

NBER WORKING PAPER SERIES

TEST SCORES AND SELF-SELECTION OF HIGHER  
EDUCATION: COLLEGE ATTENDANCE VERSUS  
COLLEGE COMPLETION

Steven Venti

David A. Wise

Working Paper No. 709

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge MA 02138

July 1981

We are indebted to David Ellwood, Win Fuller, Jerry Hausman, Charles Manski, and Michael Stoto for their comments. This work was supported by the Exxon Educational Foundation, contract number 300-78-0545 from the Office of Education; Department of Health, Education, and Welfare; and the National Science Foundation. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

Test Scores and Self-Selection of Higher Education:  
College Attendance Versus College Completion

ABSTRACT

As a companion paper to our work on students' application and colleges' admission decisions, we have estimated a joint discrete-continuous utility maximization model of college attendance and college completion. The paper is motivated by the possibility that test scores are poor predictors of who will succeed in college and thus may not promote optimal investment decisions and may indeed unjustly limit the educational opportunities of some youth. We find that: (1) College attendance decisions are strongly commensurate with college completion. Persons who are unlikely to attend college would be very likely to drop out of even their "first-choice" colleges, were they to attend. College human capital investment decisions are strongly mirrored by the likelihood that they will pay off. (2) Contrary to much of the recent criticism of the predictive validity of test scores, we find that their informational content is substantial. After controlling for high school class rank, for example, the probability of dropping out of the first-choice college varies greatly with SAT scores. (3) Individual self-selection, related to both measured and unmeasured attributes, is the dominant determinant of college attendance.

Steven Venti  
J.F.K. School of Government  
Harvard University  
Cambridge, Massachusetts 02138  
(617) 495-1456

Professor David A. Wise  
J.F.K. School of Government  
Harvard University  
Cambridge, Massachusetts 02138  
(617) 495-1178

TEST SCORES AND SELF-SELECTION OF HIGHER EDUCATION:  
COLLEGE ATTENDANCE VERSUS COLLEGE COMPLETION

by

Steven Venti and David A. Wise\*

Most economic studies of college attendance emphasize the returns to higher education as the motivation for an individual's choice to go to college, but ignore institutional constraints on possible choices. Critics of the use of test scores in the determination of college attendance on the other hand tend to emphasize the constraints on educational opportunities imposed by test scores and to ignore individual choice. Indeed scholastic aptitude tests have become an integral part of college application and admission procedures. Implicit in much of the recent criticism of them are two assumptions or claims: One is that the impact of test scores is primarily by way of their use by college admission officials to screen people out, and thus--by way of the constraints that they place on access to higher education--to limit occupational opportunities. The other is that the test scores are poor predictors of who will succeed in college and thus may not promote optimal investment decisions and may indeed unjustly limit the educational opportunities of some youth.<sup>1</sup> We shall in this paper address the

---

\* Research Fellow, J.F.K. School of Government, Harvard University; Stambaugh Professor of Political Economy, J.F.K. School of Government, Harvard University and Research Associate, NBER respectively.

1. See Nader Report by Nairn and Associates [1980], Slack and Porter [1980] and the congressional hearings on "Truth in Testing" legislation, U.S. Congress [1979].

second claim by comparing the determinants of college attendance on the one hand versus the determinants of college completion on the other, with special emphasis on the role of test scores.

With respect to the first assumption, we have found in previous research (Venti and Wise [1980]), that test scores (or more precisely, the attributes measured by them) bear a much stronger relationship to individual college application decisions than to college admission decisions. Indeed, most persons who don't apply to any college or university would have a high probability of admission at schools of average academic quality, if they were to apply. We have found that test scores in general do not narrowly proscribe by way of college admissions decisions the post-secondary educational opportunities of youth. Our previous research also suggests that student application decisions do not simply reflect expectations about college admission decisions. But it may be that to the extent that test scores determine individual human capital investment decisions, they do not provide an adequate or appropriate signal to students, and thus in large part may not be contributing to rational individual choice. It could be, for example, that persons who don't go to college--presumably in part because of test scores--would have been well advised to obtain higher education.

We shall address this latter issue as the first of two interrelated questions posed pursuant to the goals of this paper. We ask first whether individual college attendance decisions are consistent with the likelihood that a degree will be obtained. To motivate this questions, recall that simple models of investment in higher education suggest that individuals choose to attend college if the expected net return from college attendance is greater than the return from time spent by the individual in other ways.

The return from college can be thought of as the product of two components: the probability that a degree will be obtained, times the expected gain in future earnings (and non-monetary benefits) given a degree. Much has been written about the second component but little about the first.<sup>1</sup> The first, however, is of crucial importance because at least the occupational rewards to college education and probably the earnings gains as well come in large part with the degree. Whatever the determinants of college attendance, for attendance to be "rational", it should be the case that persons who are most likely to attend are also the most likely to obtain a degree and that those who are unlikely to attend would be unlikely to obtain a degree were they to go to college. Thus we investigate the relationship of test scores and other individual attributes to college attendance on the one hand, versus the relationship between these attributes and college completion on the other. We judge the extent to which individuals make "correct" college decisions by focusing on the relationship between college attendance decisions of youth and the ability of youth to benefit from college--as measured by the likelihood of graduation.<sup>2</sup> Within this context, we shall emphasize the relationship of test scores to these outcomes, and implicitly their informational value to students. Not only are universities likely to want to admit persons who will succeed, but students may be just as likely to use test scores to judge their own chances of success.

In addressing these issues we shall also emphasize student self-

---

1. Griliches [1974] elaborates on the distinction.

2. Of course, it is possible that the returns to college education are greater for the more academically able as well.

selection. Our model allows us not only to consider the extent of self-selection as explained by measured variables, but also to evaluate the extent of self-selection attributable to unmeasured individual attributes (the idea commonly denoted by self-selection in a statistical sense).

Then we ask a second, but related question. To the extent that persistence in college is a criterion for admission, what is the information value of test scores to admissions officials, and is their use of the scores consistent with this criterion? We pose the question in this way to provide a framework that allows us to compare our results with the claims of critics of the predictive validity of test scores. The question is essentially whether test scores add measurably to the information available to colleges, given a measure of high school performance, also an important determinant of college admissions decisions.

To address these questions, we shall specify and estimate a mixed "discrete-continuous" utility maximization model that is in general analogous to the profit maximization discrete-continuous production models put forth by Duncan [1980] and McFadden [1979]. The model supposes that if an individual were to attend a college or university it would be one of academic quality and cost depending on the individual's personal attributes and family background. The individual is presumed to compare the net value of opportunities with an education from a college of this quality and cost with the value of opportunities without a college education. If the supposed value to him of opportunities with a college education exceeds the value to him of the opportunities without college, he is presumed to attend. If after attendance he concludes that the net value of opportunities without a degree exceeds the value of opportunities with a degree he is presumed to drop out.

It is intended that this idea incorporate attendance and dropping out associated with searching, evaluating one's abilities and likes, monetary constraints on attendance, etc. In short, the procedure estimates jointly four outcomes--the dichotomous college attendance and college dropout (or persistence) outcomes and the two continuous college quality and college cost decisions. The college quality and cost estimates represent the preferred type of school among those available to the individual if, given his individual attributes and family background, he were to choose among college possibilities. Estimation of the college dropout relationship yields for an individual the probability of dropping out, if given his attributes he were to attend the most preferred of the college alternatives. And of course the college attendance equation yields estimates of attendance for an individual with given attributes. For the purposes of our analysis, the model is estimated in reduced form. The most important results are presented in the form of simulations based on the parameter estimates pertaining to the four outcomes.

Our estimation procedure has at least two related substantive advantages. One is that it allows explicitly for individual self-selection of higher education. Given measured attributes, persons who elect to attend college are likely to have unmeasured attributes that differ systematically from the unmeasured attributes of those who elect not to attend. For example, among persons with the same measured attributes, those who attend college are likely on average to get more from school than those who don't attend, and thus will be willing for example to pay more for a college education. This is similar to the idea captured by Willis and Rosen [1979] in distinguishing the relationship between unmeasured determinants of college education and

unmeasured determinants of success in college versus non-college occupations.

The related and concomitant advantage of the procedure is that it allows estimation of outcomes for any person in the population, not just those who have elected to attend a particular college. That is, it corrects for self-selection bias and in so doing yields estimates of population parameters. Invariably, studies of the relationship between measures of pre-college academic ability and college success are based on relationships between test scores and first year college grades, say, for persons attending a single college or university. Studies of the "validity" of the Scholastic Aptitude Test (SAT) are based on this sort of relationship, with the validity criterion usually taken to be a correlation or multiple correlation coefficient.<sup>1</sup> Not only does this process ignore the important self-selection of college versus no college, but it also ignores both the individual selection of a particular college from among many possibilities and the admissions decisions of colleges. All of these decisions tend, given SAT scores, to allocate persons between college and non-college and among colleges according to individual comparative or competitive advantage. Individuals tend to go where they will "do well" and a given college tends to select persons who will do well at that college. Thus the relationship between grades and SAT scores at a particular school underestimates the relationship in the population between SAT scores and expected academic performance. Our analysis allows estimation of the probability that an individual selected at random from the population will drop out of the type of college he would be expected to choose, should he attend.

---

1. See for example Lord and Novick [1968] or Cronbach and Gleser [1965].

In addition, validity analyses of the relationship between SAT scores and grades exclude persons who drop out of school during the first year. Our emphasis on persistence in college avoids this problem while concentrating on what may be the single most important indicator of acquisition of college credentials that will yield subsequent occupational and monetary rewards.

