

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Higher Education and Earnings: College as an Investment and Screening Device

Volume Author/Editor: Paul J. Taubman, Terence Wales

Volume Publisher: UMI

Volume ISBN: 0-07-010121-3

Volume URL: <http://www.nber.org/books/taub74-1>

Publication Date: 1974

Chapter Title: Appendix F: Success and Other Biases

Chapter Author: Paul J. Taubman, Terence Wales

Chapter URL: <http://www.nber.org/chapters/c3661>

Chapter pages in book: (p. 211 - 216)

gree (Ph.D., LL.B., and
ful individuals appear to

nination of the effects of
g conclusions. The 1969
presentation of the suc-
ty cells; hence, leaving
size, the 1969 respond-
hole Thorndike group.
correct pre-1946 educa-
ences for studying the
advantages of the larger
the measurement error,
r most of our purposes,
l as one of 8,000. In ad-
1969 that are important
ons we use the smaller
is.

Appendix F: Success and Other Biases

In this appendix we will examine the data to determine both if the people who responded in 1969 were more successful than those who did not and if the omission of the pre-1946 education data in the Thorndike files seriously affects our estimates of the effect of education on ability.

We begin our analysis with the data retained by Thorndike. After eliminating instances in which questionnaires were misplaced and in which income was zero, we have about 7,600 usable observations.¹ First, we consider the importance of the success response bias and the omission of the pre-1946 education information.²

The test to determine if there is a success bias in the 1969 sample follows. First, we assume that success generally persists over time, so that the successful in 1969 were also successful in 1955. Under the null hypothesis that there is no such bias, both the 1969 respondents and nonrespondents should represent random drawings from the same populations and should yield equations that are not significantly different. Chow's *F* test can be used to test the proposition that the two samples are drawn from the same population. Alternatively, since our equations use dummy variables to represent various levels of education and ability, our regressions yield an estimate of the mean (and standard deviation of) income in each category. We can use

¹The zero-income responses represent mostly "no answer," although there are approximately 70 students and 60 unemployed persons. Substantial numbers of medical doctors, managers, and farmers failed to report their income in 1955, although the zero-income respondents within a group were not related to education, except for M.D.'s, for whom there is only one education category.

²As noted earlier, the data retained by Thorndike do not contain information on those who terminated post-high school education prior to joining the Army.

standard tests to determine if the mean 1955 incomes in a given cell in the respondent and nonrespondent samples are significantly different. Since for the nonrespondents the only education information available is that retained by Thorndike, we use the 1955 education response for both groups. While this involves using data with a measurement error, there is no reason to expect the error (and the association bias) to be different for the two groups.

In Table F-1, we present an equation relating 1955 income to its determinants for all respondents in 1955, 1969 respondents, and 1969 nonrespondents.³ The Chow *F* test for the equation as a whole is less than 1; hence, we cannot reject the null hypothesis that the same relationship holds for the 1969 respondents and nonrespondents. We conclude, therefore, that there is no overall success bias. On the other hand, there may be a suc-

³We discuss in detail below the interpretation of, and conclusions to be drawn from, the estimates of the coefficients.

TABLE F-1
Equations to test for success bias and importance of pre-1946 education with 1955 income data: monthly income (in dollars)

	Constant	Some college	Undergraduate degree	Some graduate work	Master's	Ph.D and LL.B.†
(1) All 1955 respondents*	\$336.1 (17.2)	\$19.1 (2.0)	\$46.2 (5.0)	\$ 37.0 (2.2)	\$46.2 (2.9)	\$106.0 (6.0)
(2) Nonrespondents in 1969*	333.8 (11.9)	19.0 (1.4)	51.7 (3.8)	15.0 (.6)	51.0 (1.9)	155.9 (5.7)
(3) Respondents in 1969, E as given in 1955	337.3 (12.3)	16.3 (1.2)	38.7 (3.1)	53.5 (2.5)	38.5 (1.9)	59.6 (2.5)
(4) Respondents in 1969, E in 1955 but corrected for pre-1946	290.1 (10.7)	61.9 (4.9)	90.4 (6.8)	103.5 (4.7)	87.7 (4.2)	112.0 (4.7)

*Excludes those with zero income or missing questionnaires.

†Includes M.D.'s.

NOTE: In each equation the quintiles for the four factors and the biography variables were included, but the coefficients are not presented. Figures in parentheses are *t* statistics. Mean incomes for each education cell are calculated by adding the constant and the coefficient on the dummy together and then adjusting for different time-on-the-job coefficients.

cess bias we test groups (cantly lo pondent the less due to su of high We tu because tain pre data, we their po were jus sponses

†The data no longer

1955 incomes in a given
 dent samples are signifi-
 ondents the only educa-
 ed by Thorndike, we use
 groups. While this in-
 error, there is no reason
 bias) to be different for

relating 1955 income to
 1955, 1969 respondents,
 F test for the equation as
 cannot reject the null
 olds for the 1969 respon-
 e, therefore, that there is
 and, there may be a suc-

and conclusions to be drawn

cess bias associated with a particular level of education, which
 we test for by comparing mean incomes in the education
 groups (holding ability constant). Our tests indicate a signifi-
 cantly lower mean income for three-year-graduate-degree res-
 pondents than for nonrespondents, indicating a tendency for
 the less successful Ph.D.'s and LL.B's to respond. This may be
 due to successful Ph.D. holders' being difficult to locate because
 of high mobility.

