Education data in the NLSY79: a premiere research tool

Social science researchers widely use the NLSY79 schooling data because of its longitudinal nature and range of content

Kenneth I. Wolpin

Perhaps the most widely used data in social science research are those related to measures of education; among such measures, years of schooling is the most ubiquitous. A search of the National Longitudinal Survey (NLS) Annotated Bibliography yields 1,803 articles, book chapters, dissertations, and so forth, in which either the word "education" or "schooling" appears in the title, abstract, or as a keyword.[1] Of those, more than 1,000 were based on the National Longitudinal Survey of Youth, 1979 (NLSY79) data.[2]

Researchers' use of education measures found in the NLSY79 spans several social science disciplines, particularly economics and sociology, and, to a lesser extent, psychology. A large number of articles using NLSY79 education measures have appeared in major general audience and specialty journals. (See table 1.)[3] In economics, there were 8 such journals, totaling 78 published articles, and in sociology, 6 journals with 47 articles. In psychology, one journal specializing in child development published five articles, and one medical science journal also published five.[4]

The topics covered in these articles vary widely, as is evident from looking at the titles of the journals. These articles can be classified into two broad categories: (i) articles that study schooling decisions themselves (for example, how much schooling to complete, whether to drop out of high school, or choice of college major), and (ii) articles that study the "effect" of schooling on some other decision or outcome (for example, on wages, fertility, or alcohol consumption). In both cases, the NLSY79 data is chosen for its omnibus nature (that is, the data include information other than schooling), and because the data are longitudinal.

Although the NLSY79 was a pathbreaking survey in many ways, the collection of schooling data has been relatively standard. The education section in 1979, the first year of the survey, contained a total of just 25 questions (excluding interviewer check items), although respondents were attending school levels ranging from junior high through college.[5] In the first follow-up survey in 1980, respondents were asked just 16 questions, primarily to update the baseline schooling information obtained in the 1979 survey round. Essentially, the same questions were asked in the second follow-up in 1981, with the important addition of a monthly attendance record obtained retrospectively back to January 1980. This addition was in keeping with the event history format of the NLSY79 with respect to the collection of employment data and was continued throughout all subsequent rounds. The schooling section was unchanged until 1984, when an extensive set of questions was added to obtain information on colleges attended by respondents. Again, in keeping with the overall format of the survey, the aim was to obtain an event history of college attendance. The rostering of colleges attended continued prospectively through 1990, when the young-

Kenneth I. Wolpin is a professor at the University of Pennsylvania Department of Economics. E-mail: wolpink@ssc.upenn.edu

Table 1. Number of articles published since 1990 with "education" or "schooling" as a keyword in the title or in the abstract: journals with five or more articles by field

Field / journal	Number of articles
Economics:	
American Economic Review	10
Economics of Education Review	7
Industrial and Labor Relations Review	11
Journal of Human Resources	18
Journal of Labor Economics	7
Journal of Political Economy	5
Monthly Labor Review	10
Review of Economics and Statistics	13
Sociology, demography:	
American Sociological Review	6
Demography	8
Family Planning Perspective	8
Journal of Family Issues	6
Journal of Marriage and the Family	10
Social Forces	9
Psychology:	
Child Development	5
Others:	
Pediatrics	5

est respondents were age 25. The major elements of the schooling data in the main survey include current school enrollment status, highest grade attended and completed, high school curriculum, major field of study in college, degrees obtained, names and locations of colleges attended, college loans, and the schooling of household members.

Along with the main survey, there are several supplemental data collections pertinent to schooling. The School Survey, given in 1980, collected information from school administrators on school characteristics. The High School Transcript Survey collected and coded high school transcripts for almost 9,000 of the respondents. Finally, the Children of the NLSY79, begun in 1986, collected child development data for all children born to the women of the NLSY79.

It is beyond the scope of this article to summarize the findings from the studies cited above; instead, some data are presented that fit with the purpose of this special issue, that of celebrating the 25th anniversary of the NLSY79. In that vein, the data presented illustrate the aforementioned two most important features of the survey, the longitudinal nature and the range of content. These features are highlighted by documenting differences among completed schooling groups in a wide array of behaviorally-related characteristics. (See table 2.) Cumulative differences are shown for many characteristics over a substantial post-schooling age range. The table makes clear the rationale for the intensive study of schooling by social scientists, and for the popularity of using the NLSY79 data.