Our results indicate that persons with a low probability of attending college, if they were to attend would have a high probability of dropping out. Thus attendance is strongly consistent with the likelihood of benefitting from college by obtaining a degree. Attendance reflects self-selection explained by measured as well as unmeasured individual attributes. For example, after controlling for family economic and social background characteristics, a range of values of test scores and high school class rank yields estimated attendance probabilities ranging from a low of about 0 to a high of about .85, while corresponding dropout probabilities range from about .90 to about .10. And among persons with the same measured characteristics, persons who are ex post observed not to go to college, if they were to attend would prefer schools of lower quality and cost and would have a higher probability of dropping out than persons who are observed to attend.

If attendance and persistence are predicted on the basis of test scores and high school class rank only--without controlling for socioeconomic background and other determinants of persistence, as is the case in most validity studies--the variation in dropout probabilities with test scores is more pronounced. By these measures, test scores provide substantial distinction among individuals in their estimated persistence probabilities

were they to go to college. To the extent that individual educational investment decisions are determined by SAT scores, these decisions appear in the aggregate to be strongly related to the estimated likelihood that the investment (college attendance) would be justified ex post.

These results are in sharp contrast with much of the recent interpretation of the findings of validity studies of SAT tests. As mentioned above, these studies emphasize binary or multiple correlations between test scores and/or class rank on the one hand and college grades on the other. They also by implication emphasize the effect of test scores on college admissions decisions while largely ignoring their relationship to student choices, and they ignore student persistence decisions which may be the single most important indicator of success in college. Furthermore, they are invariably limited to relationships within a single college or university. Both self-selection by students and decisions of admissions officers tend to minimize the relationship between test scores and performance among students in a single college or university.

We hasten to add that our results should not be interpreted to mean that test scores explain a large part of the variation in academic performance among individuals. We show that the effect of test scores on persistence (a "slope parameter") is large, not that the unexplained variation in college performance is small. And we emphasize that our analysis pertains to a national random sample of high school graduates and thus of colleges and universities. Our findings may not reflect relationships that exist with a single university or college. In particular they may be less accurate at the tails of the distributions of individual and college characteristics than around their central tendencies, necessarily

more representative of the weight of the data.

Section I is a description of the model we have used. The data are described in Section II. Estimated parameters and simulations based on these estimates are discussed in Section III. Section IV contains concluding comments and discussion. Some details of the estimation procedure are presented in an appendix.

## I. The Statistical Model

We begin by supposing that each individual is characterized by a vector of attributes  $X$ , with elements describing the individual's socioeconomic background, academic ability and past performance, and the local labor market conditions that he faces. Upon high school graduation but without a college education, given  $X$ , the individual is assumed to face a set of opportunities to which he attaches a value  $U_0$ , that depends on  $X$ ,

$$(1) \quad U_0 = Xa_0 + e_0 ,$$

where  $a_0$  is a vector of parameters and  $e_0$  is an error term representing the collective contribution to  $U_0$  of unmeasured characteristics including tastes.

If an individual with attributes  $X$  were to attend college, we assume that he would prefer--among those that he could attend--a school of quality

$$(2) \quad Q = Xa_Q + e_Q .$$

And, if the individual were to attend, we assume that he would prefer a school with cost of attendance given by,

$$(3) \quad C = Xa_C + e_C ,$$

where  $a_C$  is a vector of parameters and  $e_C$  an error term. Indeed, college quality and cost are likely to be determined jointly and our estimation procedure will allow for that.

At the time of high school graduation, the individual is also assumed to attach a value  $U_1$  to the opportunities he supposes he would have if he were to attend college. The expected net benefit that an individual associates with college attendance is assumed to depend not only on his attributes but also on the quality and cost of the most preferred college among those he could attend.<sup>1</sup> Thus,  $U_1$  is assumed to be given by

$$(4) \quad U_1 = Xa_1 + \alpha_1 Q + \delta_1 C + e_1 \\ X(a_1 + \alpha_1 a_Q + \delta_1 a_C) + (e_1 + \alpha_1 e_Q + \delta_1 e_C) ,$$

if (2) and (3) are substituted for  $Q$  and  $C$ .

Finally, the individual is assumed to attend college if--at the time of high school graduation--the net value that he attaches to the opportunities that would be available to him if he had college education is greater than the value he attaches to opportunities available to him without college. The probability that he will attend college is thus

$$(5) \quad \Pr[U_1 - U_0 > 0] \\ = \Pr[X(a_1 + \alpha_1 a_Q + \delta_1 a_C - a_0) + (e_1 + \alpha_1 e_Q + \delta_1 e_C - e_0) > 0] \\ = \Pr[A = X\beta_1 + \varepsilon_1 > 0] ,$$

where  $A$ ,  $\beta_1$ , and  $\varepsilon_1$  are defined by the last equality.

---

1. Evidence on the relationship between college quality and earnings is provided by Wise [1975a, 1975b], Solmon [1975], and Morgan and Duncan [1979].

This attendance specification incorporates both student application and college admission decisions. In an earlier paper (Venti and Wise [1980]), we estimated a model that distinguished the two decisions and indeed was focused on the college admission decision versus the student application decision. We found that the student application decision is the primary determinant of college attendance.<sup>1</sup> Because our primary concern in this paper is on dropping out of college, we concluded that our analysis would not be appreciably affected by distinguishing with separate equations the application and admission decisions.

To determine the relationship between personal attributes and college "completion," we observe persons in October of the fifth year after high school graduation. Those who went to a four-year college and are either still attending or who have obtained a bachelor's degree we assume to have "persisted," and those who are not in college and have not obtained a degree we assume have dropped out. (For some, of course, dropping out may be temporary.) An individual's perception of the costs and benefits of a college education, as well as his perception of opportunities without a college degree, over time may change from his perceptions when he graduated from high school. Schooling is likely to be part of the searching and learning process that with respect to youth is most often mentioned as integral to youth job hunting and an important reason for the high job turnover rate among youth. Just as youth take jobs to "try them out," some go to college for the same reason. Some youth may enter a college or university without a clear intent of obtaining a degree.

---

1. Approximately 89 percent of applicants were admitted to the college or university that was their "first choice."

To simplify the analysis--and we feel without appreciably affecting our results or their interpretation--we assume that persons after entering college associate with college and non-college opportunities, values that may differ from  $U_0$  and  $U_1$ . Suppose that during the first five years after high school (October of the fifth year) the highest value associated with opportunities without a college degree is given by  $V_0$  and the value associated with opportunities with a degree is given by  $V_1$ . Both may depend on the quality and cost of the college a person enrolled in, but because we assume that they are affected directly by personal attributes  $X$  as well as by these same attributes indirectly through college quality and cost, we specify  $V_0$  and  $V_1$  in terms of  $X$ , without distinguishing the direct effect of college quality and cost, as

$$(6) \quad \begin{aligned} V_0 &= Xb_0 + u_0 , \\ V_1 &= Xb_1 + u_1 , \end{aligned}$$

where  $b_0$  and  $b_1$  are parameters and  $u_0$  and  $u_1$  are error terms. Then, the probability of persisting in college is given by

$$(7) \quad \begin{aligned} & \Pr[V_1 - V_0 > 0] \\ &= \Pr[X(b_1 - b_0) + (u_1 - u_0) > 0] \\ &= \Pr[P = X\beta_4 + \varepsilon_4 > 0] , \end{aligned}$$

where  $P$ ,  $\beta_4$ , and  $\varepsilon_4$  are defined by the last equality.

To recapitulate a bit, we have four equations:

$$\begin{aligned} A_i &= X_i \beta_1 + \epsilon_{1i} , \\ Q_i &= X_i \beta_2 + \epsilon_{2i} , \\ C_i &= X_i \beta_3 + \epsilon_{3i} , \\ P_i &= X_i \beta_4 + \epsilon_{4i} , \end{aligned} \tag{8}$$

where  $\beta_2 = a_Q$ ,  $\epsilon_2 = e_Q$ ,  $\beta_3 = a_C$ ,  $\epsilon_3 = e_C$ , and  $i$  indexes individuals. The college quality and cost variables  $Q$  and  $C$  are continuous and the variables  $A$  and  $P$  are latent indicator variables with the properties that individual  $i$  attends college if  $A_i > 0$  and persists in college if  $P_i > 0$ .

The college quality and college cost variables  $Q_i$  and  $C_i$  are assumed to reflect optimal college attributes for individual  $i$ , given  $X$ , if he were to attend college. Of course  $Q$  and  $C$  are observed only for persons who attend a college or university, and only persons who attend can persist (or drop out). For this reason alone, it is necessary to consider attendance along with the other decisions, since the unmeasured determinants of all of the outcomes are likely to be correlated. The empirical estimates, however, yield more than this mechanical correction for statistical bias. They yield estimates that in particular allow estimation for any high school graduate of the probability that he would drop out of college, should he decide to attend, with allowance made for the expected quality and cost of the college he would select if he were to attend. (Our ultimate goal is to compare the probability

of dropping out among persons with different attributes, and to relate this evidence to student application and college admission decisions from our earlier work. We will return to this in our discussion of the results.)

The model as specified is analogous to the discrete-continuous production (profit maximization) model presented by Duncan [1980] and also by McFadden [1979]. Our model pertains to utility maximization, however, and we will estimate it using an unconstrained maximum likelihood procedure, which incorporates fewer conditions than the analogous Duncan or McFadden versions of the production model; but the underlying model also has less structural content than these production model counterparts. Our model is in reduced form. It would be informative, for example, to know the relationship of college cost and college quality to persistence, that is, to include them as right-hand variables in the persistence equation. But both are surely determined by the same variables that determine persistence itself. Thus, without making what we believe are unrealistic exclusion or covariance restrictions, we are unable to obtain adequate identification of coefficients on these variables in the persistence equation.

