Appendix E

NELS:88 Related Data Files and Data Files Available from the National Center for Education Statistics

Studies and Files Related to NELS:88

In addition to the core sample and survey described in the main text, several other supplemental components were undertaken and data files generated under the auspices of the NELS:88 base year study. These include: several state augmentations; a supplement of hearing-impaired students, funded by Gallaudet University; a supplement of Christian schools that are members of the Christian Schools International organization, funded by the Barnabas Foundation; the NELS:88 Enhancement Survey of Middle Grades Practices, funded by the Office of Research in the Office of Educational Research and Improvement (OERI), through the Johns Hopkins University Center for Research on Elementary and Middle Schools (CREMS); the collection of transcripts for the base year teacher sample, sponsored by the National Science Foundation; and the production of a modularized version of the NELS:88 data in IBM-compatible format on floppy diskettes, sponsored by a grant from the National Science Foundation and the U.S. Department of Education. These auxiliary data files greatly expand and enrich the analytic uses of the public use data sets.

The NCES-sponsored core sample of 1,052 participating schools and 24,599 participating students was increased to 1,242 participating schools and 28,397 participating students, respectively, as a result of the state augmentations and Christian schools supplements.

Data for the state augmentations and all supplements discussed below do not appear on the NCES public release tapes for NELS:88.

Christian Schools Supplement

A sample of Christian schools that are members of the Christian Schools International (CSI) organization was drawn to supplement NELS:88. The sample was selected from CSI schools with probability proportional to eighth grade size. Two disproportionately large school units were double-sampled. Of the initially contacted 58 schools, 41 schools agreed to participate. (Due to the double-sampling of the two schools, the number of sampling units was 43.) Students, parents, teachers, and school administrators were surveyed. Students completed both the cognitive test battery and the questionnaire during the Survey Days held in their schools.

State Augmentations and Supplements

In an effort to enhance the statistical precision of their state samples, four states sponsored sample augmentations by adding schools and students in their states. Three of these states also sponsored instrument supplements in the form of additional questions pertaining to policy issues of interest to their states.

Survey of NELS:88 Base Year Dropouts

Seven months after completion of in-school data collection (in January 1989), the small number of dropouts from the base year core sample were surveyed. These were students who were eligible to participate at the time that the school roster was annotated to indicate eligibility by the school coordinator. They were drawn into the sample but then dropped out between the time of sampling and their school's Survey Day. Students who drop out of school subsequent to their base year Survey Day will be captured in the NELS:88 First Follow-up. A student was designated a "dropout" when several conditions were met: the student had been absent from the school for at least twenty consecutive days, the absence was not excused, and it was the opinion of the school coordinator that the

child would not return to school. According to this definition, chronic truants who had not taken legal action to leave school (or could not take such action owing to their age) could also be designated dropouts.

In identifying the dropouts, significant definitional problems were encountered as plans for the dropout survey progressed. On Survey Day, school coordinators identified 96 absent sample members as dropouts. However, the following autumn, it was learned that most of these students were not dropouts at all, but had transferred to other schools. Thus, during the five to seven month period following the Survey Day, when NORC staff were engaged in locating and interviewing the dropouts in the sample, it was frequently the case that students who had been originally classified as 1987-1988 school year dropouts had to be reclassified based on new information that became available. For the purposes of this survey, we attempted to collect data from all students who were dropouts or truants as of their base year Survey Day.

The sample of eligible base year dropouts, whose status was verified, contained 29 dropouts and one parent of each child. The locating task was made more difficult by the fact that, unlike those who had completed the questionnaires on Survey Day, these children had not provided any locating information. The locating information was first sought at the child's former school. If the school was not able to provide a valid current address, calls were made to directory assistance and to selected former classmates of the child. Field interviewers were able to locate 26 of the 29 students. Of the 26 locatable children, 25 participated; of the 26 locatable parents, all 26 participated. The response rate was 86 percent for the dropouts and 90 percent for their parents. Although the sample is small, it is a national probability sample of eighth grade dropouts. In the NELS:88 First Follow-up, these dropouts will be surveyed again in spring 1990.

The instruments for the dropouts differed only slightly from those used for the core sample of students. Both the base year student and base year parent questionnaires were modified in order to reflect the later administration date and changed school status of the children. Certain questions were reworded to reflect the appropriate point of reference. For example, "since the beginning of this school year" was changed to "when you were in eighth grade." Questions about school situation were deleted as no longer directly relevant to the situation of the dropout when they referred to such things as high school attendance plans and courses in which the student was currently enrolled. Student cognitive tests were not administered. Nor was teacher information collected for the dropouts.