In the table, schooling is based on completed years of schooling as of age 25 and is divided into five categories: high school dropouts (completed years of schooling reported to be less than 12) who have not received a GED, high school dropouts who have received a GED, high school graduates (completed years of schooling exactly 12, no GED), those with some college (completed years of schooling between 13 and 15), and college graduates (completed years of schooling 16 or more). The table, based on data through the 2000 survey, considers differences in earnings, welfare takeup and payments, unemployment, hours worked, financial asset holdings, health status, obesity, alcohol consumption, fertility, and birth weight and test scores of the first-born child.

About 20 percent of males and 17 percent of females in this cohort did not complete high school with a regular diploma; of those, about a third received a GED. Notice that counting a person with a GED as a high school graduate makes a nontrivial difference to the calculation of the dropout rate.[6] A little less than 40 percent graduated high school with a diploma and did not complete any further schooling. The rest, about 40 percent of the cohort, went on to college, and about half of those graduated.

The cumulative figures in the table span ages from 25 through 39. However, many of the variables are not available in every year. Thus, cumulative variables, such as total earnings or weeks of unemployment, would be understated. Averaging available data and multiplying by 15 would provide an unbiased estimate except that missing values tend to occur more frequently at the older ages, both because of attrition and because the survey became biannual after 1994, and many of the variables are age-trended. To obtain an estimate of the cumulative values, averages of three 5-year age intervals (ages 25–29, 30–34, and 35–39) were calculated, each average was multiplied by 5, and then the three totals were added together. As long as the age trends are not severe within each of the 5-year age intervals, this procedure should provide a reasonable estimate.

The most striking feature of the table is how different high school dropouts, particularly those without GED's, and college graduates are relative to the rest. Following is a breakdown of each characteristic in turn.

Earnings. There are two sources of information on earnings in the NLSY79: a global question on the amount of wage and salary income in the previous calendar year, and earnings obtained from the event history on reported jobs. The figures in the table come from the former source and are in 1996 dollars.[7] Over the 15-year period between the ages of 25 and 39, a male high school dropout without a GED earned, on average, $273,250, about $30,000 less than a high school dropout with a GED, and $150,000 less than a high school graduate with a regular diploma. On the other end of the edu-

Table 2. Selected characteristics by completed schooling

Characteristic	Males					Females				
	High school dropout		High school graduate	Some college	College graduate	High school dropout		High school graduate	Some college	College graduate
	No GED	GED				No GED	GED			
Percent	13.7	6.2	37.9	20.4	21.8	11.5	5.4	39.7	23.2	20.3
Cumulative earnings[1]										
Age 25–39	$273,250	$305,041	$429,813	$486,619	$775,206	$113,521	$187,214	$207,738	$284,859	$426,855
Cumulative welfare payments[1]										
Age 25–39	$7,004	$7,185	$2,447	$1,216	$346	$28,032	$16,477	$8,198	$4,285	$271
Percent of years receiving welfare										
Age 25–39	13.3	10.6	4.5	2.8	.5	31.0	19.7	10.7	5.4	.7
Cumulative weeks unemployed										
Age 25–39	71.9	60.3	36.5	31.0	11.3	53.3	47.8	33.7	24.0	12.2
Cumulative hours worked: full-time (2,080 hours) year equivalents										
Age 25–39	13.1	12.7	15.6	15.3	16.6	7.3	9.6	10.4	11.2	12.2
Change in financial assets between 1985 and 2000										
10th percentile	–543	–814	–1,357	–2,360	–2,474	–87	–679	–1,368	–2,035	–950
50th percentile	0	0	977	1,928	19,347	0	0	467	1,868	12,927
90th percentile	4,671	7,865	31,805	55,119	268,212	3,661	8,517	20,642	67,953	178,281
Percent reporting health limitation, age 25–39										
At least once	36.4	32.2	21.1	18.4	9.0	51.6	36.1	36.0	31.7	21.1
At least 50 percent of the years	7.4	8.1	3.1	3.2	1.0	10.1	9.8	4.4	4.8	1.9
Percent obese (BMI greater than or equal to 30), age 25–39										
At least once	35.7	25.8	33.2	29.7	18.7	39.3	30.7	32.2	27.5	17.8
At least 50 percent of the years	16.3	12.2	17.8	16.2	8.2	26.1	18.7	17.1	14.7	6.2
Number of days per month having had six or more alcoholic drinks										
Age 18–24	1.97	1.78	1.70	1.40	1.39	.80	.59	.46	.40	.54
Age 25–34	1.21	.86	.98	.93	.61	.39	.55	.27	.17	.14
Number of children ever born, greater than or equal to age 35	1.9	1.9	1.6	1.4	1.5	2.6	2.3	1.9	1.7	1.6
Percent at first birth, less than or equal to age 19	23.9	23.7	8.8	4.2	1.1	68.4	55.6	26.1	11.1	.7
Birth weight of first born (ounces)	–	–	–	–	–	113.2	115.4	117.0	116.4	117.9
PPVT percentile score of first born	–	–	–	–	–	30	41	43	49	63

