.235	.311	.229	.242
(7) Log income	...	...	...	...	...	...	1.000	.338	.114	.229
(8) Log occupational SES	...	...	...	...	...	...	...	1.000	.250	.266
(9) Father's schooling	...	...	...	...	...	...	...	...	1.000	.431
(10) Father's occupational SES	...	...	...	...	...	...	...	...	...	1.000

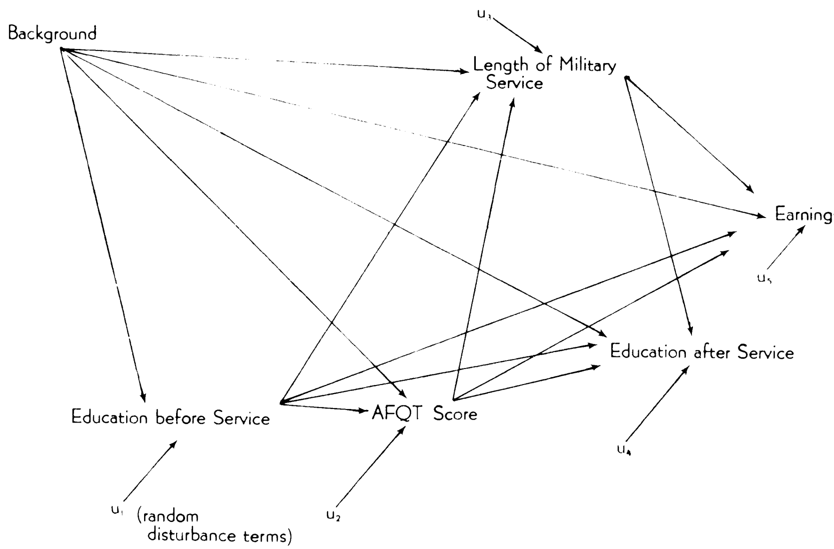

FIG. 1.—Basic causal model for determination of earnings. u_i = random disturbance terms.

the same reason, we concentrate in this section on the income equation, using the actual estimates for the rest of the causal model only to obtain a few secondary results.

The organization for the rest of this section is as follows: First we describe the sensitivity of the education coefficients to inclusion of ability and personal background characteristics in the income-generating equation. At this point we also appraise more generally the contribution of education to income. Next we describe the contributions of ability, background, color, and other variables in the income-generating equation, and, to some extent, their contributions in the model taken as a whole. Finally, we summarize some of the relationships between variables other than income.

Education.—We began this paper with a concern about the bias in the schooling coefficient due to the omission of ability. There are, however, several different ways of measuring this bias. We have already stated the need to take into account personal background factors. Doing so means that the estimated bias in the schooling coefficient due to omitting ability can be computed before or after the inclusion of personal-background factors in the regression. We also are including two schooling variables, so that there are two schooling coefficients to examine for each assessment of bias, although we have emphasized that the coefficient of schooling after service is preferable to the coefficient of schooling before service. Therefore, to derive the needed bias figures, our procedure is to regress income on (1) education; (2) education and ability; (3) education and personal-background factors; (4) education, personal-background factors, and ability. Comparisons made among these four regressions will provide the necessary figures for assessing bias.

Table 3 presents a number of regression results relating the logarithm of income to selected variables at our disposal. All of the regressions include age, length of military service, and color, so that the education, ability, and background effects are all net of color and the potential experience variable defined earlier. The first four regressions are directly relevant to deriving the reduction of the schooling coefficients due to the inclusion of personal-background factors and ability. For comparative purposes, regressions 5–8 parallel regressions 1–4 but use total schooling instead of the two separate schooling components.

Regression 1 provides the "baseline" estimates of the two schooling coefficients, estimates that do not allow for the effects of ability, father's status, and region of origin. Regressions 2 and 3, respectively, add AFQT and personal-background factors to the baseline regression. Regression 4 includes both AFQT and personal-background factors. By taking one minus the ratio of the education coefficient after including the factor to the corresponding education coefficient before including the factor, we get the proportionate bias in the schooling coefficients due to the omission of a relevant factor as a proportion. These calculations applied separately to regressions 1–4 and 5–8 provide the estimates shown in table 4.

Looking first at the figures for SB and SI, the introduction of the AFQT variable leads to a drop of 7–10 percent in the coefficient of SI and 13–17 percent in the coefficient of SB. The drop in the SB coefficient (22–25 percent) is, however, much greater than in the SI coefficient (3–6 percent) when the personal-background factors are included. Moreover, the total decline in the SB coefficient (35 percent) is nearly three times the decline in the SI coefficient (12 percent). These results were to be expected. Education before service is more highly correlated to personal-background factors and ability than education after service and more likely to be biased downward because of the absence of a measure of school quality.[16] This is why

[16] The argument concerning the effects of the left-out variable of schooling quality is slightly more complicated than that outlined in the Introduction because of the presence of *two* schooling variables. Considering only differences in the quality of schooling before military service and assuming that they are uncorrelated with both SB and SI leads to the conclusion that the introduction of the AFQT variable will bias the estimated SB coefficient downward (due to the assumed positive correlation of quality of schooling, Q, with AFQT and the observed positive correlation of AFQT with SB). The estimated coefficient of SI would remain unbiased provided that it really was uncorrelated with SB, AFQT, and the unobserved Q. The correlation of SI with AFQT is effectively zero ($r^2 = .007$), but it does have a nonnegligible negative correlation with SB. This leads also to a downward but smaller bias in the coefficient of SI; the ratio of the two biases (in the coefficient of SI relative to the bias in the coefficient of *SB*) is equal to $b_{\mathrm{SB,SI}}$, which is about .3 in our data (see Appendix).

