Social Class Measure

This memo discusses the creation of a social class measure. Social class of the family of origin is one of the central concepts in social stratification research: students from more advantaged families are more likely to experience educational and occupational success (Blau and Duncan, 1967). This memo will address the following: 1) a cursory and abbreviated discussion of the concept of social class, 2) a review of the survey questions used to create the social class indicator, 3) and a suggested coding of the social class measures, and 4) descriptive statistics for the social class index.

Social Class ~ A Very Quick Conceptual Overview :

The role of the social class of the family of origin is well defined in the social stratification literature. Numerous studies have illustrated a positive association between social class of family of origin and later life outcomes, such as educational attainment, occupational attainment, and wealth/earning (Blau and Duncan, 1967; Jencks et al, 1972; Portes and Rumbaut, 2001; Sewell and Hauser, 1975). Social class is central variable as it has been shown to consistently operate directly on later life outcomes; however, it also operates indirectly through the socialization process and social networks (Lareau, 2003).

Admittedly, there are numerous ways in which one can note the social class of the student's family of origin. However, numerous analyses tend to use measures such as mother and fathers education, mothers and father occupational prestige scores, parental income (when available), and other indicators of familial wealth, such as home ownership. The UW-BHS survey includes questions regarding the parents' level of educational attainment, parental occupation, and whether the student's family owns the home in which they live.

Social Class Indicators in the UWBHS:

This section of the memo will review the questions included in the UW-BHS survey that will be used to create a social class index.

Family Home Ownership (Question 97):

97)   Does your family own or rent the home you live in?

Own (with or without mortgage) 1

Rent 2

Don't know..3

Syntax to code this variable into a usable binary variable which denotes students' whose families own their own home (coded '1') and those who do not or don't know (coded '0') is included at the end of the memo .

Parental Education Questions: *Note that the parental education question in 2000 differs slightly from the question in 2002 to 2005.*

2002-2005 question:

122_) What is the highest degree or level of school that he (your father or father figure) has completed?

(CIRCLE ONLY ONE – THE HIGHEST LEVEL COMPLETED)

Less than 1st Grade 1

1st, 2nd, 3rd, or 4th Grade 2

5th or 6th Grade 3

7th or 8th Grade 4

9th Grade 5

10th Grade 6

11th Grade 7

12th Grade no diploma 8

High School Grad-Diploma or Equivalent 9

Some College But No Degree 10

Associate Degree-Occupational/Vocational 11

Associate Degree/ Academic Program 12

Bachelor's Degree 13

Master's Degree (i.e. MA, MS)...14

Professional School (i.e. MD, DDS, DVM) 15

Doctorate Degree (i.e. PhD, EdD).......................................16

Not Applicable (No Father or Father Figure).....................17

Note that a separate question for mothers was asked (question 130) the only difference from this question is that it asks about the students' mothers' highest level of completed education.

2000 question:

122) What is the highest degree or level of school that your FATHER or male guardian has completed?

(MARK ONLY ONE – THE HIGHEST LEVEL COMPLETED)

No Schooling .... ○	11th Grade .... ○
1st Grade .... ○	High School Graduate or GED .... ○
2nd Grade .... ○	Some College, No Degree..................... ○
3rd Grade .... ○	2-Year or Community College Degree . ○
4th Grade .... ○	4-Year College Degree .... ○
5th Grade .... ○	Master's Degree or Equivalent.............. ○
6th Grade .... ○	Ph.D., M.D., or
7th Grade .... ○	Other Professional Degree.............. ○
8th Grade .... ○	
9th Grade .... ○	Not Applicable
10th Grade .... ○	(no male parent/guardian) .... ○

Note that a separate question for mothers was asked (question 130) the only difference from this question is that it asks about the students' mothers' highest level of completed education.

As you can see, the questions vary slightly between 2000 and later years in that the 2000 survey includes each of the years of schooling from 1st through 11th grade as unique responses. In the 2002 to 2005 many of these years are collapsed.

SPSS syntax to create parental education variables is included at the end of this memo.

Parental Occupational Socioeconomic Status Scores:
The creation of the maternal and paternal occupational socioeconomic status score variables is rather complex a separate memo which details their creation was constructed. You can access it.

We are interested in finding out about the kind of work that he (your FATHER or the person most like a father to you) does.

123) IN THE LAST MONTH, did he (your father or father figure) do any work for either pay or profit?

(CIRCLE ONLY ONE)

Yes 1

Not applicable (no male parent/father figure)..............2 } SKIP TO QUESTION

No ....3 } 129_a ON PAGE 28

If he (your father or the person most like a father to you) had more than one job, describe the one at which he worked the most hours. If he had no job or business in the last month, give the information for his last job or business.