[1] 1996 dollars

NOTE: All figures use the 1979 sample weights. Dash indicates data not available.

cation spectrum, a male college graduate earned $775,206, almost $300,000 more than a male who completed some college. Female differences by education are smaller in absolute value, in part because they work less (see below). Nevertheless, a female high school graduate earned about $90,000 more than a dropout without a GED, and about $75,000 less than what a female who completed some college earned. A female college graduate earned $426,855, about $140,000 more than a female with some college.

Welfare payments and takeup. The NLSY79 collects data on takeup and payments for several welfare programs. The figures in the table aggregate respondent information about the Aid to Families with Dependent Children (AFDC), Food Stamp, and Supplemental Security Income (SSI) programs. Because of AFDC, aggregate welfare payments and takeup are much larger for females than for males. Over the 15 years, female high school dropouts without a GED received welfare on average in about 30 percent of the years, totaling $28,032. In contrast, high school dropouts with a GED received welfare in about 20 percent of the years, totaling $16,477. Comparable figures for high school graduates were about 11 percent of the years and $8,198, and for those with some college, 5 percent of the years and $4,285. Female college graduates, on average, essentially received no welfare.

Perhaps somewhat surprisingly, as seen in the table, even female high school dropouts without a GED had considerably more market earnings over the period than they received in welfare. In fact, only 35 percent of these women reported receiving more in welfare payments than they earned over the period. In contrast, 16 percent of high school dropouts with a GED, 10 percent of high school graduates, and 4 percent of those with some college received more welfare over the 15 years than they earned.

The takeup rates and average payments for males are, as noted, considerably smaller than for females; for example, male high school dropouts without a GED received some welfare in only 13 percent of the years, about the same as those with a GED. Welfare payments over the entire 15-year period were, on average, only about $7,000 for high school dropouts regardless of GED status, and only 7 percent of dropouts without a GED received more in welfare benefits than they earned.

Labor force status (unemployment and hours worked). Male high school dropouts without a GED were unemployed 72 weeks, and similar female dropouts were unemployed 53 weeks, cumulatively over the 15-year age period.[8] These figures are slightly higher than for dropouts with a GED. Those who graduated from high school or completed some college had significantly fewer cumulative weeks of unemployment. However, as with earnings and welfare, college graduates were distinctly different than the other schooling groups; male (female) college graduates spent only 11 (12) weeks unemployed over the 15 years. With respect to the amount of time spent working, measured in full-time year equivalents (2,080 hours), on average male dropouts (those with and without a GED) worked about 13 of the 15 years, while male college graduates worked more than full-time, 16.6 years or about 2,300 hours per year. In contrast, female dropouts without a GED worked less than one-half of the years, while college graduates worked four-fifths of the years.

