HS&B:80 Common Variables Public Dataset

Dataset Version: 2025-10-30

Introduction

EdSHARe's High School & Beyond Study of 1980 (HS&B:80) common variables public dataset gathers together several of the most common variables (derivable from the HS&B:80 public data) that we use in EdSHARe analyses. To the extent possible, the definitions of these variables are parallel to those in EdSHARe's National Longitudinal Study of the Class of 1972 (NLS:72) common variables public dataset.

The operationalizations used for this dataset constitute default choices, but you should always consider whether the given operationalization is most appropriate for your specific analysis. If your analysis requires a different operationalization, go back to the original data and construct your own variable, as appropriate for your analysis. Refer to the code for the HS&B:80 common variables public dataset as a starting point.

Constructed Variables

Parental Education

Father's (BB039, FY55) and Mother's (BB042, FY56) education were recoded into the standard four categories:

Original Coding	New Coding
1 NOT LIVE WITH FTHR/MTHR	(Missing)
2 LT HS GRAD	1 HS or less
3 HS GRAD ONLY	1 HS or less
4 LT 2 YRS VOC, TECH	2 Some college
5 2 OR GT YRS VOC	2 Some college
6 LT 2 YRS COLLEGE	2 Some college
7 2 OR GT YRS COLLEGE	2 Some college
8 FINISHED COLLEGE	3 College
9 MASTER^S	4 Grad School
10 PH.D., M.D.	4 Grad School
11 DON^T KNOW	(Missing)

Then, the higher of father's and mother's education was taken in each of 1980 and 1982 (using the single known value if one parent's education was missing). The final variable (paredu) is taken from the 1980 observation, with missing values filled in from 1982.

High School Family Income

High school family income is based on the 7-/8-category student report (BB101, FY111). The 3-category variable fills in only a small proportion of cases beyond the finer grained variable. Note that the categories differ between the 1980 and 1982 variables, and that the 1982 variable is available only for sophomores.

Most categories were recoded with the midpoint of the given category, except for the highest two categories. Since it is open-ended, the highest category does not have a midpoint. To obtain a reasonable value for this category, the frequencies in each category were fit with a Normal distribution using the cut points indicated by the category ranges. Then, the median of each range was obtained from truncated Normal distributions using the fitted mean and standard deviation. The medians obtained from this procedure were fairly close to the range midpoints for most categories, so the rounded median was used for the upper category, as well as the penultimate category, which deviated somewhat from the midpoint for that range.

The final codings for the 1980 and 1982 family incomes are as follows:

Variable	Original Coding	New Coding
income80	1 UNDER $7000 2 $7000-$11999 3 $12000-$15999 4 $16000-$19999 5 $20000-$24999 6 $25000-$37999 7 $38000 OR MORE	3500 <$7k 9500 $7-12k 14000 $12-16k 18000 $16-20k 22500 $20-25k 30000 $25-38k 42000 $38k+
income82	1 $7,999 OR LESS 2 $8,000-$14,999 3 $15,000-$19,999 4 $20,000-$24,999 5 $25,000-$29,999 6 $30,000-$39,999 7 $40,000-$49,999 8 $50,000 OR MORE	4000 <$8k 11500 $8-15k 17500 $15-20k 22500 $20-25k 27500 $25-30k 35000 $20-40k 44000 $40-50k 55000 $50k+

The income variables may be used categorically or continuously, depending on the needs of the analysis. The default recommendation is to use 1980 income for analysis, but to include 1982 income as an auxiliary variable in the imputation model. The 1982 income variable should be conditionally imputed only for sophomores.

Number of Siblings

In 1980, students reported detailed information about the number and ages of their siblings (BB096A-E). In a number of cases, this resulted in abnormally large sibling counts. The 1982 questionnaire asked for the number of siblings in a single item (FY106/FE70). The 1982 report is preferred for the number of siblings, with missing values filled in using the 1980 total, capped at “6 or more” siblings to match the response categories from 1982.

High School Disability

Students reported their disabilities in both 1980 (BB087A-G) and 1982 (FY103A-G, FE83A-G). For seniors, we prefer the 1980 report, filling in missing values with the 1982 report. For sophomores, we prefer the 1982 report, filling in missing values with the 1980 report. Missing values were filled in for each item before constructing the aggregate indicator.

The learning disability indicator (learndisab) was taken directly from the "specific learning disability" item (A). The physical disability indicator (physdisab) was positive if a student indicated any disability <u>other</u> than a learning disability (item A) or a visual disability (item B).