TABLE 3

REGRESSION EQUATIONS WITH LOG INCOME AS DEPENDENT VARIABLE

REGRESSION No.	COEFFICIENT (STANDARD ERROR) OF					OTHER VARIABLES IN EQUATION*	R^2
	Color	SB	SI	ST	AFQT		
1	.2548 (.0472)	.0502 (.0042)	.0528 (.00702)	...	...	Age, AMS	.1666
2	.2225 (.0479)	.0418 (.0049)	.0475 (.0072)	...	.00154 (.00045)	Age, AMS	.1732
3	.1904 (.0473)	.0379 (.0045)	.0496 (.0070)	...	...	Age, AMS, fa. stat., reg. bef.	2129
4	.1714 (.0479)	.0328 (.0050)	.0462 (.0071)	...	.00105 (.00045)	Age, AMS, fa. stat., reg. bef.	2159
5	.2544 (.0471)	...	...	.0508 (.0039)	...	Age, AMS	.1665
6	.22245 (.04793)	...	...	.0433 (.0044)	.00150 (.00045)	Age, AMS	.1729
7	.1907 (.0473)	...	...	.0408 (.0041)	...	Age, AMS, fa. stat., reg. bef.	.2115
8	.1732 (.0479)	...	...	.0365 (.0046)	.00097 (.00044)	Age, AMS, fa. stat., reg. bef.	.2141
9	.1335 (.0487)	...	...	...	.00252 (.00041)	Age, AMS, fa. stat., reg. bef.	.1794
10	.1742 (.0488)	...	...	...	...	Age, AMS, fa. stat., reg. bef.	.1578
11	.2052 (.0456)	.0320 (.0048)	.0445 (.0068)	...	.00115 (.00045)	Age, fa. stat., reg. bef., reg. now, curr. exp., AMS	.2979
12	.2240 (.0449)	.0372 (.0046)	.0468 (.0068)	...	.00129 (.00043)	Age, reg. now, curr. exp., AMS	.2851

NOTE.—See table 1 for definitions.
* Variable groups are denoted as follows: fa. stat. = fa. occ. and fa. schooling; reg. bef. = ROS, POC, POS; reg. now = RNS, RNW, SMSA; curr. exp. = NM, LCJ.

TABLE 4

ESTIMATED BIAS IN SCHOOLING COEFFICIENTS

PROPORTIONAL CHANGE IN THE COEFFICIENT OF			
SB	*SI*	*ST*	VARIABLES ADDED
.17	.10	.15	AFQT
.25	.06	.20	Fa. stat., reg. bef.
.35	.12	.28	AFQT, fa. stat., reg. bef.
.13	.07	.11	AFQT added after fa. stat., reg. bef.
.22	.03	.16	Fa. stat., reg. bef. added after AFQT

we prefer the coefficient of SI as an estimate of the effect of an incremental change in schooling. But, even using total schooling, the decline (28 percent) in the education coefficient is not all that great. Of the total decline in the coefficient for ST, 11–15 percent can be attributed to the introduction of the AFQT variable; the rest is due to parental background and region and size of city of origin, variables that are likely to be closely related to the omitted school-quality dimension.

For analysis of the contribution of education to economic growth, the most appropriate estimate is that given by the coefficient of incremental schooling in regression 4, a regression which includes background and ability measures but does not contain any later current experience and success variables. The value of this coefficient is .0462, and we have already observed that this is only 12 percent lower than the .0528 given by the first regression, which includes no background or ability measures. Thus, while the usual estimates of the contribution of education may be biased upward due to the omission of such variables, this bias does not appear to be large and is much smaller than the 40 percent originally suggested by Denison (1962).

Education does, of course, make some significant independent contribution to the explanation of income, as may be seen by comparing regression 9 with regression 4. And comparison of regressions 4 and 8 indicates that even though the two schooling variables are acquired at different times and under different circumstances, their effects on income are similar. In fact, the difference between the two schooling coefficients in regression 4 is not statistically significant at the conventional 5 percent level, although this difference is significant at about the 8 percent level (which the computed $F = 3.2$ satisfies). We would expect the difference to be more highly significant with a larger sample, and we also would expect the inclusion of a school-quality measure to eliminate it completely.

Finally, recall that our model postulates the dependency of postservice schooling, length of service, and performance of the AFQT on schooling before service. It might be argued, quite apart from SB's sensitivity to the omission of school quality from regression 4, that the correct comparison of the effects of SB and SI on income would take account of SB's indirect contribution to income through SI, AMS, and AFQT, and that if we made this comparison we would discover SB's effect on income to be greater than SI's. As it turns out, the hypothesis that excluding the paths of SB through SI, AMS, and AFQT to income stacks the cards in favor of the coefficient of SI is incorrect. For, taking into account SB's effects on SI, AMS, and AFQT, we obtain a total coefficient of .0319 for SB's effect on income, which is slightly less than the direct coefficient of .0328 for SB in regression 4.[17] The explanation for this is, of course, that there is a negative relationship between SI and SB; the further a man goes in school before service, the less he needs to go after leaving service, and the less he *can* go after leaving service.

AFQT.—Given the current resurgence of interest in the role of intelligence in the achievement process and the common use of the AFQT as a measure of IQ, the performance of this variable is more modest than we had expected. While it is relatively highly intercorrelated with schooling before military service and with the other personal-background variables, its own *net* contribution to the explanation of the variance in the income of individuals is very small. For example, introducing AFQT into regression 2 increases the R^2 by only .007 (relative to regression 1). Introducing it into regression 4 would only increase the R^2 by .003 (relative to regression 3). Even if one attributed all of the joint schooling-intelligence effects (including schooling before service and hence before the date of these tests) to the AFQT variable, one would raise its contribution to the R^2 to only .022 (regression 9 vs. regression 10).[18]

[17] Given the causal ordering embodied in equations (1)–(5), the total effect of SB on income net of all prior factors can be decomposed into a direct contribution (given in regression 4) and an indirect contribution, obtained by computing the contribution of SB to income *through* SI, AMS, and AFQT. Decompositions of this sort are part of the results of the method of path regressions or path coefficients (Duncan 1966). Or, as we demonstrate later, they also can be derived by application of the excluded-variables formula given in the introduction. Dividing the variables listed in (1)–(5) into SI (S), SB (B), AFQT (T), AMS (M), other (O), and calling income y, we can think of the total effect of SB on y as given by $b_{yB.O}$. The decomposition of this coefficient implied by our model is given by the following expression: $b_{yB.O} = b_{yB.TOMS} + b_{SB.TOM} \cdot b_{yS.TOMB} + b_{MB.TO} \cdot (b_{yM.TOBS} + b_{yS.TOMB} \cdot b_{SM.TOB}) + b_{TB.O} [b_{yT.MOBS} + b_{ST.MOB} \cdot b_{yS.TOMB} + (b_{yM.TOBS} + b_{yS.TOMB} \cdot b_{SM.TOB}) b_{MT.BO}]$. The first term on the right-hand side gives the net, direct effect of SB on income, and is equal to .0328 as indicated by regression 4. Each of the other terms on the right-hand side gives the indirect contribution of SB to income through SI, AMS, and AFQT, respectively. The sum of these indirect effects is —.0009. Therefore $b_{yB.O} = .0319$.