Financial asset accumulation. The NLSY79 began collecting asset data in 1985. Table 2 presents statistics on the amount of financial assets accumulated from the first report (when respondents were 20 to 28 years of age) to the 2000 report (when respondents were 35 to 43).[9] Differences in financial savings by schooling are large, although the relationship is quite skewed. Median financial savings are zero for high school dropouts of both sexes, slightly negative at the 10th percentile, and less than $10,000 at the 90th percentile. Median savings is less than $2,000 for those with some college, but savings increase to $55,000 for males and $68,000 for females at the 90th percentile. College graduates are again different than other schooling groups in terms of their financial savings. Median financial savings of males is almost $20,000, but more than $250,000 at the 90th percentile; for females, the median is about $13,000 and the 90th percentile, $178,000.

Health. In each round of the NLSY79, respondents were asked separately about whether their health limited the amount or the kind of work they could perform. Table 2 presents data on the distribution of the number of times between the ages of 25 and 39 that the respondent answered affirmatively to either question.[10] The first row reports on the percent who gave an affirmative response at least once, and the second row, on the percent who gave an affirmative response in at least 50 percent of the years. High school dropouts without a GED are the most prone to report a health limitation affecting work at least once; more than a third of the men and more than half of the women reported ever having a limitation. At the other extreme, only 9 percent of male college graduates, and 21 percent of female college graduates, reported such an occurrence at least once. Chronic health conditions, as measured by having reported a health limitation in at least half of the years, are much less prevalent in all schooling groups, but non-negligible for high school dropouts. Between 7 percent and 8 percent of male high school dropouts, and about 10 percent of female high school dropouts, have a persistent health condition, while this is true of only 1 percent of male and 2 percent of female college graduates.

Obesity. The NLSY79 collected self-reported information on a respondent's height and weight in 1981, 1982, and 1985,

and on the respondent's weight in all rounds since 1986, except 1987 and 1991. Body mass index (BMI) was calculated in each year using the latest height information that was available.[11] As with health limitations, table 2 reports the distribution of the number of times over the 15-year age period, 25–39, that respondents have a BMI of 30 or more, indicating obesity, again by schooling groups for each sex. Although there is some prevalence of obesity among males with schooling below college completion, it is not large. Among males with some college, 30 percent were obese in at least 1 year, and 16 percent were obese in at least one-half of the years. The same figures for high school dropouts are 36 percent and 16 percent. Differences are larger for females, with 28 percent of those with some college, and 39 percent of high school dropouts without a GED, being obese in at least 1 year. College graduates again are substantially different from all of the other groups. For both sexes, less than 20 percent were obese in any year, and less than 10 percent in at least half of the years.

Alcohol consumption. In several years (1982, 1983, 1984, and 1994), the NLSY79 has included a section of questions, funded by the National Institute of Alcohol Abuse and Alcoholism (NIAAA), on alcohol use. A common question in those rounds has been the number of days in the last month that the respondent consumed six or more drinks. Table 2 shows the average number of days over all of the observations in those years for two age intervals, 18–24 and 25–34, by schooling and sex. In all of the schooling groups, for both males and females, alcohol consumption measured in this way fell substantially with age. Male high school dropouts without a GED averaged just less than 2 days a month between the ages of 18 and 24, but 30 percent less between 25 and 34. Females within that schooling group averaged only .8 days at the earlier age, but that fell in half at the later age. College graduates are less of an outlier in terms of alcohol consumption at the earlier age, although their consumption falls proportionately more at the later age.

Fertility. Differences by schooling in the number of children ever born by age 35 are large. Female high school dropouts who have not earned a GED have had 2.6 children on average by age 35, .3 more than their counterparts with a GED, .7 more than those with a high school diploma, .9 more than those with some college, and a full child more than those with a college degree. Males at all schooling levels report having had fewer children, and this is particularly pronounced for high school dropouts. Even so, those male respondents with some college report having fathered half a child less than those who were high school dropouts.

Schooling differences in fertility are even more striking in terms of the prevalence of teenage childbearing. More than two-thirds of female high school dropouts without a GED, and more than one-half of those with a GED, gave birth to a child as a teen. This figure drops to one-quarter for high school graduates, to about 10 percent for those with some college, and to less than 1 percent for college graduates.