The wording for the "visual handicap" item changed from 1980 ("visual handicap") to 1982 ("visual handicap (not corrected by glasses)"), with a substantial decrease in student endorsement. This seems to suggest that many students in 1980 may have been interpreting imperfect vision as a "visual handicap," while this item was probably intended to capture more substantial visual impairments, such as partial or full blindness.

Body Mass Index

Body mass index in 1980 (bmi80) was constructed from student self-reported height (BB085) and weight (BB086). BMI is body mass in kilograms divided by the square of body height in meters. The factor 703 converts from pounds per square inches to the BMI's units of kilograms per square meters. Students indicated their weight in 5-pound ranges; the midpoint of each range was used for the calculation, and students indicating a weight below 100 pounds or above 226 pounds were treated as missing.

Post-secondary Education

The HS&B:80 data on students' post-secondary educational experiences is complex and piecemeal, with most measures covering only specific time periods. Combined with varied patterns of non-response and the different meaning of post-secondary enrollment in a given data collection period across the sophomore and senior cohorts so shortly after graduation from high school, obtaining a coherent view of students' early educational attainment after high school is a challenge.

To assist with imputation of educational attainment, evidence of post-secondary educational experiences has been encoded in a suite of variables in the HS&B:80 common variables dataset (pse*). **Generally, you should <u>not</u> use these variables in your analysis**

directly. They are intended only to provide auxiliary data for imputation of a carefully constructed educational attainment variable (e.g., edattain21). If you would like to analyze early post-secondary educational experiences, you should instead wrestle with the original data to define a variable operationalization that is most appropriate for your analysis.

The post-secondary education variables each have the following coding, based on the evidence provided by the source data:

- 1 No post-secondary education
- 2 Some post-secondary education (enrollment)
- 3 Two-year degree attainment
- 4 Four-year degree attainment
- 5 Graduate degree attainment

Note that these variables only code for four-year and graduate degree attainment when they are explicitly evidenced by the source data—enrollment in a graduate degree is not taken as evidence of four-year degree attainment, for example.[1] Attainment of a certificate is coded as "some post-secondary education."

The post-secondary education variables are provided separately for each cohort to account for the different meaning of the evidence of post-secondary education shortly after high school. The coding of these variables allows them to be entered together in the same regression model for both cohorts, creating implied interaction terms with cohort.

For variables from the student surveys in 1984 and 1986, evidence of some post-secondary education was derived from items SE15/SY15 and TE20/TY20 (as well as item FE31 for the senior cohort in 1982). Evidence of degree attainment was derived from the reports of each post-secondary education institution attended (items SE/SY18-20 and TE/TY21-22), if a degree was completed (part J in 1984, part I in 1986) based on the degree type (part I in 1984, part H in 1986). Both vocational and academic programs were counted for two-year degrees. If a degree was completed but the type of degree was missing, a two-year, four-year, or graduate degree was assumed if the institution type (part A) was "junior college," "college/university," or "graduate/professional school," respectively.

[1] Graduate program enrollment is not frequently attested in the early waves of HS&B:80.

Educational Attainment

The educational attainment variable (edattain86) is constructed based on self-report at multiple points in time as of 1986 and uses the following coding:

- 1 HS or less
- 2 Some college (attendance/certificate)
- 3 Two-year degree attainment
- 4 Four-year degree attainment
- 5 Graduate degree attainment

Educational attainment as of 1986 (edattain86) is based on student's report of post-secondary education in HS&B:80/84 and HS&B:80/86 (pse84*, pse86*), each of which covered distinct periods of time. If neither of these variables is missing, edattain86 is taken to be the larger of the two. If either of these variables is missing, taking the non-missing value may potentially under-report attainment as of 1986. The variable edattain86 is similar to the HS&B:80 variable EDATTAIN, but differs in terms of the some college and two-year degree categories, as well as handling of missing values.

In general, the variable edattain86 should <u>not</u> be used in multiple imputation directly. This variable is provided as a convenience for analyses that do not use imputation. Instead, for analyses that make use of imputation, impute the variables pse84* and pse86*, and reconstruct edattain86 based on the imputed values. In this setup, the missing values for the pse* variables just mentioned should <u>not</u> be replaced with 0; but the remaining pse* variables (pse82sr, psetran*) <u>are</u> auxiliary, and the missing values of these auxiliary variables may be replaced with 0.