[18] Another way to look at the relation between income and AFQT is to decompose the correlation between them into components, using path coefficients. Doing so is equivalent to a repeated application of the excluded-variables formula, with all the variables scaled to have mean zero and a unit standard deviation. The advantage of such a decomposition is that it is additive, whereas a decomposition in terms of changes in R^2 is not. This decomposition does presuppose a causal ordering, for which we shall

One final consideration is of interest here in discussing the role of AFQT in determining earnings. The literature on the "residual factor" and economic growth (Denison 1964, for example) has frequently involved adjusting, rather arbitrarily, observed income distributions for variation presumed due to a genetic substrate. Relevant variation on this substrate is usually held to be measured best by variation in performances on intelligence tests and to some extent by variation in parental social status. Since, in this paper, we have measures of these variables, we are in a position to question how much they contribute, taken together, to the explanation of income differences. We can then use our estimate as an upper bound for the (presently) measurable effects of this part of genetic heredity on income. This, in turn, provides us with another way of looking at the bias in education due to omitting intelligence and parental status.

With our data, adding AFQT and fa. stat. to this list of regressors in a regression of income on age, color, and reg. bef. increases the R^2 by only .052; while adding *color*, AFQT, and fa. stat. to the list of regressors in a regression of income on age and reg. bef. increases the R^2 by only .061. The increment in explained variance due to these "heredity"-associated variables is thus only about a fifth of the total "explainable" variance in income (the maximal R^2 in predicting income is given in table 5 as .31). And this makes no allowance for the effects of quality of schooling and discrimination that are confounded with color, regional origin, and parental-status variables. The *measurable* potential effects of genetic diversity on income, in the sense described above, appear to be much smaller than is usually implied in debates on this subject. And it follows, therefore,

---

use equations (1)–(5) (our model). Dividing and labeling our variables into AFQT (T), SI (S), AMS (M), and other (O), calling income y, and using the left-out variables formula repeatedly, we get the path coefficients decomposition of: $r_{yT} = \beta_{yT\cdot MSO} + \beta_{yM\cdot SOT} \cdot \beta_{MT\cdot O} + \beta_{yS\cdot TOM} (\beta_{ST\cdot MO} + \beta_{SM\cdot OT} \cdot \beta_{MT\cdot O}) + r_{OT} [\beta_{yO\cdot TSM} + \beta_{yS\cdot TOM} (\beta_{SO\cdot TM} + \beta_{SM\cdot TO} \cdot \beta_{MO\cdot T}) + \beta_{yM\cdot SOT} \cdot \beta_{MO\cdot T}]$, where the "β's" are the standardized partial regression coefficients and $\beta_{ij} = r_{ij}$. The first term of the right-hand side is the net effect of T on y, the second and third terms together give the effect of T via M and S, and the last term gives the effect of T which is "due to" or "joint with" the other variables (O).

The decomposition of r_{yT} via path coefficients yields the conclusion that more than half of the observed simple correlation between income and AFQT is "due to" or "joint with" the logically prior variables of color, fa. stat., reg. bef., SB, and age. The estimates for equations (1)–(5) of our model imply that $r_{\text{earnings'AFQT}} = .2355 =$ (.0657 net) + (.0361 through SI and AMS) + (.1337 joint with, or due to, other factors) = (.102 attributable to AFQT net of prior factors) + (.133 attributable to correlations between AFQT and prior factors). In terms of the model used here, over half of the initial correlation between income and AFQT is explained by factors in the model which are prior to AFQT. And, even if schooling before service and the background variables were not taken as predetermined with respect to AFQT, over half of the zero-order correlation still would be allocated to *joint* influence with these other independent variables. Note also that $r = .1$ (the approximate role of AFQT net of prior factors) is equivalent to $r^2 = .01$.

TABLE 5

REGRESSION OF LOG INCOME ON ALL AVAILABLE RELEVANT VARIABLES

Variable	Coefficient	*t*-Ratio
Age	.0126	(4.3)
Color	.1970	(4.4)
FO	.0016	(3.2)
FS	−.0038	(−1.2)
POC	.0325	(1.4)
POS	.0971	(2.4)
ROS	−.0238	(−0.7)
SB	.0244	(4.9)
AFQT	.00095	(2.2)
SI	.0352	(4.8)
RNS	−.0751	(−2.3)
RNW	.1173	(4.5)
SMSA	.1365	(6.7)
LCJ	.0013	(5.7)
NM	−.1496	(−5.7)
LOSES	.0804	(5.3)
AMS	−.0011	(2.0)
(Constant)	3.6483	...
(R^2)	.3114	...

that since most of the effects of heredity are indirect, there is little bias in an estimate of a schooling coefficient that does not take heredity into account. Heredity will affect the distribution of schooling attained, but the estimated schooling coefficient measures its contribution correctly, whatever the source of a change in schooling.

Additional details and relationships.—By including almost all of the variables available to us (see table 5) we can account for about a third of the observed variance in the logarithm of income. This is comparable with the results of other studies based on observations of individuals (for example, Hanoch 1967), but it is clear that the bulk of the variance in individual income is not accounted for by our equations, even when using a rather long list of variables.

We may use the regression displayed in table 5 to provide some more information on our results. Since the dependent variable is the logarithm of income, these coefficients (times 100) give the percentage effect of a unit change in the respective variables on income. The more interesting findings here are: (1) The nonsignificance of the father's schooling variable in the presence of father's occupational SES score. This is also true in most of the other regressions. (2) The relative importance of current location (being in an SMSA and in the West). (3) The rather surprising strong negative effect of not having married. And (4), the negative effect of time spent in the military and the implied positive effect of potential experience in the labor force on income.[19]

[19] Since, except for constants, potential experience = age − SB − SI − AMS/12, in a regression that already contains age, SB, and SI, its coefficient is given by *the*

In table 6 we gather some results on the interrelationships between the other variables in our model. Among the more interesting of these are the highly significant (and rather large) effects of region, color, and schooling before service on AFQT, and the barely significant (and minor) effects of the parental-status variables. This is hardly consistent with Jensen's (1969) treatment of variance in AFQT scores as primarily heritable. The other interesting fact is that using occupational status rather than income as the dependent variable gives similar results: significance for the schooling variables, and only marginal importance for parental status and AFQT.