A covariance matrix  $\Sigma$  describes the relationships among the random terms in (8), with

$$(9) \quad \Sigma = \begin{bmatrix} 1 & & & \\ \sigma_{12} & \sigma_2^2 & & \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & 1 \end{bmatrix} .$$

Note that the variances of both  $\epsilon_1$  and  $\epsilon_4$  are set to 1.<sup>1</sup>

Three outcomes are possible: (i) Individual  $i$  does not attend college so that  $A_i < 0$ . (ii) Individual  $i$  attends a college of quality  $Q_i$  and cost  $C_i$  and persists so that  $A_i > 0$ ,  $Q_i$  and  $C_i$  are observed, and  $P_i > 0$ . (iii) Individual  $i$  attends a college but drops out so that  $A_i > 0$ ,  $Q_i$  and  $C_i$  are observed, and  $P_i < 0$ .

If the  $\epsilon_{ki}$  are assumed to be multivariate normal,  $f$  is taken to be the bivariate density of  $Q_i$  and  $C_i$ , and  $g$  is the conditional bivariate density of  $A_i$  and  $P_i$ , given  $Q_i$  and  $C_i$  (more precisely, of  $\epsilon_{1i}$  and  $\epsilon_{4i}$ , given  $\epsilon_{2i}$  and  $\epsilon_{3i}$ ) then the probabilities of the three outcomes are given by:

$$(10-i) \quad \begin{aligned} \Pr(A_i < 0) &= \Pr(\epsilon_{1i} < -X_i\beta_1) = 1 - \Phi[X_i\beta_1] \\ &= P_{1i} \end{aligned}$$

$$(10-ii) \quad \begin{aligned} &\Pr(A_i > 0, Q_i, C_i, P_i > 0) \\ &= \Pr(\epsilon_{1i} < X_i\beta_1 \text{ and } \epsilon_{4i} < X_i\beta_4 | Q_i, C_i) \cdot f(Q_i, C_i) \\ &= \int_{-\infty}^{X_i\beta_1} \int_{-\infty}^{X_i\beta_4} g(\epsilon_1, \epsilon_4 | \epsilon_2, \epsilon_3) d\epsilon_4 d\epsilon_1 \cdot f(Q_i, C_i) \\ &= P_{2i} \end{aligned}$$

---

1. It is not possible to identify their ratio if  $\beta_1$  and  $\beta_4$  are allowed to differ.

$$\begin{aligned}
 & \Pr(A_i > 0, Q_i, C_i, P_i < 0) \\
 &= \Pr(\varepsilon_{1i} < X_i\beta_1 \text{ and } \varepsilon_{4i} > X_i\beta_4 | Q_i, C_i) \cdot f(Q_i, C_i) \\
 (10\text{-iii}) \quad &= \int_{-\infty}^{X_i\beta_1} \int_{X_i\beta_4}^{\infty} g(\varepsilon_1, \varepsilon_4 | \varepsilon_2, \varepsilon_3) d\varepsilon_4 d\varepsilon_1 \cdot f(Q_i, C_i) \\
 &= P_{3i}
 \end{aligned}$$

The log-likelihood function for N observations is given by

$$(11) \quad L = \sum_{i=1}^{N_1} \ln P_{1i} + \sum_{i=1}^{N_2} \ln P_{2i} + \sum_{i=1}^{N_3} \ln P_{3i}$$

The function L is maximized with respect to  $\beta_1, \beta_2, \beta_3, \beta_4$ , and the eight variance-covariance terms. We will report the two variance estimates  $\sigma_2^2$  and  $\sigma_3^2$ , and the six relevant correlations (instead of the covariances). Some details of the estimation procedure are presented in an appendix.

## II. The Data

We have based our estimates on data obtained from the National Longitudinal Study of the High School Class of 1972. During the Spring of 1972, approximately 23,000 high school seniors were surveyed. The data collected included information on each student's family background, high school performance, and a host of other student characteristics. The students also took a battery of six aptitude tests developed by the Educational Testing Service (ETS). Three follow-up surveys were used to obtain information on post-secondary school and work decisions as well as other related data. We have based our estimates on a random sample of 5,726 of the total sample.

We have confined our attention to four-year colleges and universities. For example, the probability of attendance refers to the probability of attendance at a four-year college; some of those who didn't go to four-year schools went to two-year or vocational schools.

Our analysis is based on several groups of variables. The first group includes variables that describe the individual's academic aptitude and high school performance, as well as non-academic achievement in high school.

The Scholastic Aptitude Test (SAT) score of an individual is presumed to have a substantial effect on his post-secondary school preferences and on available alternatives. Although SAT scores are recorded for many students in our sample, not all of them took the SAT test. Many took no college admissions tests at all, while others took the American College Testing Program (ACT) test. Our method of analysis requires that we have an academic ability measure for all persons in the sample, not only for those who took the SAT test--undoubtedly for most

because they planned to apply to some school. For this reason, we have used an "SAT" score predicted on the basis of the ETS test scores that are available for each person in our sample.

High school class rank is also presumed not only to influence students' preferences for college but to measure individual attributes related to college persistence as well. An individual's class rank is determined not only by his ability, but also by the ability of others in his high school. We do not have a definitive measure of high school quality, but have tried to control for it in part by including in our analysis the percent of students from an individual's high school who go to college. Finally, we have used two measures of non-academic achievement in high school: leadership in student government and athletic achievement.

The second group of variables is intended to measure an individual's socioeconomic background; it includes family income and parents' education. Race is also included, with the effect of race allowed to interact with geographic region--South versus non-South.

The third group includes two measures of local labor market conditions: a local wage rate and a local unemployment rate. These variables enter the attendance equation only. The analysis also includes an indicator of sex, and an indicator of rural versus city or urban high school location.

The variables are defined as follows:

SAT Score: predicted SAT scores based on 5 of the 6 ETS tests administered to the National Longitudinal Study sample.

College SAT: the average of the SAT scores of freshman students at the college. We often refer to this as college quality.<sup>1</sup>

College Cost: tuition and living expenses, minus grant aid to the student.

High School Class Rank: the percentile rank of a person in the person's high school, based on course grades.

High School Student Government Leader: one if the person was a "leader" in high school student government and zero otherwise.

High School Athlete: one if the student was a "leader" in high school athletics and zero otherwise.

Percent of High School Class Going to College: the percent of students from an individual's high school who go to colleges.

Parents' Income: family income as reported by the youth respondent.

Education of Mother (Father) less than High School: one if the individual's mother (father) had less than a high school education and zero otherwise.

Education of Mother (Father) College Degree or More: one if the individual's mother (father) had a college degree or more education, and zero otherwise. The excluded category is a high school degree but less than a college degree.

---

1. Solmon [1975] discusses several alternative measures of college quality and expresses a preference for SAT scores over survey measures such as the Gourman ratings. See also Astin and Henson [1977].

Siblings: the number of non-adult dependents of the individual's parents.

Black in the South: one if the person is black and went to high school in the South.

Black in the Non-South: defined analogously to black in the South.

White in the South: defined analogously to black in the South.

The excluded category is white in the non-South.

Local Wage Rate: average 1972 wage of manufacturing workers in the SMSA of the individual's high school, or if not available, the state average.

Local Unemployment Rate: SMSA 1972 unemployment rate (or state rate where not available).

Male: one if the individual is male and zero if female.

Rural: one if the individual went to high school in a rural area and zero otherwise.

Parents' Income Missing: one if parents' income was not reported and zero otherwise.

Means and standard deviation of the variables are shown in Appendix Table 1.

---

1. South is assumed to include the following states: Alabama, Arkansas, Delaware, Washington, D.C., Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, South Carolina, Virginia, West Virginia.

### III. Results

We shall present first the parameter estimates and discuss the implied partial effects of specified changes in each of the right-hand variables. Then we shall present more detailed simulations based on the parameter estimates--the first to describe the effect of changes in SAT scores and high school class rank and the second to describe the effect of changes in parents' income and education. The quantitative implications of self-selection are presented next. Finally, predictions based on SAT scores and class rank only are discussed, with emphasis on their relationship to current criticism of SAT scores.

#### A. The Parameter Estimates and Normalized Partial Effects of Individual Variables

The parameter estimates and their asymptotic standard errors are shown in Table 1. We shall not present standard error estimates for the simulations, but we emphasize that the major coefficients in the model are measured with considerable precision, as shown by the standard errors in Table 1. For example, many of the implied  $t$ -statistics on the SAT and class rank variables are 10 or more. To evaluate the partial effects of each of the variables, in Table 2 are shown the predicted shifts in each of the four outcomes due to specified changes in each of the right-hand variables.

The correlations among the unmeasured determinants of each of the outcomes are shown at the bottom of Table 1. It is clear from these correlation estimates that unmeasured attributes that increase an individual's propensity to attend college are strongly related to the unmeasured attributes that determine the cost and the quality of the

college he would choose, were he to attend; the relevant correlations are .94 and .45 respectively. Given measured attributes, the more value an individual attaches to college attendance, the more he is willing to pay for a college education and the more likely he is to go to a higher quality school. Of course, both the average SAT score of entering students and the cost of attendance may be measures of college quality. (It can also be seen that the unmeasured determinants of college quality and college cost are positively related as casual observation would suggest.)

The results in Table 2 show the effects of the specified changes in each of the variables, assuming that all other variables are held constant at their respective population means. If the variable is continuous, the outcome with this variable one standard deviation above its mean is compared with the outcome with the variable one standard deviation below its mean. For example, holding other variables at their sample means, a person with an SAT score one standard deviation above the mean for all persons is estimated to have a probability of college attendance .234 higher than a person with an SAT score one standard deviation below the mean. Evaluated at the mean of all the right-hand variables, the expected attendance probability is .21, college quality is 908, college cost is -\$314, and the persistence probability is .47.