The data collection procedures also differed from those used in the main study. Both student/dropout and parent questionnaires were completed by telephone interviews or, for the significant number of respondents without telephones, in personal interviews by NORC field staff. Locating and data collection was conducted between November, 1988 and January, 1989.

CREMS NELS:88 Enhancement Survey of Middle Grades Practices

The Survey of Middle Grades Practices enhances the NELS:88 base year school questionnaire by collecting new information to monitor middle grades reform in the schools attended by NELS:88 eighth graders. The questionnaire for this supplemental survey was designed by the Center for Research on Elementary and Middle Schools (CREMS) of the Johns Hopkins University and the data collection was conducted by NORC. The school principals who provided base year information in the NELS:88 school questionnaire were asked to participate in this enhancement survey between late October 1988 and February 1989. The enhancement survey augments the information in the base

year school questionnaire with details on school and classroom characteristics and practices, including school organization, guidance and advisory practices, rewards for and evaluations of student performance, curriculum and instructional practices, transition to high school, middle grade programs, parent involvement, and team teaching.

Included in the enhancement survey is an alternative version of an item on classroom organization. This item from the CREMS data has been appended to the base year school file. It should be noted that the original question on the organization of classroom instruction (see school codebook, BYSC18) was asked during the 1987-1988 school year, while the correction item was asked during and references the 1988-1989 school year.

The completion rate for the enhancement survey was 98.63 percent.

Collection of NELS:88 Teacher Transcripts

In order to assess teacher qualifications in science and mathematics, NELS:88 participating teachers were asked for permission to obtain copies of their college transcript records. The NSF will use the transcripts to conduct research on college coursetaking patterns of teachers in order to assess and improve teacher education and training programs.

Under a grant from the NSF, Westat began collecting the college transcripts in the fall of 1988. Based on the NELS:88 design, a total of 1,881 mathematics and science teachers (or the total number of those who gave permission to obtain their college transcripts) are participating in the Transcript Study, requiring transcript collection and follow-up efforts at registrars' offices at approximately 1,200 postsecondary institutions. Two data files will be developed to facilitate the analysis of the relationship between transcript-based measures of teacher qualifications and teacher characteristics and practices. One file will link the teacher transcript measures with applicable teacher and school survey data sets from NELS:88. The second file will link the teacher transcript measures to NELS:88 student questionnaire and cognitive test data.

Modularized Version of NELS:88 Data for Floppy Diskettes

An education longitudinal analysis group at the University of Chicago, sponsored by the National Science Foundation and the U.S. Department of Education, will produce a modularized version of the NELS:88 base year data for floppy diskettes. The modularized version of the data will be appropriate for modern IBM-compatible computing environments and it will make the data easily and more economically accessible for research and policy-related use by a wider audience. The modularized NELS:88 data will be made available by NCES.

Past Studies and Data Files Related to NELS:88 Available from NCES

Data from the earlier NCES longitudinal studies--NLS-72 and HS&B--may also be of some interest to users of the NELS:88 data. These data sets will be of special interest in later waves of NELS:88, when cross-cohort comparisons will be possible (for example, comparisons of the NELS:88 1990 sophomores and the HS&B 1980 sophomores; comparison of the 1992 NELS:88 seniors and the HS&B sophomore and senior cohorts in 1982 and 1980, and NLS-72 seniors in 1972).

In addition to the core surveys for HS&B and NLS-72, briefly described earlier, records studies have been undertaken, including the collection of the high school transcripts of the sophomore cohort and the collection of postsecondary education transcripts and financial aid data for the seniors.

Data files for these studies and other HS&B data, such as parent surveys, school surveys, teacher comments, etc., are described below. Users manuals or other forms of documentation are available from NCES for all the data files. These auxiliary data files greatly expand the analytic capabilities of the core data sets, and researchers are encouraged to become familiar with them.

HS&B Base Year Files

The **Language File** contains information on each student who during the base year reported some non-English language experience either during childhood or at the time of the survey. This file contains 11,303 records (sophomores and seniors combined), with 42 variables for each student.

The **Parent File** contains questionnaire responses from the parents of about 3,600 sophomores and 3,600 seniors who are on the Student File. Each record on the Parent File contains a total of 307 variables. Data on this file include parents' aspirations and plans for their children's postsecondary education.