IV. Errors in the AFQT Variable and Other Extensions

In this section we reestimate the income-generating equation assuming that AFQT is subject to random errors to get an idea of the results we might obtain with a better measure of ability.[20] To do so, we shall have to revise somewhat the model sketched out in the previous section and introduce an unobservable ability or achievement variable. Since we have no direct knowledge of the errors in the AFQT, the discussion which follows is an essay: We assume the AFQT measures adult ability with random errors. We specify a model for the explanation of earnings that takes into account these random errors. If these assumptions are correct then the results of our reestimation also are correct.

Let us postulate the following simple linear model, summarized in table 7 and diagrammed in figure 2, where the time subscripts 0, 1, 2, represent measurements taken before the start of formal schooling (approximately age 6), before entering military service (approximately age 18), and at the

---

negative of the coefficient of AMS times 12. In this case, it comes out to .0132, and this is also the predicted coefficient for age. Since the actual coefficient for age is .0126, the two are consistent and support the interpretation that both calendar age and time spent in military service influence income via their effect on "experience." Another way of testing this is to constrain the coefficient of age to equal 12 times minus the coefficient of AMS. The computed F-statistics for such constrained versions of regressions 1 and 4 are 3.7 and 2.8, respectively, indicating that the data are consistent with the validity of such a constraint at the conventional 5 percent significance level (the critical F is 3.8). For regression 4, the constrained version implies that a year of experience is worth a 2.3 percent increase in income, on the average, and that holding "experience" (but not age) constant leads to estimated 7.3 and 7.8 percent increases in income per year of schooling, for pre- and postservice schooling, respectively.

[20] The sources of random error in the AFQT are presumed to be grouping, reliability, aggregation, and left-out components of ability. Grouping would create random errors if the actual scores are distributed evenly within intervals. Reliability errors, though doubtless present, probably are minor because of the grouping procedure. Aggregation, in the sense of using a global index instead of its separate components, could create random differences between the ability index which maximally predicts income and the AFQT index. Left-out components of ability also could differ randomly from the AFQT. *Nonrandom* errors could be due to the differential distribution by parental SES of test-wiseness, motivation, and experience with the kinds of material the test uses (culture-boundedness of the test). We are unable to adjust for nonrandom errors.

TABLE 6

INTERRELATIONS BETWEEN DETERMINANTS OF INCOME
(t-RATIOS)

DEPENDENT VARIABLE	INDEPENDENT VARIABLES											
	Color	FO	FE	POC	POS	ROS	SB	SI	AFQT	AGE	AMS	R^2
SB	*	8	6	4	*	5	...	...	...	...	...	.152
SB	—4	6	4	3	*	3	...	...	17	...	...	.289
AFQT	5	5	5	2	*	6	...	...	...	...	...	.139
AFQT	6	2	3	*	*	4	17	...	...	...	...	.271
AMS	*	*	2	4	*	*	5	...	*	9	...	.083
SI	*	3	4	*	*	3	11	...	8	4	4	.130
NM	*	—2	*	*	*	—3	*	*	*	—9	*	.073
LCJ	*	*	*	*	*	2	3	3	*	18	4	.208
RNS	3	*	*	*	*	46	*	*	*	*	*	.625
RNW	*	*	*	—4	*	—5	3	2	*	*	3	.051
SMSA	—3	*	*	12	6	—4	2	3	*	*	2	.145
LOSES	*	2	3	5	*	*	12	10	3	5	*	.290

* In the equation but estimated t-ratio less than 2.

TABLE 7

SCHEMATIC MODEL OF INTERRELATIONSHIPS BETWEEN SCHOOLING, ABILITY, AND INCOME

(1)	$G_0 = a_1B + a_2H$
(2)	$T_0 = G_0 + t_0$
(3)	$S_1 = b_1B + b_2H + e$
(4)	$G_1 = G_0 + \gamma S_1$
(5)	$T_1 = G_1 + t_1$
(6)	$S_2 - S_1 = c_1S_1 + c_2B + w$
(7)	$G_2 = G_1 + \gamma(S_2 - S_1)$
(8)	$I_2 = \beta G_2 + u$

NOTE.—G = achievement, or ability to earn income, unobservable directly; B = background factors including social class of parents (fa. stat.) and location of adolescence (reg. bef.); H = heredity, or genotype, unmeasured; T = test score, purporting to measure G (T_1 = AFQT); S = schooling (S_1 = SB, S_2 = ST, $S_2 - S_1$ = SI); I income (LINC); e,t,w,u = random forces, uncorrelated with each other and with the causally prior exogenous variables of the system, that is, the t's are assumed to be uncorrelated with each other and with all the other variables in the model except the T's; e is assumed to be uncorrelated with B and H, w also with S_1 and u also with $S_2 - S_1$.

time of the survey (age in 1964), respectively. The symbols are intended to be mnemonic; random disturbances appear only in equations with observable dependent variables. We also assume that all variables are measured around their mean levels, obviating the need for constants in these equations. Basically we have an unobservable ability or achievement (or human-capital) variable, which is augmented by schooling, and the stock of which (G) is estimable (subject to error) via test scores (T). We assume in this model that all of the influence of class and heredity is indirect, via the early-achievement variable. Note that we assume equal contributions of a unit change in SI ($S_2 - S_1$) to achievement and of a unit change in S_1 (SB), and we also assume that the schooling increment is uncorrelated with the error in observed test scores (t_1) and with that part of heredity (H) not already reflected in S_1 or correlated with B. These assumptions (equality of the coefficients of S_1 and S_2 and no correlation between $S_2 - S_1$ with t_1 and H net of S_1 and B) are the important identifying restrictions in our model.