When imputing pse*so and pse*sr, the imputation must be conditional on the corresponding cohort. That is, replace these variables with 0 for the inapplicable cohort (e.g., replace pse84sr = 0 for the sophomore cohort) and include the `mi impute conditional` option to specify the corresponding cohort. This is illustrated in the EdSHARe multiple imputation template code in the population-representative estimation guide.

Once you have imputed the source variables, you may reconstruct edattain86 as follows:

```
drop edattain86
mi passive: egen edattain86 = rowmax(pse84so pse86so pse84sr pse86sr)
```

Recoded Variables

Variable	Source Variables	Original Coding	New Coding
hsregion	SCHREGN	1 New England 2 Mid-atlantic 3 South Atlantic 4 East South Atlantic 5 West South Atlantic 6 East North Central 7 West North Central 8 Mountain 9 Pacific	1 Northeast 1 Northeast 2 South 2 South 2 South 3 Midwest 3 Midwest 4 West 4 West
hstype	SCHSAMP	0 Regular Public School 1 Alternative School 2 Cuban Hispanic School 3 Other Hispanic School 5 Regular Catholic School 6 Black Catholic School 7 Cuban Hispanic Catholic 8 Elite Other Private School 9 Other Private School	1 Public 1 Public 1 Public 1 Public 2 Catholic 2 Catholic 2 Catholic 3 Private 3 Private
cohort	COHORT	1 SENIOR 2 SOPHOMORE	1 Senior 0 Sophomore
sex	SEX	1 MALE 2 FEMALE	1 Male 0 Female
race	RACE	5 WHITE 4 BLACK 1 HISPANIC OR SPANISH 2 AMERINDIAN 3 ASIAN, PCFC ISLNDR 6 OTHER	1 White 2 Black 3 Hispanic 4 Other 4 Other 4 Other
home80	BB102	1 OWN	1 Owned

		2 RENT 3 OTHER ARRANGEMENT	0 Not owned 0 Not owned
hsprog	BB002, FY2	1 GENERAL 2 ACADEMIC PREP 3 AGRICULTURE 4 BUSINESS, OFFICE 5 DISTRIBUTIVE ED 6 HEALTH 7 HOME ECONOMICS 8 TECHNICAL 9 TRADE, INDUSTRIAL	1 General 2 Academic 3 Vocational 3 Vocational 3 Vocational 3 Vocational 3 Vocational 3 Vocational 3 Vocational
alg1 alg2 geom trig calc physics chemistry biology	EB005A, FY5A EB005B, FY5B EB005C, FY5C EB005D, FY5D EB005E, FY5E EB005F, FY5F EB005G, FY5G EB005H, FY5H	1 YES, HAVE TAKEN 2 NO, HAVE NOT TAKEN	1 (Yes) 0 (No)
highmath		Last of: !missing(alg1)|!missing(geom) alg1 == 1	geom == 1 alg2 == 1 trig == 1	calc == 1	 0 Below Algebra 1 1 Algebra 1 or Geom 2 Algebra 2 3 Trig or Calculus
gpa12 gpa10	BB007, FY6 BB007 (soph)	1 As Mostly 2 As & Bs 3 Bs Mostly 4 Bs & Cs 5 Cs Mostly 6 Cs & Ds 7 Ds Mostly 8 Below D	4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.0
marital86	TY41, TE41	5 MARRIED 6 SINGLE LIKE MARRIED 2 DIVORCED 3 WIDOWED 4 SEPARATED	1 Married 1 Married 2 Previously married 2 Previously married 2 Previously married

		1 NEVER MARRIED	3 Never married

Additional Variables

Variable	Source Variables	Codes
RAWWT	RAWWT	
SCHSAMP	SCHSAMP	0 REGULAR PUB SCHOOL 1 ALTERNATIVE SCHOOLS 2 CUBAN HISPANIC PUBL 3 OTHER HISPANIC PUBL 5 REGULAR CATHOLIC 6 BLACK CATHOLIC 7 CUBAN HISPANIC CATH 8 ELITE OTHER PRIVATE 9 OTHER PRIVATE
SCHREGN	SCHREGN	1 NEW ENGLAND 2 MIDDLE ATLANTIC 3 SOUTH ATLANTIC 4 EAST SOUTH CENTRAL 5 WEST SOUTH CENTRAL 6 EAST NORTH CENTRAL 7 WEST NORTH CENTRAL 8 MOUNTAIN 9 PACIFIC
hsurb	SCHURB	1 Urban 2 Suburban 3 Rural