124) **For whom did he (your father or father figure) work?** If now on active duty in the Armed Forces, print the branch of the Armed Forces.

_____ (Name of Company or Business)

125) **What kind of business or industry was this? Describe the activity at location where employed. (For example: hospital, newspaper publishing, mail order house, auto repair shop, bank)**

_____

126) Is this mainly –

Manufacturing? 1

Wholesale trade? 2

Retail trade? 3

Other (agriculture, construction, service, government, etc.)? 4

127) **What kind of work was he (your father or father figure) doing?** (For example: registered nurse, personnel manager, supervisor of order department, auto mechanic, accountant)

_____

128) **What were his most important activities or duties?** (For example: patient care, directing hiring policies, supervising order clerks, repairing automobiles, reconciling financial records)

_____

129) **Was he (your father or father figure)–**

Employee of a private company or business?.......................................1

Government employee? 2

Self-Employed?...3

Working without pay in a family business or farm? 4

Note that a separate set of question for mothers' occupation was asked (question 131 to 137) the only difference from this question is that it asks about the students' mothers' occupational information.

Again for an explanation of how the parental occupational information was transformed into occupational socioeconomic status score memo.

Operationalization of Social Class:
There are multiple ways in which one can operationalize social class. For example, one can use home ownership, fathers' education, and fathers' occupational SES score as separate indicators of social class. Alternatively, one could use home ownership, maximum level of education achieved by either parent, and the maximum occupational SES score of either parent. The advantage of utilizing each of the measures separately is that one can examine how each of them independently operates on the outcome of interest. However, a drawback to utilizing these measures independently is that they require additional degrees of freedom. Additionally, if one is estimating models with smaller samples (e.g. stratifying data by generational status or race/ethnicity) or if they are estimating interaction effects (social class measures and another predictor variable), a singular social class indicator may be more efficient and it may allow for easier interpretation.

One rather straightforward way to make a singular indicator of social class is to use factor analysis to extract the underlying factor that explains the most variance across the social class measures (mothers and fathers education, mothers and fathers occupational SEI score, and familial homeownership). The SPSS syntax to create a social class indicator using factor analysis is included below.

When the aforementioned social class measures are included in a factor analysis, the first factor has an Eigenvalue of 2.551 and it explains 51% of the cumulative variance across the social class measures. The other 4 factors are not extracted as they have Eigenvalues below one and they each explain less than 18% of the variance. All of the measures appear to hang together

relatively well which is indicated by the factor loading scores for each of the variables (below). Admittedly, the score for familial homeownership is lower than one would desire, but it is adequate. The mean for the factor score is 0 and the standard deviation is .898.

Factor Matrix[a]

	Factor
	1
iownhome imputed, family owns home-1 yes, 93 are imputed	.312
imothedu imputed mother parental education, q130	.767
ifathedu impued, father education q122	.759
isfocced	.623
ismocced	.613

Extraction Method: Principal Axis Factoring.

a. 1 factors extracted. 6 iterations required.

Although there are a lot of positive aspects to using a singular measure of social class, one potential drawback is the number of missing cases. The number of cases missing on the parental occupational SEI measures is rather high (3,133/3400 missing on father/mother SEI measure), as students did not complete these questions if they did not have a mother or father in their life. Further, for the students that had both parents in their life, coding the students' responses was difficult in that the students did not always know enough about their parents' jobs to allow for the assignment of an occupational code. Lastly, the high number of missing cases is partially due to the fact that students whose father or mother had not worked for pay or profit in the last month were instructed to skip the parental occupational questions. Thus, if one opts to create a singular factor score for SES they should consider imputing missing values for the inputs or imputing values for the factor score after it is created, as the number of valid cases, using listwise deletion, is 4,737.

SPSS syntax to create a social class indicator using factor analysis is included below. .