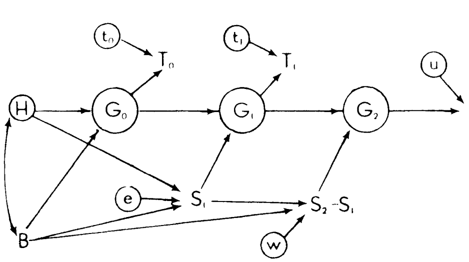

FIG. 2.—Revised causal model of income determination. Circled items = unobservables.

The present data are not sufficient to estimate this model in its entirety. We have no measures of A, T_0, and H. Yet, we can mesh our data with this model in a way which may allow us to escape the effect of errors in AFQT. Substituting equations (4) and (1) into (5) (see table 7) gives

$$T_1 = \gamma S_1 + a_1 B + a_2 H + t_1, \tag{9}$$

and substituting equations (7) and (5) into (8) results in

$$\begin{aligned} I_2 &= \beta[\gamma(S_2 - S_1) + (T_1 - t_1)] + u \\ &= \beta\gamma(S_2 - S_1) + \beta T_1 + u - \beta t_1. \end{aligned} \tag{10}$$

Since the error (t_1) in T_1 is not observable, we have again an errors-in-variables problem (or a simultaneity problem in the sense of a nonzero correlation of T_1 with the new disturbance $u - \beta t_1$). To solve this problem we can use the observable predetermined variables (S_1 and B) not appearing in equation (10) in a two-stage instrumental-variables procedure. In the first stage we estimate equation (9), ignoring the unavailable H variable and get a predicted value of T_1, $\hat{T}_1$ (AFQT *Hat*), based on the observed predetermined variables. This predicted value replaces T in equation (10). In the second stage, we regress I_2 (LINC) on $S_2 - S_1$ (SI) and $\hat{T}_1$ (AFQT *Hat*) to estimate $\beta\gamma$ and β.[21] This procedure solves the problem of error in T_1, assuming that our model is correctly specified, but does little about the effect of the omitted variable H (except for its influence via S_1). Here we have to count on the presumed relative independence of the increment in schooling from H, net of their joint relationship with S_1 and the variables contained in B.

Table 8 summarizes the two-stage calculations. Comparing regressions 13 and 14 with 4, 11, and 12 (table 3), we note that the estimated coefficient of incremental schooling does not decrease. Constraining the model so that background factors and schooling before service work through the unobserved achievement variable gives the same results for the remaining schooling variable as the unconstrained regressions. Allowing for direct effects of measured AFQT, schooling before service, and social background improves the fit only marginally (regressions 4 vs. 13 or 11 vs. 14). Thus, the approach taken here suggests that our initial estimate of the schooling effect on income is robust with respect to the presence of (random) measurement errors in AFQT. Moreover, the comparable levels of fit in the error model and the unconstrained regressions support the model outlined in table 7.

Considering next the AFQT *Hat* variable, note that its coefficient in regressions 13 and 14 is much larger and more highly significant than those

[21] Note that color, age, and AMS also are included because they are assumed to have an independent effect on income. As in Section I, AMS also could be entered explicitly into the model. To do so, however, would not change the results of interest and would detract from the clarity of the model's central features.

TABLE 8

TWO-STAGE AND OTHER REGRESSIONS

Regression No.	Coefficient (Standard Error) of: Color	SI	AFQT *Hat**	AFQT	Other Variables in Equation	R^2
Dependent variable = Log Income:						
13	.0351 (.0494)	.0504 (.0069)	.01051 (.00078)	...	Age, AMS	.1876
14	.0730 (.0468)	.0483 (.0065)	.00889 (.00078)	...	Age, reg. now, curr. exp., AMS	.2855
15	.1982 (.0458)	.0331 (.0067)	...	.00298 (.00038)	Age, reg. now, curr. exp., AMS	.2526
Dependent variable = Log Occupation SES:						
16	—.3979 (.0815)	.1320 (.0114)	.02554 (.00129)	...	Age, AMS	.2636
17	—.3517 (.0809)	.1277 (.0113)	.02626 (.00134)	...	Age, reg. now, curr. exp., AMS	.2880
18	.0157 (.0831)	.0843 (.0121)	...	.00809 (.00069)	Age, reg. now, curr. exp., AMS	.1779
19	.1014 (.0787)	.1151 (.0117)	...	.00253 (.00073)	Age, reg. now, curr. exp., fa. stat., reg. bef., SB, AMS	.3034

* The equation used to define this variable is:

$$\text{AFQT } Hat = -19 + \underset{(2.83)}{17.85}\ \text{color} + \underset{(.0309)}{.0735}\ \text{FO} + \underset{(.1481)}{.5505}\ \text{FS} + \underset{(.262)}{4.434}\ \text{SB} - \underset{(1.282)}{5.472}\ \text{ROS}$$

with $R^2 = .271$.

for the original AFQT measure (table 3). "Purging" AFQT of errors thus increases its contribution to income, even though it does not modify the estimated contribution of education. Observe also that a bound can be set on the effect of ignoring the H variable in equations (9) and (10) derived from the error model. In particular, the gain in predicting income with the estimate of error-free AFQT more than offsets the loss due to lack of a measure of the direct influence of H. That is, the *ignored* systematic part of ability, the part of heredity that is uncorrelated with the variables defining AFQT *Hat*, has a smaller variance than the variance of error in observed AFQT, since the R^2 in regression 15 is greater than in regression 14.[22]

The only novel result in table 8 pertains to the coefficient of the white-black dummy variable in the presence of the AFQT *Hat* variable. It is insignificant now, indicating that all of the color effects were captured by AFQT *Hat*. Taken at face value, this result implies that discrimination against blacks does not affect white-black differences in income once person-to-person differences in ability and achievement are adjusted for random-measurement error. This outcome could not have been forecast on the basis of any previous literature. Since the number of blacks in the sample is very small, the result cannot be taken for anything more than an invitation to further work along the above lines.

Having set up the model outlined in table 7, we could add additional equations connecting other indicators of success, such as occupational SES, to the unobserved G_2 (achievement in 1964) variable. Such an extension is presented in figure 3. It implies a proportionality of coefficients in

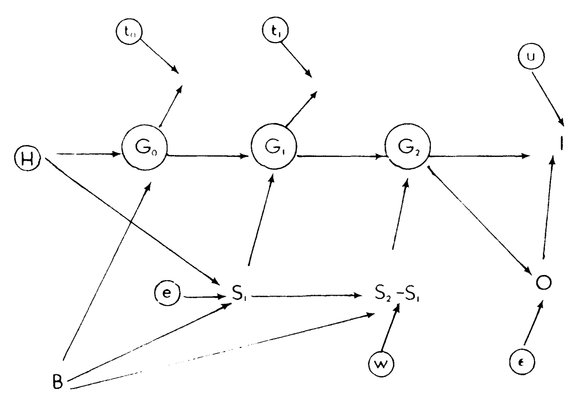

FIG. 3.—Extension of revised model of income determination to include current success variables. $O =$ occupational SES or other measures of current success.