Birth weight of children. Starting with the 1982 survey round, with funding from the National Institute of Child Health and Human Development (NICHD), the NLSY79 collected information about the pregnancy outcomes of the female respondents.[12] Table 2 compares mean birth weight of first-born children across the schooling groups.[13] The differences observed among the respondent's own outcomes, such as earnings and health, also emerge among the children of the respondents. High school dropouts without a GED have the lowest birth weight of children, almost 5 ounces less than the first births of the college graduates.

Peabody Picture Vocabulary Test (PPVT) scores of children. The Children of the NLSY79 survey was started in 1986 and obtains information about the children born to the women of the NLSY79, including cognitive test scores. Table 2 reports the mean percentile scores (age adjusted) of first-born children by mother's schooling. As seen repeatedly in the table, the lowest and highest schooling groups, high school dropouts without a GED and college graduates, appear to be outliers. Children of the former group have mean scores that fall in the 30th percentile, while those of the latter fall in the 63rd percentile. The other three schooling groups fall in between and differ only slightly from each other.

THIS OVERVIEW analyzes the contributions of the NLSY79 to the study of issues related to education. Space limitations, combined with the vastness of the literature, precluded a critical assessment of the scientific contributions of the research based on NLSY79 schooling data. Instead, this article demonstrates why the schooling data has played such an important role in making the NLSY79 one of the premiere research instruments in the social sciences. □

Notes

[1] Education alone yields 1,600 matches, schooling alone 588, with 385 matches using both. A search for the words "earnings" or "wages" yielded 1,437 matches, and for "income," 1,027 matches. All of these figures are somewhat overstatements because of duplications that arise when papers appear in multiple forms, for example, as a working paper and later as a journal publication.

[2] When referring to the NLSY79, I include the Children of the NLSY79. Recall also that the NLS consists of five additional cohorts, four begun between 1966 and 1968, and the latest begun in 1997.

[3] I thank Terry Fahey at the Center for Human Resource Research (CHRR) for performing the specialized search of the Bibliography's data-

base on which these figures are based.

[4] This is the only psychology journal publishing at least five such articles. There are about a dozen other psychology journals that have published fewer than five, though most have published only one article.

[5] In the NLSY97, on the other hand, the schooling section contained several hundred questions in the first round. Schooling data in the 1997 survey are obtained as a dated event history similar to the employment data.

[6] The Current Population Survey (CPS) counts both those with a GED and those with a regular diploma as a high school graduate.

[7] The samples for the earnings and welfare statistics are both based on respondent-years in which there is no missing information either on earnings or on welfare benefits. In addition, the sample is restricted to those who had at least one report in each 5-year age interval.

[8] As with earnings and welfare, the sample is restricted to those who have a valid report in at least 1 year of each 5-year age interval.

[9] Financial assets consist of moneys in savings or checking accounts, money market funds, credit unions, U.S. savings bonds, individual retirement accounts, 401K or pre-tax annuities, certificates of deposits, personal loans to others or mortgages held, stocks, bonds, mutual funds, and rights to an estate or investment trust.

[10] The sample is based on respondent-years in which there is no missing information on the health limitation question. The average numbers of years over the 25–39 age range in which these data are reported for each schooling category are: 10.3, 10.0, 10.5, 10.4, and 10.3 for males, and 10.9, 10.6, 10.7, 10.8, and 10.7 for females.

[11] More specifically, the 1985 height report was used whenever available (about 90 percent of the cases). If that was missing, the 1982 height report was used, and if both were missing, the 1981 height report was used. Differences in reported height in the 3 years are small; mean reported height in 1981 is .17 inches less than in 1985 for those that report both. The BMI is defined as weight in kilograms divided by the square of height in meters. Obesity is defined as having a BMI of 30 or more.

[12] Information was also collected on maternal behaviors during each pregnancy, such as prenatal care, smoking, drug use, and drinking.

[13] Looking at first-born children avoids differences that might arise if birth weight is related to birth order, given that less educated women have more children as shown above. A similar comment applies to the Peabody Picture Vocabulary Test scores addressed in the following section.