Descriptive Statistics for the Social Class Index:

This section of the memo compares descriptive statistics for social class index using listwise deletion and an index in which the missing values are imputed. The memo will first examine the descriptive statistics and the density plot for the index that uses listwise deletion.
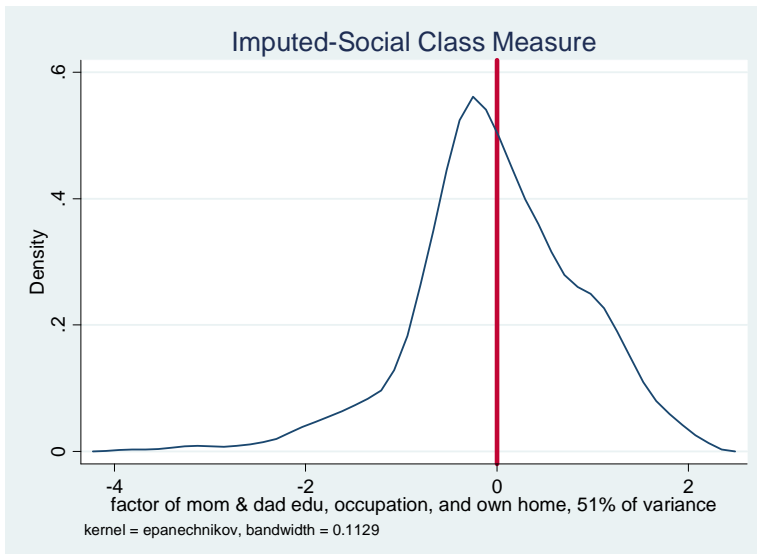
Density Plot:

Social Class Factor

factor of mom & dad edu, occupation and own home, 51% of variance
kernel = epanechnikov, bandwidth = 0.1487
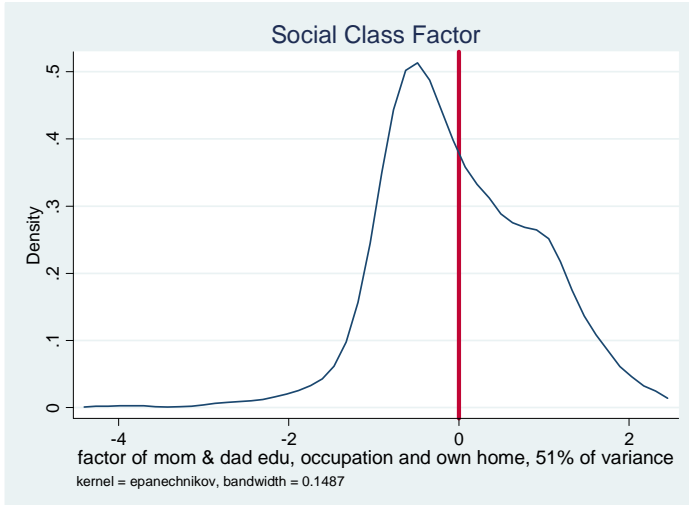
```
     Variable |        Obs        Mean     Std. Dev.         Min         Max
-------------+---------------------------------------------------------------
          ses |       4737     5.85e-11     .8977652    -4.256303    2.300696
--4,921 cases, 51% of valid cases, are missing on this measure.
```

The social class measure appears to be almost normally distributed around a mean of (nearly)
zero. The density plot and the min and max values indicate that it is slightly left skewed.
As most analyses will impute values this memo includes the density plot and descriptive
statistics for an SES measure that is imputed, using single regression imputation.

Imputed-Social Class Measure

factor of mom & dad edu, occupation, and own home, 51% of variance
kernel = epanechnikov, bandwidth = 0.1129

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         ses |      9649     2.63e-10    .8786643   -4.112996   2.374859
```

The descriptive statistics and the density plot are relatively similar for the non-imputed and the imputed measures.

Mean Value on Source Variables for SES Measure by SES Quartiles (using imputed values)					
SES Quartiles	Family Owns Home	Mother's Education	Father's Education	Mother's SEI Score	Father's SEI Score
High	.91	16.4	17.4	81.0	81.8
Mid to High	.78	13.8	14.2	66.0	62.1
Mid to Low	.66	12.6	12.8	53.5	47.4
Low	.45	8.4	11.3	32.5	35.1

To check that the SES index is correctly capturing the source variables, the table above illustrates the mean values for the source variables across SES quartiles. As expected the mean values on the source variables increase as one moves from low to high on the SES index.

Citations

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

Jencks, Christopher, Marshall Smith, Henry Alcand, Mary Jo Bane, David Cohen, Herbert Gintis, Barbara Heyns, and Stephan Michelson. 1972. *Inequality: A Reassessment of the Effect of Family and Schooling in America* New York: Basic Books

Lareau, Annette. 2003. *Unequal Childhoods: Class, Race, and Family Life.* Berkeley, CA: University of California Press.

Portes, Alejandro and Ruben Rumbaut. 2001. *Legacies: The Story of the Immigrant Second Generation.* Berkeley: University of California Press.

Sewell, William and Robert Hauser. 1975. *Education, Occupation, and Earnings.* New York: Academic Press.

Syntax:

Own home
Below is SPSS code to create a binary variable which converts s097 from the student survey into a binary variable which indicates whether the respondent's family owned their home. A value of 1 on the ownhm variable would indicate the student's parents owned their home, while a value of zero indicates the student's parents did not own their home.