[22] Let $G = S + H$, and H be defined so as to be uncorrelated with S. Then using the observed T as a variable implies leaving out from the regression $-\beta t$, the error of measurement in T. Using $\hat{T} = S$ implies the leaving out of βH. The latter causes a smaller reduction in the explained variance than the former.

equations with different success measures as dependent variables that could be used in another estimation round to get a constrained but more efficient set of estimators for the coefficients of the independent variables (see Zellner 1970; Goldberger 1971). Since we are primarily interested in the effect of these variables on earnings, we have not pursued this further here. We doubt, however, that it is reasonable to impose such a proportionality assumption across the coefficients of *all* the variables in our data. It would not be surprising if variables such as marital status or color have different relative effects on income and occupational status. The last set of regressions in table 8 points up the problem. With log occupational SES as the dependent variable, the coefficients of incremental schooling and AFQT (or AFQT *Hat*) are roughly proportional to those with log income as the dependent variable. Comparing regression 16 with regression 13, the coefficients stand in the ratio 2.6 and 2.4 for the SI and AFQT *Hat* variables, respectively. In regressions 18 and 15 the ratios are 2.5 and 2.7, respectively. This is not too bad. But the color coefficients stand in a ratio of .5 for the second comparison, and they are actually of opposite sign for the first. Thus, proportionality across *all* the coefficients is not apparent in the data and is also unlikely for such variables as color and marital status. Procedures are available for dealing with these more complicated models but we do not pursue this topic further here.[23]

V. Discussion and Summary

We have tried to compare our results with those of other similar studies but without too much success. None of the other studies uses an incremental-schooling variable, a distinction on which much of our results rest. Also, such studies tend to treat years of school as the conceptually right and error-free measure of educational attainment, a position that is hardly tenable in light of the extreme diversity of the education system in the United States.

Duncan's (1968) major study uses the same basic data set as we do, but defines the subsample of interest as white males ages 25–34, includes both veterans and nonveterans, and introduces early intelligence and

[23] See Hauser and Goldberger (1970) for more details. To rationalize these facts we must assume that there also is some direct effect of variables such as color and marital status on income outside and beyond their contribution to the unobserved achievement variables. In terms of figure 3, color would be contained in B but might have additional independent and different paths to I_2 and 0. Similarly, marital status could be interposed between G_2 and I_2 and 0, having differential effects on the latter two variables. In general, if income and occupational success depend not only on cognitive achievement (AFQT, schooling, and related measures) but also on "motivation" where motivation may be a function of previous achievement, some of the same background variables, and other random variables, then only smaller subsets of coefficients are subject to proportionality constraints.

number of siblings from other sources. Instead of actual income, he uses expected income. For this sample so defined, the coefficient of total schooling declines about 31 percent when parental status, number of siblings, and early intelligence variables are introduced into a regression with expected income as the dependent variable. We cannot, however, be sure that this difference between Duncan's study and ours is due to the difference in populations sampled, because expected and actual income are imperfectly correlated (in our sample the correlation between the logarithms of these two variables is about .7), and his results do not control for differences in labor-force participation or the effects of different regions of origin nor do they allow for the correlation of the parental-status variables with the left-out school-quality variable.[24]

Hansen, Weisbrod, and Scanlon (1970) analyze a sample of 17–25-year-old men rejected by the Selective Service System because of low AFQT scores, and conclude that schooling is a relatively unimportant income determinant. In their data, the education coefficient drops about 50 percent when the AFQT variable is introduced into the regression of income on age, color, size of family of origin, whether the family of origin was intact, and education. The education coefficient drops even further when such current-success variables as job training and marital status are added. Their sample is peculiar in that it concentrates on the very young and on blacks (about half of their sample is nonwhite vs. 9 percent in our subsample). It is well known that schooling-income differentials are rather low at the beginning of the labor-force experience and that there is little evidence for a strong schooling-income relationship among blacks (see Hanoch 1967). Both facts could help to explain the differences between these two samples. Moreover, the correlation between AFQT and the omitted variable of school quality is likely to be higher for this population than for higher-ability groups, so that including AFQT in the regression overstates the bias in the education coefficient due to neglecting ability. For these reasons, then, we are not ready to conclude that using a larger number of low-ability men than was available to us within our own sample would alter our estimate of the bias in the education coefficient due to omitting ability. All of these considerations do remind us again, though, that we cannot take our sample as representative of the entire labor force.[25]

[24] In addition to collating information from several samples, Duncan's study also uses correlations between the AFQT and other variables based on an extrapolation from the veterans' subsample to the total sample. The use of these adjusted correlations would seem to partly explain the discrepancy between our own results and those implied by Duncan's data. Although the assumptions which underlie the adjusted correlations appear reasonable, they do remain open to question.

[25] Several studies of high-SES samples have also shown a relatively small bias in the schooling coefficient due to left-out ability variables (Ashenfelter and Mooney 1968; Weisbrod and Karpoff 1968; Rogers 1969; Taubman and Wales 1970). This last study also can be interpreted so as to show a rather significant effect of variation in the quality of college schooling.

Our findings support the economic and statistical significance of schooling in the explanation of observed differences in income. They also point out the relatively low independent contribution of measured ability (AFQT scores). Holding age, father's status, region of origin, length of military service, and the AFQT score constant, an additional year of schooling would add about 4.6 percent to income in our sample. At the same time a 10 percent improvement in the AFQT score would only add about 1 percent to income.

Using a "clean" schooling variable, incremental schooling, we concluded that the bias in its estimated coefficient due to the omitted ability dimension is not very large (on the order of 10 percent). The earlier (before military service) schooling coefficient falls more, but we interpret this to be the consequence of the interrelationship between test scores and father's status variables with the other important omitted variable—the quality of schooling. Unfortunately, given the nature of our sample, restricted as it is by the selectivity inherent in being a veteran and the relatively young (under 35) age of males included, these results cannot be taken as representative for all males. Nevertheless, this is one of the largest samples ever brought to bear on this problem and we would expect it to survive extension to a more complete population.