By these measures SAT scores and high school class rank are of approximately equal importance in determining college attendance, college quality, and college cost. But the relationship of class rank to persistence is over three times as great as the relationship to persistence of SAT scores. That is, once persons enter the schools optimal for

them, given their measured characteristics, high school class rank is a stronger indicator of persistence than are SAT scores, although both are statistically significant. This result is not surprising if it is assumed that given academic ability or achievement as reflected in SAT scores, class rank reflects in part differences in individual effort or desire to perform well. Recall that the model also controls for high school "quality," to the extent that it is measured by the proportion of students from an individual's high school who go to college. It can also be seen in Table 2 that holding constant other variables, including SAT score and high school class rank, persons from "better" high schools are considerably more likely to persist in college.

It might be expected that given an indication of how well a student performed academically as well as an indication of the abilities of those with whom he was competing, the additional information that the SAT score provides would be limited. The former two measures provide observations on the sort of performance that SAT scores are intended to predict, but in high school instead of college. Because together they provide an observation that is similar to what the SAT score is intended to predict, it should not be surprising that they are better predictors of subsequent but similar student performance than SAT scores are. We will see below that if predictions are based solely on SAT scores and class rank, the importance of knowing SAT scores is considerably greater.

The difference in the probability of college attendance, as well as college quality and cost, associated with college educated parents

Table 1. Parameter Estimates (and Asymptotic Standard Errors), by Equation.

Variable	Probability of Attendance	College Quality	College Cost	Probability of Persistence
SAT Score ÷ 1000	2.233 (0.160)	0.225 (0.021)	1.700 (0.193)	0.786 (0.305)
High School Class Rank ÷ 100	1.454 (0.106)	0.142 (0.016)	1.093 (0.132)	1.613 (0.227)
Proportion of High School Class Going to College	0.809 (0.120)	0.060 (0.015)	0.691 (0.135)	1.025 (0.226)
High School Student Leader	0.246 (0.072)	0.023 (0.008)	0.154 (0.082)	-0.042 (0.106)
High School Athlete	0.370 (0.061)	0.027 (0.007)	0.286 (0.067)	0.140 (0.105)
Parents' Income	0.220 (0.049)	0.019 (0.006)	0.401 (0.055)	0.357 (0.085)
Education of Mother Less than High School	-0.129 (0.056)	-0.010 (0.008)	-0.158 (0.068)	0.029 (0.114)
Education of Mother College Degree or More	0.138 (0.063)	0.014 (0.007)	0.172 (0.066)	0.108 (0.094)
Education of Father Less than High School	-0.244 (0.055)	-0.008 (0.007)	-0.221 (0.068)	-0.072 (0.106)
Education of Father College Degree or More	0.254 (0.057)	0.025 (0.007)	0.242 (0.062)	0.140 (0.091)
Siblings	-0.026 (0.012)	-0.002 (0.002)	-0.028 (0.014)	-0.012 (0.023)
Black in the South	0.749 (0.105)	-0.127 (0.011)	0.594 (0.147)	0.510 (0.181)
Black in the Non-South	0.925 (0.100)	0.033 (0.012)	0.655 (0.116)	0.006 (0.196)
White in the South	0.038 (0.055)	-0.033 (0.007)	-0.111 (0.061)	-0.162 (0.086)
Rural High School	-0.016 (0.045)	-0.013 (0.006)	0.055 (0.052)	-0.027 (0.080)

Table 1. Parameter Estimates (and Asymptotic Standard Errors), by Equation (continued).

Variable	Probability of Attendance	College Quality	College Cost	Probability of Persistence
Local Wage	-0.136 (0.031)	--	--	--
Local Unemployment Rate	0.000 (0.008)	--	--	--
Male	0.193 (0.044)	0.021 (0.005)	0.132 (0.051)	0.032 (0.079)
Parents' Income Missing	0.231 (0.030)	0.025 (0.010)	0.521 (0.095)	0.415 (0.144)
Constant	-3.680 (0.199)	0.601 (0.032)	-3.081 (0.234)	-2.509 (0.458)

Correlation Matrix

Attendance	1.000			
College Quality	0.451 (0.056)	1.000		
College Cost	0.935 (0.011)	0.389 (0.046)	1.000	
Persistence	0.155 (0.099)	0.047 (0.048)	0.135 (0.074)	1.000

Standard Error of College Quality ( $\sigma_2$ )	0.098 (0.005)	Total Sample	5726
Standard Error of College Cost ( $\sigma_3$ )	1.122 (0.055)	Number Attending	1622
Log-Likelihood Value	-532.1	Number Persisting	1155

a. College quality for purposes of estimation is in 1000's, college cost is in 1000's, student SAT is in 1000's, and parents' income is in 10,000's.

Table 2. Effects of Two Standard Deviation and Category Changes in Variables, by Equation

Variable	Probability of Attendance	College Quality	College Cost	Probability of Persistence
SAT Score 1 S.D. Above the Mean Versus 1 S.D. Below the Mean	.234	81.0	611.8	.112
High School Class Rank, 1 S.D. Above the Mean Versus 1 S.D. Below the Mean	.231	77.8	597.3	.340
High School Student Leader Versus Not	.078	23.1	154.4	-.017
High School Athlete Versus Not	.121	27.2	285.7	.056
Percent of High School Class Going to College, 1 S.D. Above the Mean Versus 1 S.D. Below the Mean	.090	22.7	260.2	.153
Parents' Income, 1 S.D. Above the Mean Versus 1 S.D. Below the Mean	.070	20.5	430.8	.152
Education of Mother and Father College Degree or More Versus Education of Mother and Father Less Than High School	.233	56.9	791.9	.116
Number of Siblings, 1 S.D. Above the Mean, Versus 1 S.D. Below the Mean	-.005	-.007	-93.8	-.017

Table 2. Effects of Two Standard Deviation and Category Changes in Variables, by Equation (continued).

Variable	Probability of Attendance	College Quality	College Cost	Probability of Persistence
Black in the South Versus White in the Non-South	.265	-127.1	594.4	.198
White in the South Versus White in the Non-South	.011	-33.4	-110.8	-.064
Black in the Non-South Versus White in the Non-South	.336	32.7	654.5	.002
-----				
Male Versus Female	.057	20.7	132.3	.013
Rural Versus Non-Rural High School	-.005	-13.4	55.2	-.011
-----				
Local Wage, 1 S.D. Above the Mean Versus 1 S.D. Below the Mean	-.047	--	--	--
Local Unemployment Rate, 1 S.D. Above the Mean Versus 1 S.D. Below the Mean	.000	--	--	--

versus parents without high school degrees is quantitatively similar to the effect of a two-standard deviation change in class rank and SAT scores. A two-standard deviation change in parents' income on the other hand has a relatively small effect on college attendance and type of college, but has a greater effect on persistence than the change (just described) in parents' education. Parents' education is apparently much more important than income in determining preferences for education, but after entering college, income is relatively more important than parents' education in determining persistence.

Black youth are more likely than white youth, given other attributes, to go to college. In the South, the probability that black youth attend is .25 higher than the probability for white youth with the same attributes. In the non-South, the probability is .34 higher for black youth. While black youth in the South go to colleges with average SAT scores about 94 points lower than white youth in the South, the persistence rate of black youth is about .26 higher than the rate for white youth. The persistence rate of black youth in the South is about .20 higher than the rate of black youth in the non-South. While black youth in the South go to colleges with average SAT scores about 160 points lower than black youth in the North, their probability of persistence is about .20 higher than the persistence probability of black youth in the non-South. In the non-South, black youth go to colleges with student SAT scores that average a bit higher than those of schools attended by white youth, given academic and socioeconomic characteristics.

That black youth are more likely than white youth to go to college, given comparable academic and family background attributes, may reflect

the relatively greater returns to higher education for blacks, as found for example by Freeman [1976, 1978]. The atmosphere created by emphasis on affirmative action may also have encouraged more black youth to go to college. The average SAT scores of students in most predominantly black schools (largely in the South) are low relative to the least selective schools in the non-South. Thus some students in the South may be attracted to a college alternative that doesn't exist in other regions of the country. In addition, some black youth who go to four-year colleges in the South might be more likely to go to two-year colleges if they lived in the non-South.