```
***family owns home***.
missing values s097 (-99797 thru -999) .
compute ownhm = 999.
if s097 eq 2 ownhm = 0.
if s097 eq 3 ownhm = 0.
if s097 eq 1 ownhm = 1.
variable label ownhm 'family owns home, 1 is yes'.
missing values ownhm(999).
```

Comment [npc1]: This was -999 in own home memo

Parental Education Variables
Because of the change in response categories between 2000 and 2002, care must be taken when creating variables for parental education. We have created the following three variables that are equivalent between the years. The code used to create these variables can be found at the end of this document.

We will create three recoded variables measuring maternal and paternal education. The three variables are 1) a categorical measure distinguishing between commonly completed levels of educational attainment (maedcata, paedcata), 2) years of schooling completed using the midpoint of the response categories (paed, maed), and 3) years of schooling completed using the highpoint of the response categories (paedh, maedh).

Please note: Sometimes students would circle two adjacent answers. We believe that this was to indicate that they were not sure which of the two answers was correct. We coded these answers as the average of the two answers. For example, if in 2002, 7 and 8 were circled, we would code the answer as 7.5. In creating the variables that are consistent between the years, we first round the variables so that they become integers.

1) Categorical measure distinguishing between commonly completed levels of educational attainment (maedcata , paedcata)

We have code that can creates categorical measures distinguishing among parents who completed less than high school, graduated from high school, completed some college, and completed at least college, with an additional category for parents who are absent. Note, if one uses this variable as a categorical/continuous measure in an analysis, he or she is assuming that not having a mother or father figure corresponds to a lower value on this construct than having less than a 12th grade education.

2) Years of schooling completed using the midpoint of response categories

This code will recode mother's and father's education into the number of years of schooling one must complete for each educational level. This measure uses the midpoint for the response categories indicating an educational attainment level at the elementary level or below for years 2002-2005 (for example, "7th or 8th grade" is recoded as 7.5 years). For educational attainment levels greater than the elementary school level the average years of schooling for the category was used (for example, a Bachelor's Degree is coded as 16 years of schooling).

The two variables to be created are paed and maed:
paed 'Fath educ, grade equiv based on midpoint of category'
maed 'Moth educ, grade equiv based on midpoint of category'
Note that for this measure students that state that they have no father/mother figure are considered to be
missing. Code for creating paed and maed reproduced at the end of this document.

3) Years of schooling completed using the high point of the response categories

This code will recode mother's and father's education into the number of years of schooling one must
complete for each educational level. This measure uses the highpoint for the response categories
indicating an educational attainment level at the elementary level or below for years 2002-2005 (for
example, "7th or 8th grade" is recoded as 8 years). For educational attainment levels greater than the
elementary school level the average years of schooling for the category was used (for example, a
Bachelor's Degree is coded as 16 years of schooling).

The two variables to be created are paedh and maedh.
paedh 'Fath educ, grade equiv based on highpoint of category'
maedh 'Moth educ, grade equiv based on highpoint of category'
Note that for this measure students that state that they have no father/mother figure are considered to be
missing. Code for creating paedh and maedh below.

```
*** parental education .

RECODE s122_
(1 THRU 8=1) (9=2) (10 THRU 12=3) (13 THRU 16=4) (17=0) into paedcata .
RECODE s122
(1 THRU 12=1) (13=2) (14 THRU 15=3) (16 THRU 18=4) (19=0) into paedcata .
VARIABLE LABELS paedcata 'Fathers Education recoded categorically (s122)'.
RECODE s130_
(1 THRU 8=1) (9=2) (10 THRU 12=3) (13 THRU 16=4) (17=0) into maedcata .
RECODE s130
(1 THRU 12=1) (13=2) (14 THRU 15=3) (16 THRU 18=4) (19=0) into maedcata .
VARIABLE LABELS maedcata 'Mothers Education recoded categorically (s130)''.
VALUE LABELS paedcata maedcata
0 'No mother/father figure'
1 'Less then 12th Gr'
2 'HS Graduate'
3 'Some College'
4 'College Grad +'.

if missing(paedcata ) and s122  = -99797  paedcata = -99797 .
if missing(paedcata ) and s122_ = -99797  paedcata = -99797 .
if missing(maedcata ) and s130  = -99797  maedcata = -99797 .
if missing(maedcata ) and s130_ = -99797  maedcata = -99797 .

if missing(paedcata ) paedcata = -999 .
if missing(maedcata ) maedcata = -999 .