Our results also throw doubt on the asserted role of genetic forces in the determination of income. If AFQT is a good measure of IQ and if IQ is largely inherited, then the direct contribution of heredity to current income is minute. Its indirect effect also is not very large. Of course, the AFQT scores may be full of error and heredity may be very important, but then previous conclusions about the importance of heredity are also in doubt since they were drawn on the basis of similar data.

Appendix

All of the formulae used in the text are repeated variations on the "left-out variable" formula.[26] Let the true equation be

$$y = \beta_1 x_1 + \beta_2 x_2 + e,$$

where all the variables are measured around their means (and hence we ignore constant terms) and e is a random variable uncorrelated with x_1 and x_2.

Now, consider the least-squares coefficient of y on x_1 *alone*:

$$\begin{aligned} b_{y1} &= \Sigma x_1 y / \Sigma x_1^2 = \Sigma x_1 (\beta_1 x_1 + \beta_2 x_2 + e)/\Sigma x_1^2 \\ &= \beta_1 + \beta_2 \, \Sigma x_1 x_2 / \Sigma x_1^2 + \Sigma x_1 e / \Sigma x_1^2. \end{aligned}$$

Since the expectation of the last term is zero, we can write

[26] These formulas are given, in a different context, in appendix C of Griliches and Ringstad (1971). See Yule and Kendall (1950, chap. 12) for the notation used here.

$$E(b_{y1}) = \beta_1 + \beta_2 b_{21},$$

where

$$b_{21} = \Sigma x_1 x_2 / \Sigma x_1^2$$

is the (auxillary) least-squares coefficient of the left-out variable x_2 on the included x_1.

Moreover, if e were to refer to the computed least-squares residuals, $\Sigma x_1 e$ would equal zero by construction. Hence, the same formula also holds as an *identity* between computed least-squares coefficients of different order. That is,

$$b_{y1} = b_{y1\cdot 2} + b_{y2\cdot 1} b_{21}.$$

This same formula, with a suitable change in notation, applies also to higher-order coefficients:

$$b_{y1\cdot 2} = b_{y1\cdot 23} + b_{y3\cdot 12} \cdot b_{3\cdot 12}.$$

In what follows we shall assume that we are talking either about least-squares coefficients or about population parameters, and we will not carry expectation signs along. The discussion could be made somewhat more rigorous by inserting the plim (probability limit) notation at appropriate places.

The model we deal with can be written as

$$\begin{aligned} y &= \beta_1 E + \beta_2 T + e \\ &= \beta_1 S + \beta_2 T + \beta_1 Q + e, \end{aligned}$$

where $E = S + Q$ is education, S is quantity of schooling, Q is quality of schooling, and T is a measure of ability (here assumed to be error-free); Q is uncorrelated with S but is correlated with T. Then, estimating the equation with both T and Q out, leads to

$$b_{yS} = \beta_1 + \beta_2 b_{TS} + \beta_1 b_{QS} = \beta_1 \beta_2 b_{TS}$$

since $b_{QS} = 0$ by assumption. Including T in the equation also gives

$$b_{yS\cdot T} = \beta_1 + \beta_1 b_{QS\cdot T}.$$

Now, while b_{QS} is zero, $b_{QS\cdot T}$ need not be zero. Given our assumptions we can write,

$$b_{QS} = b_{QS\cdot T} + b_{QT\cdot S}\, b_{TS} = 0,$$

which implies that

$$b_{QS\cdot T} = -b_{QT\cdot S}\, b_{TS} < 0,$$

since both $b_{QT\cdot S}$, the partial relationship of school quality to test scores, and b_{TS}, the relationship between test scores and levels of schooling, are expected to be positive. We also have

$$b_{QT} = b_{QT\cdot S} + b_{QS\cdot T} \cdot b_{ST}.$$

Substituting the formula for $b_{QT\cdot S}$ into the formula for b_{QT}, we get

$$b_{QT} = b_{QT\cdot S} - b_{QT\cdot S} b_{TS} \cdot b_{ST}.$$

Solving for $b_{QT\cdot S}$ and remembering that $b_{TS}b_{ST} = r^2_{ST}$ gives

$$b_{QT\cdot S} = b_{QT}/(1 - r^2_{TS}),$$

which then gives

$$b_{QS\cdot T} = -b_{QT} \cdot b_{TS}/(1 - r^2_{TS}).$$

The algebra gets a bit more complicated when S is divided into two components, which for notational convenience will be called B (before) and A (after) here. The model now is

$$y = \beta_1 B + \beta_1 A + \beta_2 T + \beta_1 Q + e.$$

Then

$$b_{yB\cdot AT} = \beta_1 + \beta_1 b_{QB\cdot AT}$$

and

$$b_{yA\cdot BT} = \beta_1 + \beta_1 b_{QA\cdot BT}.$$

Assume, as is approximately true in our sample, that A is uncorrelated with T. Since we have already assumed that Q is uncorrelated with both A and B, we have:

$$b_{QB\cdot A} = b_{QB\cdot AT} + b_{QT\cdot AB} \cdot b_{TB\cdot A} = 0;$$
$$b_{QA\cdot B} = b_{QA\cdot BT} + b_{QT\cdot AB} \cdot b_{TA\cdot B} = 0;$$

and hence

$$b_{QB\cdot AT} = -b_{QT\cdot AB} b_{TB\cdot A};$$
$$b_{QA\cdot BT} = -b_{QT\cdot AB} b_{TA\cdot B}.$$