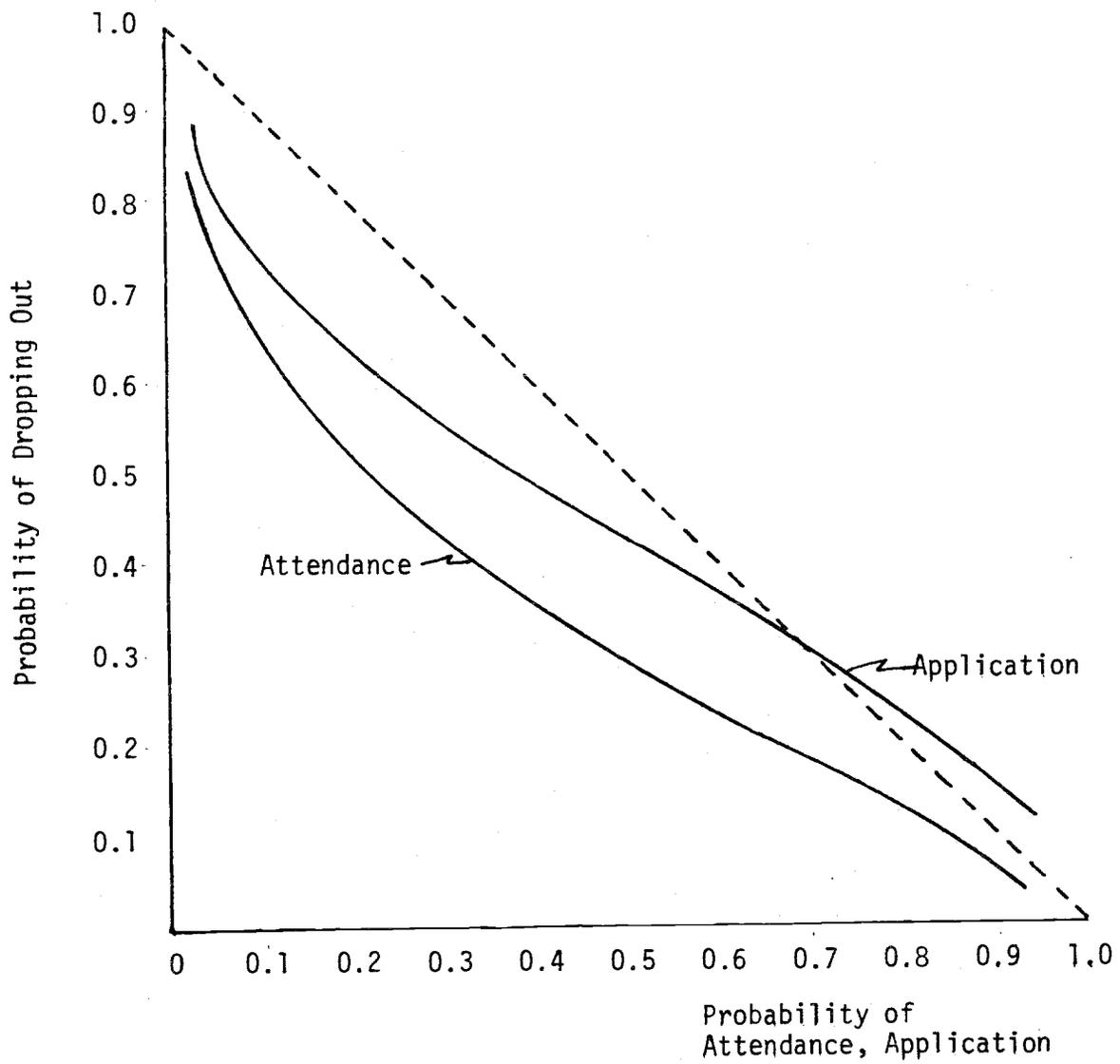
The greater returns to education for black youth are consistent also with the finding that given other attributes black youth pay more to go to college than white youth. The estimated effect of the race-region variables on cost assumes of course that other attributes--in particular SAT scores--are held constant. When these attributes are not held constant we find the average college cost for blacks is lower than for whites in both the South and the non-South, while in both regions the average of black SAT scores is about 170 points lower than the white average. These and other means of selected variables by race and region are shown in Appendix Table 2.

B. The Probability of Attendance versus the Probability of Dropping Out

In the next section we shall present simulated partial effects of selected variables on each of the four outcomes. These simulations, however, don't provide an overall indication of the relationship between attendance and persistence. To provide such a summary we have presented in Figure 1 a graph of the probability of attendance versus the probability of dropping out. To relate the probability of dropping out more purely to student decisions, we have also graphed this probability against the probability of application, taken from Venti and Wise [1980]. That is, for each individual in our sample, we have predicted attendance (and application) and dropout probabilities. Then, for each attendance (or application) interval, we have calculated the average dropout probability. These predictions are thus based on all measured determinants of the outcomes.

The figure indicates that persons who are unlikely to attend college, on average would be very likely to drop out were they to attend. The same pertains to those who are unlikely to apply to any college. On the other hand, persons who are likely to apply, and attend, are very unlikely to drop out. In this sense, student application and attendance decisions seem to be quite consistent with persistence outcomes and more generally with rational educational investment decisions. To the extent that the expected benefit from college is determined by whether a degree is obtained, persons with high expected benefit are likely to go while those with low expected benefit are unlikely to go.

Figure 1. Probability of Attendance and Application Versus Probability of Dropping Out.



C. Simulated Effects of Selected Variables

1. SAT Scores and Class Rank.

Estimated probabilities of attendance and dropping out, together with estimated college quality and cost, are presented in Table 3 for selected SAT and class rank values. Differences among the entries in the table represent estimated partial effects of SAT and class rank, with other variables in the model assigned their population means.

Persons with SAT scores and class ranks that make them unlikely college entrants would have a high probability of dropping out were they to go to college. For example, persons with SAT scores of 900 or lower and in the bottom half of their high school classes have at most a .24 probability of going to college. If persons in this group were to attend, their dropout probabilities would be very high, at least .53 and ranging up to .88 for those with the lowest test scores and poorest academic records. On the other hand, persons with scores and class rank that yield high estimated attendance probabilities are much less likely to drop out than those with low scores and class rank. While attendance probabilities range from .01 to .82, corresponding dropout probabilities range from .88 to .15.

We see also from Table 3 that if persons were to choose among college possibilities, the qualities of the schools they would choose would vary with their own academic abilities, as casual observation would suggest. Recall that estimated college quality and college cost represent the characteristics of the school a person with given attributes--randomly selected from the population--would choose, were he to choose

Table 3. Simulated Effects of SAT and High School Class Rank on Attendance, College Quality and Cost and on the Probability of Dropping Out

SAT	Class Rank (Percentile)				
	0	25	50	75	100
	Probability of Attendance				
500	.01	.03	.06	.11	.19
700	.03	.07	.13	.22	.34
900	.08	.15	.24	.37	.51
1100	.16	.27	.40	.55	.68
1300	.30	.43	.58	.71	.82
	College Quality				
500	755	791	826	862	898
700	800	836	871	907	943
900	845	881	916	952	988
1100	890	926	961	997	1033
1300	935	971	1006	1042	1076
	College Cost				
500	-1372	-1099	-825	-552	-279
700	-1032	-759	-486	-212	61
900	-692	-419	-146	128	401
1100	-352	-79	194	468	741
1300	-12	261	534	807	1081
	Probability of Dropping Out				
500	.88	.79	.65	.49	.34
700	.85	.74	.59	.43	.28
900	.81	.68	.53	.37	.23
1100	.76	.62	.47	.31	.19
1300	.71	.56	.40	.26	.15

among colleges. Thus it is also not surprising that persons who are unlikely to find school attractive would pay little to attend. Indeed these estimates indicate that for many persons the optimal college would be one that they were paid to attend; the estimated cost is negative. "Desired" college cost is not constrained to be positive.<sup>1</sup>

As mentioned above, the partial effects--after controlling for a measure of the academic level of an individual's high school, as well as socioeconomic background and other variables--show that the effect on persistence of high school class rank is much greater than the effect of SAT scores on persistence. If the percent of students from a person's high school who are college bound is not controlled for, however, the corresponding variation in dropout probabilities by SAT score is much greater than is shown in Table 3. (Such estimates are presented below.)

The attendance and dropout probabilities in Table 3 may be integrated with the application and admission probabilities from our earlier paper (Venti and Wise [1980]), based on a model with the same variables used in this paper. These results are shown in Table 4.

Based on the application and admission probabilities shown in Table 4, we concluded in our earlier paper that SAT scores were related much more to individual application than to college admission decisions. Most persons would have a high probability of admission were they to apply to a college of average quality. It is clear also from Table 4

---

1. Because unmeasured determinants of attendance are positively related to quality, cost, and persistence, the conditional quality and cost estimates are higher than the unconditional ones, and the dropout probabilities are lower. This idea is pursued further in Section D below.

Table 4. Simulated Application, Admission, Attendance, and Dropout Probabilities,<sup>a</sup> for Selected SAT and Class Rank Values<sup>a</sup>.

SAT	Class Rank (Percentile)				
	0	25	50	75	100
	Probability of Application				
500	.03	.06	.11	.19	.30
700	.07	.14	.23	.35	.49
900	.17	.27	.40	.54	.68
1100	.31	.45	.59	.72	.83
1300	.50	.64	.76	.86	.92
	Probability of Admission <sup>b</sup>				
500	.52	.63	.74	.82	.89
700	.64	.74	.82	.89	.93
900	.74	.82	.89	.93	.96
1100	.83	.89	.94	.96	.98
1300	.89	.94	.96	.98	.99
	Probability of Attendance				
500	.01	.03	.06	.11	.19
700	.03	.07	.13	.22	.34
900	.08	.15	.24	.37	.51
1100	.16	.27	.40	.55	.68
1300	.30	.43	.58	.71	.82
	Probability of Dropping Out				
500	.88	.79	.65	.49	.34
700	.85	.74	.59	.43	.28
900	.81	.68	.53	.37	.23
1100	.76	.62	.47	.31	.19
1300	.71	.56	.40	.26	.15

a. The probability of application and the probability of admission from Venti and Wise [1980].

b. At a college of average quality (with an average SAT of 1012).

that some persons with a high estimated probability of admission have a low estimated probability of attendance. In particular, this is true of persons whose SAT scores, and/or class ranks are below the sample averages.<sup>1</sup> This result also supports the finding that to the extent that SAT scores determine college attendance, their effect is more by way of individual than by way of college admission decisions.

And whether we consider the individual application decision or the attendance decision outcome--that incorporates both student and college decisions--the relationship between test scores and these decisions seems supported by the dropout probabilities. To the extent that test scores affect student and college decisions, in the aggregate, they tend to be consistent with the dropout probabilities. We will pursue this point in Section E.

## 2. Income and Parents' Education.

Simulated attendance and dropout probabilities together with college quality and cost for selected levels of parents' education and family income are shown in Table 5. Again, they are unconditional estimates--representing the predicted choices of a person with the given attributes selected from the population: college characteristics and dropout probabilities conditional on attendance differ from these, as shown in Section D below.