** for 2000 .
do repeat
a = s122 s130 /
b = paed maed /
c = paedh maedh .
```

```
compute a = rnd(a) .
if a = 1 b = 0 .
if a = 2 b = 1 .
if a = 3 b = 2 .
if a = 4 b = 3 .
if a = 5 b = 4 .
if a = 6 b = 5 .
if a = 7 b = 6 .
if a = 8 b = 7 .
if a = 9 b = 8 .
if a = 10 b = 9 .
if a = 11 b = 10 .
if a = 12 b = 11 .
if a = 13 b = 12 .
if a = 14 b = 13 .
if a = 15 b = 14 .
if a = 16 b = 16 .
if a = 17 b = 18 .
if a = 18 b = 20 .
compute c = b .
end repeat .
** for 2002, 2003 .
do repeat
a = s122_ s130_ /
b = paed maed /
c = paedh maedh .
compute a = rnd(a) .
if a = 1 b = 0 .
if a = 2 b = 2.5 .
if a = 3 b = 5.5 .
if a = 4 b = 7.5 .
if a = 5 b = 9 .
if a = 6 b = 10 .
if a = 7 b = 11 .
if a = 8 b = 11 .
if a = 9 b = 12 .
if a = 10 b = 13 .
if a = 11 b = 14 .
if a = 12 b = 14 .
if a = 13 b = 16 .
if a = 14 b = 18 .
if a = 15 b = 20 .
if a = 16 b = 20 .
if a = 1 c = 0 .
if a = 2 c = 4 .
if a = 3 c = 6 .
if a = 4 c = 8 .
if a = 5 c = 9 .
if a = 6 c = 10 .
if a = 7 c = 11 .
if a = 8 c = 12 .
if a = 9 c = 12 .
if a = 10 c = 13 .
if a = 11 c = 14 .
if a = 12 c = 14 .
if a = 13 c = 16 .
if a = 14 c = 18 .
```

```
if a = 15 c = 20 .
if a = 16 c = 20 .
end repeat .

** First set missing to "survey incomplete" if it is .
if missing(paed ) and s122  = -99797  paed = -99797 .
if missing(paed ) and s122_ = -99797  paed = -99797 .
if missing(paed ) and s130  = -99797  maed = -99797 .
if missing(maed ) and s130_ = -99797  maed = -99797 .

if missing(paedh ) and s122  = -99797  paedh = -99797 .
if missing(paedh ) and s122_ = -99797  paedh = -99797 .
if missing(maedh ) and s130  = -99797  maedh = -99797 .
if missing(maedh ) and s130_ = -99797  maedh = -99797 .

** Then set it to missing for some other reason - skipped, refused/na.

if missing(paedh ) paedh = -999 .
if missing(paed  ) paed  = -999 .

if missing(maedh ) maedh = -999 .
if missing(maed  ) maed  = -999 .

add value labels paed paedh maed  maedh paedcata maedcata
 -99797 'No more valid answers (possibly incomplete student survey)'
 -999   'Missing for some other reason (DK/Refused/Skipped Question)'  .

variable labels
paed 'Fath educ, grade equiv based on midpoint of category (s122)'
maed 'Moth educ, grade equiv based on midpoint of category (s130)'
paedh 'Fath educ, grade equiv based on highpoint of category (s122)'
maedh 'Moth educ, grade equiv based on highpoint of category (s130)' .

* paed paedh maed  maedh paedcata maedcata .
```

For the SES index feel free to use either the parental education indicators that use the maximum value or those that use the midpoint—they should work equally well.

SES FACTOR:
Socioeconomic status factor scores can be created using values on homeownership, mother's and father's education level, and the average level of education within each parent's occupation. The code below can be used to create SES factor scores.

```
***ses factor***.
** you might include are   ownhm  paedh maedh SFocced SMocced
missing values SFocced SMocced  (-99797 thru -999) .
FACTOR
/VARIABLES VARIABLES YOU WANT TO INCLUDE
/MISSING LISTWISE /ANALYSIS  VARIABLES YOU WANT TO INCLUDE
/PRINT UNIVARIATE INITIAL CORRELATION SIG KMO EXTRACTION ROTATION
/PLOT EIGEN
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PAF
/CRITERIA ITERATE(25)
```

```
/ROTATION VARIMAX
/SAVE REG(ALL)
/METHOD=CORRELATION.
compute ses = FAC1_1.
variable label ses 'factor of mom & dad edu, occ and own home' .
execute.
del var FAC1_1.
execute.
```