Thus we can see immediately that the relative magnitude of the biases in the two schooling coefficients depends on the size of $b_{TB\cdot A}$ relative to $b_{TA\cdot B}$. Now because

$$b_{TA} = b_{TA\cdot B} + b_{TB\cdot A} b_{BA} = 0,$$

by assumption, we have

$$b_{TA\cdot B} = -b_{TB\cdot A} b_{BA},$$

which we can substitute in

$$b_{TB} = b_{TB\cdot A} + b_{TA\cdot B} \cdot b_{AB}$$

to yield

$$b_{TB\cdot A} = b_{TB}/(1 - b_{AB}b_{BA}) = b_{TB}/(1 - r^2_{AB})$$

and

$$b_{TA\cdot B} = -b_{TB}b_{BA}/(1 - r^2_{AB}).$$

Now, if A (schooling after service) were entirely uncorrelated with B (schooling before service), $b_{BA} = 0$, and its coefficient in the income-generating equation ($b_{yA \cdot BT}$) would be unbiased:

$$b_{QA \cdot BT} = -b_{QT \cdot AB} \cdot b_{TA \cdot B}$$
$$= +b_{QT \cdot AB} \cdot b_{TB} \cdot b_{BA}/(1 - r^2_{AB}),$$
$$= 0$$

while the coefficient of schooling before service in the income-generating equation would be biased downward. In our sample, however, b_{BA} is actually negative and on the order of -.3, implying that the coefficient of A is also biased downward, but only by about a third of the bias in the coefficient of B.

References

Ashenfelter, Orley, and Mooney, Joseph D. "Graduate Education, Ability and Earnings." *Rev. Econ. and Statis.* 50 (February 1968): 78–86.

Blau, Peter M., and Duncan, Otis Dudley. *The American Occupational Structure.* New York: Wiley, 1967.

Bowles, Samuel. "Schooling and Inequality from Generation to Generation." Unpublished paper presented at the eastern meetings of the Econometric Society, Tokyo, June 1970.

Davis, James W., Jr., and Dolbeare, Kenneth M. *Little Groups of Neighbors: The Selective Service System.* Chicago: Markham, 1968.

Denison, Edward F. *The Sources of Economic Growth in the United States and the Alternatives before Us.* Supplementary Paper no. 13. New York: Committee Econ. Development, 1962.

———. "Measuring the Contribution of Education." In *The Residual Factor and Economic Growth*, pp. 13–55, 77–100. Paris: Org. Econ. Co-operation and Development, 1964.

Duncan, Otis Dudley "A Socioeconomic Index for All Occupations," and "Properties and Characteristics of the Socioeconomic Index." In Albert J. Reiss, Otis Dudley Duncan, Paul K. Hatt, and Cecil C. North, *Occupation and Social Status*, chaps. 6 and 7, pp. 109–61. New York: Free Press, 1961.

———. "Path Analysis: Sociological Examples." *American J. Sociology* 72 (July 1966): 1–16.

———. "Ability and Achievement." *Eugenics Q.* 15 (March 1968): 1–11.

———. "Inheritance of Poverty or Inheritance of Race?" In *On Understanding Poverty*, edited by Daniel P. Moynihan, pp. 85–110. New York: Basic, 1969.

Goldberger, A. S. "Maximum-Likelihood Estimation of Regression Models Containing Unobservable Variables." Social Systems Research Institute, Unpublished paper, EME 7101, Univ. Wis., 1971.

Griliches, Zvi. "Notes on the Role of Education in Production Functions and Growth Accounting." In *Education, Income and Human Capital*, edited by W. Lee Hansen, pp. 71–127. Studies in Income and Wealth, vol. 35. New York: Nat. Bur. Econ. Res., 1970.

Griliches, Zvi, and Ringstad, V. *Economies of Scale and the Form of the Production.* Amsterdam: North-Holland, 1971.

Hanoch, Giora. "An Economic Analysis of Earnings and Schooling." *J. Human Resources* 2 (Summer 1967): 310–29.

Hansen, W. Lee; Weisbrod, Burton A.; and Scanlon, W. J. "Schooling and Earnings of Low Achievers." *A.E.R.* 60 (June 1970): 409–18.

Hauser, Robert M., and Goldberger, A. S. "The Treatment of Unobservable Variables in Path Analysis." Social Systems Research Institute Workshop Paper, EME 7030. Univ. Wis., 1970.

Jensen, Arthur R. "How Much Can We Boost IQ and Scholastic Achievement?" *Harvard Educ. Rev.* 39 (Winter 1969): 1–123.

Karpinos, Bernard D. "The Mental Qualification of American Youths for Military Service and Its Relationship to Educational Attainment." *Proc. of Social Statis. Sect. American Statis. Association, 1966.* Washington: ASA, 1966.

———. "Mental Test Failures." In *The Draft*, edited by Sol Tax, pp. 35–53. Chicago: Univ. Chicago Press, 1967.

Klassen, A. D. *Military Service in American Life since World War II: An Overview.* Report no. 117. Chicago: Nat. Opinion Res. Center, Univ. Chicago, 1966.

Mason, William. "Working Paper on the Socioeconomic Effects of Military Service." Unpublished paper, Univ. Chicago, 1968.

———. "On the Socioeconomic Effects of Military Service." Ph.D. dissertation, Univ. Chicago, 1970.

Rivera, Ramon J. "Sampling Procedures on the Military Manpower Surveys." Chicago: Nat. Opinion Res. Center, Univ. Chicago, 1965.

Rogers, Daniel C. "Private Rates of Return to Education in the U.S.: A Case Study." *Yale Econ. Essays* 9 (Spring 1969): 89–134.

Siegel, Paul M., and Hodge, Robert W. "A Causal Approach to the Study of Measurement Error." In *Methodology in Social Research*, edited by Hubert M. Blalock, Jr., and Ann B. Blalock, pp. 28–59. New York: McGraw-Hill, 1968.

Taubman, Paul, and Wales, Terence. "Net Returns to Education." In *Economics —a Half-Century of Research, 1920–1970; 50th Annual Report*, pp. 65–66. New York: Nat. Bur. Econ Res., 1970.

U.S. President's Task Force on Manpower Conservation. *One-third of a Nation: A Report on Young Men Found Unqualified for Military Service.* Washington: Government Printing Office, 1964.

Weisbrod, Burton A., and Karpoff, Peter. "Monetary Returns to College Education, Student Ability, and College Quality." *Rev. Econ. and Statis.* 50 (November 1968): 491–97.

Yule, G. Udny, and Kendall, M. G. *An Introduction to the Theory of Statistics.* 14th ed., rev. London: Griffin, 1950.

Zellner, Arnold. "Estimation of Regression Relationships Containing Unobservable Independent Variables." *Internat. Econ. Rev.* 11 (October 1970): 441–54.