---

1. In part, these estimates pertain to SAT and class rank combinations that would be unlikely to be observed empirically. For example, because SAT and class rank are positively correlated, very high SAT and very low class rank is uncommon.

Table 5. Simulated Effects of Parents' Income and Education on Attendance, College Quality and Cost, and on the Probability of Dropping Out.

Parents' Education	Parents' Income				
	\$6,000	\$9,000	\$12,000	\$15,000	\$18,000
	Probability of Attendance				
Less than High School	.11	.13	.14	.15	.17
College Degree or More	.33	.35	.38	.40	.43
	College Quality				
Less Than High School	879	885	890	896	902
College Degree or	936	941	947	953	959
	College Cost				
Less Than High School	-747	-627	-507	-387	-267
College Degree or More	45	165	285	405	525
	Probability of Dropping Out				
Less Than High School	.62	.57	.53	.49	.45
College Degree or More	.50	.46	.42	.38	.34

Relative to parents' education, parents' income appears to affect dropout probabilities much more than attendance or college quality or college cost. Apparently persons whose parents earn less are substantially more likely to drop out of college for monetary reasons than persons from wealthier families, if family income is interpreted strictly as a determinant of individual budget constraints. This observation is confirmed by the estimated effects of two standard deviation shifts reported in Table 2. Indeed parents' income has considerably less effect than SAT scores except in the determination of dropping out. Parents' education, on the other hand, is by these comparisons relatively less important in the determination of dropping out than in the determination of attendance and college characteristics. Judging by the evidence in Table 2, however, the three measures of individual academic ability--SAT scores, the percent of the individual's high school class who go to college, and high school class rank--taken together are much more important than family income in the determination of persistence. Nonetheless, our results suggest that if educational attainment were not to be related to family income, for example, low family income would have to be offset by external funds, even if income itself were not a major determinant of college attendance.<sup>1</sup>

---

1. Recall that we have used net college cost, subtracting aid from college tuition plus living cost. Holding other determinants constant, persons from poorer families obtain more aid than persons from wealthier families, although in 1972 academic performance was an important determinant of aid. In subsequent years after the introduction of the federal Basic Equal Opportunity Grant program, relatively more aid was allocated in inverse relation to family income. (See Fuller, Manski, and Wise [1980a, 1980b].)

D. The Importance of Self-Selection

The foregoing results highlight the very substantial relationship between measured individual attributes on the one hand and college selection and persistence on the other. The effect of measured attributes on outcomes, however, represents only one of the components of self-selection. The other is the effect of unmeasured individual attributes which may be expected to have comparable effects on self-selection. In particular we proposed above that given measured attributes, individuals who elect to go to school are likely to be disproportionately those whose relative opportunities will be most enhanced by college or who most enjoy school, etc., and who are the most likely to succeed in college. The estimated correlations among the disturbance terms reported in Table 1 support this proposition. The correlations may be interpreted as an indication of self-selection bias.<sup>1</sup>

To provide a better measure of the importance of this form of self-selection, we have estimated the difference in college quality, college cost, and the probability of dropping out, conditional on attendance versus conditional on non-attendance. The results evaluated at the population means of all variables are as follows:<sup>2</sup>

	<u>Conditional on Attendance</u>	<u>Conditional on Non-Attendance</u>	<u>Difference</u>
	(1)	(2)	(1)-(2)
College Quality	966	893	73
College Cost	1188	-720	1908
Probability of Dropping Out	0.45	0.55	-0.11

---

1. See Hausman and Wise [1979], for example.

2. The probabilities have been rounded to the nearest hundredth.

Given personal attributes, persons who elect not to go to college, if they were to attend, would choose schools of lower quality than the schools selected by persons with the same attributes who go to college. Also, among persons with the same attributes, those who don't go would have a higher probability of dropping out than persons who are observed to go. Possibly most noticeable is the very large difference between the average cost of schools selected by those who attend versus the cost "desired" by students who don't attend. On average, students who attend are prepared to pay much more for college than those who don't attend.<sup>1</sup>

E. The Information Content of SAT Scores and Class Rank and Implications for Test "Validity."

The simulations above are based on a model that we believe incorporates measures of the major determinants of college-going decisions, namely, it incorporates both academic and family socioeconomic variables. Measures of academic ability and past performance are assumed to be reflected in SAT scores, high school class rank, and the proportion of a person's high school classmates who go to college. Thus all of the above estimates are of partial effects after controlling for other variables. In particular, unreported estimates suggest that when the percent of an individual's high school classmates who go to college is

---

1. Mechanically, these effects are due to the size of the Mills ratio for the different values of X. For example, the expected value of cost given attendance is

$$X\beta_3 + \rho_{13}\sigma_3(\phi(X\beta_1)/\Phi[X\beta_1]) ,$$

where  $X\beta_3$  is shown in Table 3,  $\rho_{13} = .935$  and  $\sigma_3 = 1122$ ;  $\phi$  is the unit normal density function and  $\Phi$  is the corresponding distribution function.

added to the model, the relationship between high school class rank and dropping out of college becomes larger and the partial relationship between SAT score and dropping out smaller. The percent of high school classmates who go to college appears to be an alternative measure of the academic ability of the individual; or class rank relative to SAT score means more if we have an indication of the abilities of the students against whom he was competing.

But much of the recent discussion on the subject is motivated by claims that the predictive "validity" of the tests is low, in particular relative to high school class rank. To address these claims, we have estimated our model using only SAT scores and high school class rank, together with the three race-region variables. Simulations analogous to those in Table 4, but based on this limited predictive model, are shown in Table 6. (Parameter estimates analogous to those in Table 1 but based on the simplified model, together with simulations analogous to those in Table 3, are shown in Appendix Tables 3 and 4 respectively.)

It is clear from the bottom section of the table that given high school class rank, substantial predictive information is contained in SAT scores. The change in the dropout probability associated with a 200 point change in the SAT score is about the same as the change in the dropout probability associated with a 25 point change in class rank. For example, begin with a person at the 50th percentile in his high school class and with an SAT of 900. If the person instead were at the 75th percentile in class rank the estimated dropout probability would be .11 lower; similarly if instead the person had an SAT score of 1100,

Table 6 Simulated Application, Admission, Attendance, and Dropout Probabilities, for Selected SAT and Class Rank Values, Based on Simplified Model<sup>a</sup>.

SAT	Class Rank (Percentile)				
	0	25	50	75	100
	Probability of Application				
500	.02	.04	.07	.11	.17
700	.09	.14	.21	.29	.39
900	.25	.34	.44	.54	.64
1100	.49	.60	.69	.78	.85
1300	.74	.82	.88	.92	.95
	Probability of Admission <sup>b</sup>				
500	.45	.59	.71	.81	.89
700	.56	.69	.79	.88	.93
900	.66	.78	.86	.92	.96
1100	.76	.85	.91	.96	.98
1300	.83	.90	.95	.98	.99
	Probability of Attendance				
500	.01	.02	.03	.06	.10
700	.04	.07	.12	.18	.25
900	.14	.21	.29	.38	.49
1100	.33	.43	.53	.63	.73
1300	.58	.68	.76	.83	.89
	Probability of Dropping Out				
500	.86	.79	.70	.60	.49
700	.78	.69	.59	.48	.37
900	.68	.58	.47	.36	.26
1100	.57	.46	.35	.25	.17
1300	.44	.34	.24	.16	.10

a. Based on a model that includes only SAT and class rank together with the race-region variables.

b. At a college of average quality (average SAT of 1012).

The dropout probability would be .12 lower.<sup>1</sup>

Comparisons based on high school class rank and SAT score are important because these are the measures that are most often considered in evaluation or criticism of the predictive validity of test scores, or the information content of the scores. The criticism, however, is often justified on the basis of the limited increase in the multiple correlation between college grades on the one hand, and high school class rank and SAT scores on the other, over the simple correlation between college grades and high school class rank.<sup>2</sup> Even though the relatively high correlation between SAT scores and high school class rank (.50 in our sample) means that expected dropout probabilities can be predicted well with only high school class rank, given class rank the difference in expected dropout probabilities varies greatly with SAT scores. This means that although on average colleges could predict reasonably well on the basis of class rank alone, these predictions would be far from expected dropout probabilities in some instances. In particular, individuals with SAT scores very different from the expected SAT score given class rank would be either substantially disad-

---

1. A 200 point change in SAT scores and a 25 point change in class rank are both roughly equivalent to one standard deviation. Calculations based on an exact one standard deviation change yield a .11 change in the dropout probability for each of them.

2. The validity coefficient is the binary or multiple correlation between the criterion variable and one or more predictor variables. A recent survey of over 800 studies of college grades versus SAT scores and high school performance found validity coefficients of about 0.52 for high school performance, 0.41 for SAT score, and 0.58 for both predictors together. See Ford and Campos [1977].

vantaged or substantially advantaged if admissions were based on class rank alone.

It is also informative to compare the estimated weights associated with the two measures in individual decisions versus college admission decisions. According to our formulation of the problem the estimated coefficients in the student attendance equation represent the relationship between the associated variables and the individual's evaluation of opportunities with education at the school he could attend versus opportunities without college. A similar interpretation can be given to estimated coefficients in the application and dropout equations. In the admission equation, the coefficients may be interpreted as the weights that colleges and universities assign to these measures in determining an individual's academic potential.<sup>1</sup>

For each estimated relationship we have calculated the ratio of the SAT to the class rank coefficient; that is, the number of class rank percentile points that is equivalent to 10 SAT score points. The results are as follows:<sup>2</sup>

Probability of Application	3.14
College Quality (Application)	3.94
College Quality (Attendance)	3.10

---

1. See Venti and Wise [1980].

2. The ratios in the probability of application, college quality (application), and the probability of admission equations are taken from equations analogous to those in Venti and Wise [1980], but based only on SAT scores and class rank together with the race-region variables.