We turn next to the effect of the measurement error that arises
 because the questionnaires retained by Thorndike did not con-
 tain pre-1946 education.⁴ Of course, in the Thorndike-Hagen
 data, we cannot distinguish between people who completed
 their post-high school education prior to 1946 and those who
 were just high school graduates. We can correct the 1955 re-
 sponses on the basis of the information available in the

⁴The data were available to Thorndike and Hagen in Air Corps records that are
 no longer extant.

Some college	Undergraduate degree	Some graduate work	Master's	Ph.D. and LL.B. ¹	M.D.	Teacher	Time on job	Sample size	R ² /S.E.
\$19.1 (2.0)	\$46.2 (5.0)	\$ 37.0 (2.2)	\$46.2 (2.9)	\$106.0 (6.0)	\$208.6 (6.1)	\$-148.2 (8.2)	\$13.8 (7.7)	7,618	.055 279
19.0 (1.4)	51.7 (3.8)	15.0 (.6)	51.0 (1.9)	155.9 (5.7)	79.7 (1.4)	129.8 (4.2)	13.4 (5.1)	3,873	.035 241
16.3 (1.2)	38.7 (3.1)	53.5 (2.5)	38.5 (1.9)	59.6 (2.5)	302.6 (7.2)	155.9 (7.1)	14.8 (5.9)	3,743	.080 266
61.9 (4.9)	90.4 (6.8)	103.5 (4.7)	87.7 (4.2)	112.0 (4.7)	298.2 (7.1)	-163.4 (7.5)	17.2 (7.0)	3,743	.090 264

naires.

and the biography variables were
 res in parentheses are *t* statistics.
 adding the constant and the coeffi-
 erent time-on-the-job coefficients.

1969-1970 reinterview of the 500 people for whom post-high school education was completed before 1946.⁵ In equation 4 of Table F-1, we present the results for the 1969 respondents using such an adjusted education measure. The effects of the missing education data are quite large. One not unexpected result is that the \bar{R}^2 in equation 4 is higher—by .01—than that in equation 3. Of more interest are changes in the mean-income estimates as presented in Table F-2. The constant term, which is an estimate of the income earned by a high school graduate in the lowest fifth of ability and biography and zero years on the job, declines by \$47 per month when the correct education data are used. The average income for the other education groups for zero time on the job is almost unchanged except for a \$22 per month increase at the Ph.D. level. However, a more meaningful comparison

⁵It is not possible to use the 1969 education responses directly in estimating the equation for 1955 income because some education was completed after 1955, and for those cases it is not currently possible to determine education obtained prior to 1955.

TABLE F-2
Average monthly earnings in 1955 for the lowest fifth before and after correcting for pre-1946 education (in dollars)

	Pre-1946 education data missing		Pre-1946 education data used	
	Time on job = 0	Average age = 33	Time on job = 0	Average age = 33
High school	337	470	290	444
Some college	353	471	352	490
College degree	375	479	380	501
Some graduate work	386	475	393	506
Master's	375	464	378	481
Ph.D. and LL.B.	380	454	402	488
Teacher with college degree	219	323	217	338
M.D.	683	756	685	786

NOTE: The time-on-the-job variable used in these regressions was based on postwar experience. Many people in the sample who were 33 in 1955 had college training before the war. Thus, we used the following figures for time on the job: high school, nine years; some college, eight; college degree, seven; some graduate work and master's, six; and Ph.D., five. Changing these years by one or two would not affect the comparisons across rows, though it obviously would affect the between-row comparisons. Better estimates are derived in Chapter 5.

SOURCE: Equations 3 and 4, Table F-1.

require
job due
ucation
we take
ings ri
Because
in other
signific
informa
are near
As in
to base
1969.

ple for whom post-high
 e 1946.⁵ In equation 4 of
 1969 respondents using
 he effects of the missing
 unexpected result is that
 than that in equation 3.
 an-income estimates as
 m, which is an estimate
 graduate in the lowest
 ears on the job, declines
 ation data are used. The
 groups for zero time on
 \$22 per month increase
 meaningful comparison

requires the inclusion of the effect of differences in time on the job due to education. We find that those with a high school education only have an average monthly income of \$26 less when we take account of the pre-1946 education, while average earnings rise by about \$20 in the other education categories. Because of the decrease in high school incomes and the increase in other incomes, the return to education is greater and more significant (in equation 4, Table F-1) when we use the pre-1946 information. Except for time on the job, the other coefficients are nearly the same.

As indicated in the text, these tests indicate that is is better to base our analysis solely on the people who responded in 1969.

uses directly in estimating the
 on was completed after 1955,
 determine education obtained

*Pre-1946 education
 data used*

<i>Time on job = 0</i>	<i>Average age = 33</i>
290	444
352	490
380	501
393	506
378	481
402	488
217	338
685	786

ons was based on postwar expe-
 college training before the war.
 high school, nine years; some
 id master's, six; and Ph.D., five.
 mparisons across rows, though
 Better estimates are derived in

Ap

This a
dike a