College Cost (Attendance)	3.78
Probability of Attendance	3.05
Probability of Admission	1.02
Probability of Persistence	1.38

In the light of recent criticism of the use of SAT scores in admissions, it is striking that the class rank percentile point equivalent of 10 SAT points is a bit less in admissions than it is in predicting persistence. Thus should persistence be the criterion for admission, colleges and universities on average may if anything assign a somewhat lower weight to SAT scores versus high school performance than persistence predictions suggest.

Note also that relative to class rank, SAT scores bear a much greater relationship to the values that individuals assign to opportunities with and without college than they do to the "average" university's evaluation for admission purposes. The importance of the individual attributes reflected in SAT scores shows up not only in individual application decisions but in college quality and cost decisions as well. On average these (by and large) individual decisions are also much more strongly related to test scores than college admission decisions are. The same may be said for ultimate attendance decisions, largely individual choices but affected in part by college admission decisions.

Apparently, the relationship between test scores and what individuals think they can get out of school is much more pronounced than colleges and universities--in effect society's--evaluation of the opportunity that should be available to them should they want to enter college,

indeed much greater than the relationship between test scores and academic success, measured by observed persistence. Individuals are likely to evaluate schooling opportunities not only in terms of persistence but in terms of human capital development while in school (and subsequent earning gains) and simply enjoyment of schooling as well.<sup>1</sup> Both would suggest that student schooling decisions are determined more by academic ability than university admission decisions are. Individuals not only weight academic ability highly in deciding whether to go to college at all, but are also willing to pay more for a college education if they have greater academic ability. And persons with greater academic ability are also likely to choose a more competitive and higher quality college. Both are consistent with rational individual human capital decisions. Simply put, the results are consistent with the hypothesis that the net returns to education are low for persons with low academic ability or interest. Colleges on the other hand don't insist that every entrant do well; neither do they insist that every entrant like school. Individuals whose lives are personally affected by such things, apparently are much more inclined to take them into account in making their decisions. Apparently, the provision of individual college opportunities in general is considerably more in the hands of individuals than at the behest of institutions, and apparently the individual decisions are strongly related

---

1. One explanation may be complementary between ability (as measured by test scores) and schooling in the determination of earnings. The value of a college degree may be more for students higher in the attributes measured by test scores. These attributes may also be more closely related to the consumption aspects of schooling.

to the attributes reflected in SAT scores.

We note also that the relative weight of SAT scores is much lower in persistence than in the other individual decisions (supposing that persistence is largely an individual and not an institutional choice). SAT scores of course provide a comparison of individual attributes with those of a much broader population than is provided by class rank which compares an individual only with his high school classmates. Since in making post-secondary schooling decisions an individual is likely to think of his attributes relative to those of a group much broader than his high school classmates, it is not surprising that the attributes reflected in SAT scores should weight relatively more heavily than high school class rank. On the other hand, once individuals have allocated themselves to college versus no college and among colleges, in large part according to the attributes reflected in SAT scores, it is also not surprising that given this allocation, class rank should weigh relatively more heavily in determining persistence than in determining college attendance and college type.

We emphasize that the calculations in this section derive from predictions based on SAT scores and high school class rank only. As we have already mentioned, relative to class rank the partial effect of SAT scores on dropping out is considerably less in the more behavioral model (Table 3 for example) than in the simplified specification. In large part the difference results from the inclusion in the behavioral specification of the percent of an individual's high school class who go to college. Apparently though, SAT scores in the simple model also pick

up to some extent individual attributes contributing to success in college but that should be ascribed to family socioeconomic background rather than academic ability. In particular, family income affects persistence and its effect is apparently picked up in part by SAT scores when income is not included in the model. In any case, it is clear that the additional predictive information contained in SAT scores, given high school class rank, would be less if family background variables were also used to predict dropout probabilities. Such predictions used in admission decisions, however, would surely be ethically inappropriate and indeed legally proscribed.

#### IV. Concluding Comments and Discussion

Our earlier work demonstrated that a very large proportion of college applicants gain admission to their "first-choice" school. The self-selection by students of colleges and universities is striking. In particular, SAT scores are more closely related to the student than to university choice. These findings together with the results reported in this paper suggest that test scores (or the attributes reflected in them) are at least as important in student decisions as in admission decisions. It is not only the university that would like to compare one student to another, but the student who would like to judge his academic abilities relative to other students. Our findings suggest that SAT scores are likely to contribute in substantial measure to rational college attendance decisions and indeed may aid student choice more than university choice. In a statistical sense, this self-selection can be thought of in two components. One is the part that is explained by observed individual characteristics. The other is associated with unobserved characteristics and is reflected in positive correlations between the random determinants of college attendance and the random determinants of persistence in college (as well as the random determinants of college quality and college cost). For example, an individual with average attributes who attends college will drop out with estimated probability .45, while an individual who is observed not to attend would have a .55 probability of dropping out. This is analogous to the self-selection component hypothesized by Willis and Rosen [1979] with respect to college attendance and earnings with and without a college degree.

College attendance and non-attendance decisions are consistent with the strong relationship between the likelihood of attendance and the likelihood of benefitting from college by obtaining a degree. Persons who are unlikely to attend are very likely to drop out without a degree should they attend and persons who are likely to attend are unlikely to drop out. In addition, the relationship between SAT scores and high school class rank and the probability of college attendance is strongly mirrored by an opposite relationship between these measures and the probability of dropping out of college, if a college should be attended. Our results also confirm that the quality of school persons attend--largely by choice--is strongly related to their aptitude and past performance in high school. This finding is in accord with casual observation, but it is also consistent with evidence based on choice models of selection among schools to which persons have been admitted.<sup>1</sup> Possibly less obvious, but consistent with economic reasoning, is the result that persons who are unlikely to succeed in school, if choosing among colleges would not only select low quality schools but low cost schools as well. Indeed for many, the preferred school would be one that individuals were paid to attend. Persons who are unlikely to succeed in school on average would not be as inclined to pay a lot for the privilege.

To the extent that colleges base admissions decisions on test scores, their selection procedures distinguish persons according to expected probability of success. Indeed the relative weight assigned to SAT scores

---

1. See Radner and Miller [1975], Kohn, Manski, and Mundel [1976], and Fuller, Manski, and Wise [1980a].

versus high school class rank by college admissions officers is somewhat less than the relative weight of SAT scores in the prediction of the probability of college persistence. (This is not to say of course that test scores necessarily explain a great deal of the variation in academic performance.)

Our findings are inconsistent with the implications of much of the recent discussion and criticism of test scores. Predictions based on SAT scores and high school class rank yield estimates that distinguish rather dramatically the school preference and college success probabilities of high school graduates. Our previous research indicated that the relationship between these measures and college attendance outcomes was due much more to their relationship to individual application decisions than to their effect on college admission decisions. The relationship between these measures and expected probabilities of success in college tend to support the individual application decisions since occupation and monetary advantages of going to college are limited if a degree is not obtained.

Together these results provide two primary pieces of evidence that are in contradistinction to much of the criticism of SAT scores. First, the test scores are much more related to student application decisions than to admission decisions of colleges. Second, the scores are rather dramatically related to the probability that persons will drop out of college, a consideration not given weight by critics of the predictive validity of SAT scores, who have emphasized the relationship between test scores and first year college grades. Furthermore, the studies of the predictive validity of tests do not account for the self-

selection of students among colleges in such a way as to match their interests and capabilities with the characteristics of the colleges, as human capital considerations would suggest. According to our findings, students self-select themselves to colleges and universities in large part on the basis of SAT scores. This alone would suggest that the correlations between SAT scores and grades within individual colleges or universities--the basis for many of the conclusions of critics such as Slack and Porter and the Nader organization--are pushed toward zero. Thus, even within the limited criteria of correlation coefficients, they tend to underestimate the relationship between success in college and test scores.

APPENDIX ON ESTIMATION

Because the estimation procedure is not new, we will not go through it in detail, but will point out only those features that may be particularly relevant--although not unique--to our application.

Suppose that given Q and C, the conditional distribution of A and P is multivariate normal with mean vector  $(\mu_1, \mu_2)$  and covariance matrix

$$\Omega = \begin{pmatrix} \omega_{11} & \\ \omega_{12} & \omega_{22} \end{pmatrix}.$$

For convenience we will deviate from the text notation by supposing that the covariance matrix  $\Sigma$  pertains to (A, P, Q, C), rather than (A, Q, C, P).

The probability of not going to college is given by

$$\Pr(A < 0) = 1 - \Phi[X_1\beta_1].$$

The probability of going to college and persisting may be written as

$$\begin{aligned} & \Pr(A > 0, P > 0, Q, C) \\ &= \int_{-\infty}^{\mu_1/\omega_1} \int_{-\infty}^{\mu_2/\omega_2} h(\lambda_1, \lambda_2; \rho^*) d\lambda_2 d\lambda_1 \cdot f(Q, C), \end{aligned}$$

where h is a standardized bivariate normal density function with correlation coefficient  $\rho^*$ . and  $\omega_j = \omega_{jj}^{1/2}$ .



$$\sigma_{11} = 1$$

$$\sigma_{12} = P_2 = \cos \theta$$

$$\sigma_{13} = P_4$$

$$\sigma_{14} = P_7$$

$$\sigma_{22} = P_2^2 + P_3^2 = \cos^2 \theta + \sin^2 \theta = 1$$

.

.

.

etc.

The likelihood function is maximized with respect to  $\beta_1, \beta_2, \beta_3, \beta_4, \theta,$  and  $P_4$  through  $P_{10}$  using an algorithm proposed by Berndt, Hall, Hall, and Hausman [1974].

Appendix Table 1. Variable Means and Standard Deviations for All Students and Students Attending College.

Variable	All Students		Attending College	
	Mean	Std. Dev.	Mean	Std. Dev.
SAT Score	828	180	959	175
High School Class Rank	.551	.273	.717	.220
Proportion of High School Class Going to College	.503	.188	.558	.196
High School Student Leader	.067	.250	.126	.332
High School Athlete	.102	.302	.161	.368
Parents' Income <sup>a</sup>	11186.	5380.	12805	5584
Education of Mother Less than High School	.269	.444	.134	.340
Education of Mother College Degree or More	.114	.318	.218	.413
Education of Father Less than High School	.324	.468	.163	.369
Education of Father College Degree or More	.180	.385	.342	.474
Siblings	2.037	1.705	2.012	1.608
Black in the South	.064	.2440	.056	.229
Black in the Non-South	.041	.197	.043	.202
White in the South	.214	.410	.212	.409
Rural High School	.498	.500	.440	.496
Local Wage Rate	3.922	.588	--	--
Local Unemployment Rate	3.833	1.852	--	--

Appendix Table 1. Variable Means and Standard Deviations for All Students and Students Attending College (continued).

Variable	All Students		Attending College	
	Mean	Std. Dev.	Mean	Std. Dev.
Male	.475	.499	.502	.500
Parents' Income Missing	.196	.397	.169	.375
-----				
Total Sample	5726			
Number Attending College	1622			
Number Persisting	1155			

a. Calculation excludes observations with missing values.

Appendix Table 2. Means of Selected Variables by Region and Race.

Variable	Non-South		South	
	Whites	Blacks	Whites	Blacks
For Persons Attending College: <sup>a</sup>				
SAT <sup>b</sup>	992	749	953	741
High School Class Rank (Percentile)	72.5	56.3	73.2	68.4
College Quality	1042	986	998	821
College Cost (\$'s)	1394	1214	1162	1031
Proportion Dropping Out	0.26	0.49	0.34	0.30
For Total Sample <sup>c</sup>				
SAT <sup>b</sup>	855	678	819	663
Proportion Attending College	.29	.30	.28	.25

a. Sample size = 1622.

b. Predicted on basis of ETS survey tests.

c. Sample size = 5726.

Appendix Table 3.

Simplified Dropout Model: Parameter Estimates (and Asymptotic Standard Errors), by Equation<sup>a</sup>.

Variable	Probability of Attendance	College Quality	College Cost	Probability of Persistence
SAT Score	3.162 (0.142)	0.313 (0.023)	2.848 (0.204)	1.547 (0.314)
High School Class Rank	1.037 (0.091)	0.101 (0.014)	0.754 (0.120)	1.119 (0.189)
-----				
Black in the South	0.567 (0.096)	-0.143 (0.010)	0.266 (0.142)	0.285 (0.167)
Black in the Non-South	0.841 (0.097)	0.030 (0.012)	0.539 (0.116)	-0.118 (0.176)
White in the South	0.084 (0.049)	-0.033 (0.007)	-0.130 (0.062)	-0.185 (0.083)
-----				
Constant	-4.004 (0.118)	0.616 (0.030)	-2.975 (0.223)	-1.845 (0.398)

Correlation Matrix

Attendance	1.000			
College Quality	0.470 (0.055)	1.000		
College Cost	0.938 (0.011)	0.421 (0.041)	1.000	
Persistence	0.194 (0.094)	0.109 (0.047)	0.187 (0.071)	1.000

Appendix Table 3. Simplified Dropout Model: Parameter Estimates (and Asymptotic Standard Errors), by Equation<sup>a</sup> (continued).

---

Standard Error of College Quality ( $\sigma_2$ )	0.102 (0.055)	Total Sample	5726
Standard Error of College Cost ( $\sigma_3$ )	1.185 (0.060)	Number Attending	1622
Log-Likelihood Value	-842.8	Number Persisting	1155

---

a. College quality for purposes of estimation is in 1000's, college cost is in 1000's, student SAT is in 1000's, and parents' income is in 10,000's.

Appendix Table 4. Simplified Dropout Model: Simulated Effects of SAT and Class Rank on Attendance, College Quality and Cost, and on the Probability of Dropping Out.

SAT	Class Rank (Percentile)				
	0	25	50	75	100
	Probability of Attendance				
500	.01	.02	.03	.06	.10
700	.04	.07	.12	.18	.25
900	.14	.21	.29	.38	.49
1100	.33	.43	.53	.63	.73
1300	.58	.68	.76	.83	.89
	College Quality				
500	758	783	808	833	858
700	820	845	870	896	921
900	883	908	933	958	983
1100	945	970	995	1021	1046
1300	1008	1033	1058	1083	1108
	College Cost				
500	-1540	-1351	-1163	-974	-786
700	-970	-781	-593	-404	-216
900	-400	-212	-23	165	354
1100	169	358	546	735	923
1300	739	928	1116	1305	1493
	Probability of Dropping Out				
500	.86	.79	.70	.60	.49
700	.78	.69	.59	.48	.37
900	.68	.58	.47	.36	.26
1100	.57	.46	.35	.25	.17
1300	.44	.34	.24	.16	.10

REFERENCES

- Astin, Alexander W., Predicting Academic Performance in College, Free Press, New York, 1971.
- Berndt, Ernest, Bronwyn Hall, Robert Hall, and Jerry Hausman, "Estimation and Inference in Nonlinear Structural Models," Annals of Economic and Social Measurement, pp. 653-665, 1974.
- Cronbach, Lee J. and Goldine C. Gleser, Psychological Tests and Personnel Decisions, Urbana Illinois, University of Illinois Press, 2nd edition, 1965.
- Duncan, Gregory M., "Formulation and Statistical Analysis of the Mixed, Continuous/Discrete Dependent Variable Model in Classical Production Theory," Econometrica, Vol. 48, No. 4, pp. 835-852, May 1980.
- Eckland, Bruce K., Louise Henderson, Ann Tickamyer, and William Trent, National Longitudinal Study: Constructed Education Variables, NCES, Washington, D.C., March 1979.
- Ford, Susan and Sandy Campos, "Summary of Validity Data from the Admissions Testing Program Validity Study Service," ETS, Princeton, N.J., 1977.
- Freeman, Richard B., Black Elite, New York, McGraw-Hill, 1976.
- \_\_\_\_\_ "Black Economic Progress After 1964: Who Has Gained and Why," NBER Working Paper Number 282, 1978.
- Fuller, Winship, Charles Manski, and David Wise, "New Evidence on the Economic Determinants of Post-Secondary Schooling Choices," J.F.K. School of Government Discussion Paper Number 90D, June 1980a.
- \_\_\_\_\_ "The Impact of the Basic Educational Opportunity Grant Program on College Enrollments," J.F.K. School of Government Discussion Paper Number 94D, July 1980b.

- Grilliches, Zvi, "Economic Problems of Youth," in James Coleman (ed.) Youth: Transition to Adulthood, Chicago, University of Chicago Press, 1974.
- Hausman, Jerry and David Wise, "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," Econometrica, Vol. 47, No. 2, pp. 455-473, 1979.
- Kohn, Meir G., Charles Manski, and David Mundel, "An Empirical Investigation of Factors Which Influence College-Going Behavior," Annals of Economic and Social Measurement, pp. 391-420, 1976.
- Levinsohn, Jay R. et al., National Longitudinal Study Base Year, First, Second, and Third Followup Data File Users Manual, NCES, Washington, D.C., 1978.
- Lord, Frederic M. and Melvin R. Novick, Statistical Theories of Mental Test Scores, Reading, MA, Addison-Wesley, 1968.
- McFadden, Daniel, "Econometric Net Supply Systems for Firms With Continuous and Discrete Commodities," presented at the joint Harvard-MIT Econometrics Seminar, Spring 1979.
- Morgan, James N. and Greg J. Duncan, "College Quality and Earnings," in Ismail Sirageldin (ed.), Research in Human Capital and Development, Vol. 1, pp. 103-121, 1979.
- Nairn, Allan and Associates, "The Reign of ETS: The Corporation That Makes Up Minds," 1980. (The Ralph Nader Report on the Educational Testing Service).
- Radner, Roy and Leonard Miller, Demand and Supply in U.S. Higher Education, New York, McGraw-Hill, 1975.

Slack, Warner V. and Douglas Porter, "The Scholastic Aptitude Test: A Critical Appraisal," Harvard Educational Review, Vol. 50, pp. 154-175, May 1980.

Solmon, Lewis C., "The Definition of College Quality and Its Impact on Earnings," Explorations in Economic Research, Vol. 2, No. 4, pp. 537-587, 1975.

U. S. Congress, Subcommittee on Elementary, Secondary, and Vocational Education, Hearings on the Truth in Testing Act of 1979 and the Educational Testing Act of 1979, October 1979.

Venti, Steven and David A. Wise, "Test Scores, Educational Opportunities, and Individual Choice," J.F.K. School of Government Discussion Paper Number 92D, June 1980, Revised February 1981.

Willis, Robert and Sherwin Rosen, "Education and Self-Selection," Journal of Political Economy, Vol. 87, pp. S7-S36, October 1979.

Wise, David A., "Academic Achievement and Job Performance," American Economic Review, pp. 360-366, 1975a.

\_\_\_\_\_ "Personal Attributes, Job Performance, and Probability of Promotion," Econometrica, Vol. 42, No. 5-6, pp. 913-931, 1975b.