The **Twin and Sibling File** contains base year responses from sampled twins and triplets; data on non-sampled twins and triplets of sample members; and data from siblings in the sample. This file (2,718 records) includes all of the variables that are on the HS&B student file, plus two additional variables (family ID and SETTYPE--type of twin or sibling).

The **Sophomore Teacher File** contains responses from 14,103 teachers on 18,291 students from 616 schools. The **Senior Teacher File** contains responses from 13,683 teachers on 17,056 students from 611 schools. At each grade level, teachers had the opportunity to answer questions about HS&B-sampled students who had been in their classes. The typical student in the sample was rated by an average of four different teachers. Preliminary analyses by NCES indicate that the files contain approximately 76,000 teacher observations of sophomores and about 67,000 teacher observations of seniors.

The **Friends File** contains identification numbers of students in the HS&B sample who were named as friends of other HS&B-sampled students. Each record contains the IDs of sampled students and IDs of up to three friends. Linkages among friends can be used to investigate the sociometry of friendship structures, including reciprocity of choices among students in the sample, and to trace friendship networks.

Merged HS&B Base Year, First, Second and Third Follow-Up Files

The **First Follow-Up Sophomore File** contains responses from 29,737 students and includes both base year and first follow-up data. This file includes information on school, family, work experiences, educational and occupational aspirations, personal values, and test scores of sample participants. Students are also classified in terms of high school status as of 1982 (that is, dropout, same school, transfer, or early graduate).

The **First Follow-Up Senior File** contains responses from 11,995 individuals and includes both base year and first follow-up data. This file includes information from respondents concerning their high school and postsecondary experiences and their work experiences.

The **Second Follow-Up Sophomore File** has all base year, first follow-up, and second follow-up data for 14,825 members of the sophomore cohort. Data cover work experience, postsecond-

ary schooling, earnings, periods of unemployment, and so forth, for the sophomore cohort, who by this time had been out of high school for two years.

The **Second Follow-Up Senior File** encompasses all base year, first follow-up, and second follow-up data for the 11,995 individuals who constitute this follow-up sample. Data cover work experience, postsecondary schooling, earnings, periods of unemployment, and so forth, for the senior cohort, who by this time had been out of high school for four years.

The **Third Follow-Up Sophomore File** includes all base year, first follow-up, second follow-up, and third follow-up data for the 14,825 members of the sophomore cohort. Data cover marriage and family formation, work experience, postsecondary schooling and interest in graduate degree programs, earnings, periods of unemployment, and alcohol consumption for this cohort, who by 1986 had been out of high school for four years.

The **Third Follow-Up Senior File** includes all base year, first follow-up, second follow-up, and third follow-up data for the 11,995 individuals who constitute this follow-up sample. Data cover marriage and family formation, work experience, postsecondary schooling and interest in graduate degree programs, earnings, periods of unemployment, and alcohol consumption for the senior cohort, who by 1986 had been out of high school for six years.

Other HS&B Files

The **High School Transcript File** describes the coursetaking behavior of 15,941 sophomores of 1980 throughout their four years of high school. Data include a six-digit course number for each course taken, along with course credit, course grade, and year taken. Other items of information, such as grade point average, days absent, and standardized test scores, are also contained on the file.

The Offerings and Enrollments File contains school information, course offerings, and enrollment data for 957 schools. Each course offered by a school is identified by a six-digit course number. Other information, such as credit offered by the school, is also contained on each record.

The **Updated School File** contains base year data (966 completed questionnaires) and first follow-up data (956 completed questionnaires) from the 1,015 participating schools in the HS&B sample. First follow-up data were requested only from those schools that were still in existence in the spring of 1982 and had members of the 1980 sophomore cohort currently enrolled. Each high school is represented by a single record that includes 230 data elements from the base year school questionnaire, if available, along with other information from the sampling files (e.g., stratum codes, case weights).

The **Postsecondary Education Transcript File** for the HS&B seniors contains transcript data on dates of attendance, fields of study, degrees earned, and the titles, grades, and credits of every course attempted at each school attended, coded into hierarchical files with the student as the highest level of aggregation. Although no survey forms were used, detailed procedures were developed for extracting and processing information from the postsecondary school transcripts that were collected for all members of the 1980 senior cohort who reported attending any form of postsecondary schooling in the first or second follow-up surveys. (Over 7,000 individuals reported over 11,000 instances of school attendance.)

The **Senior Financial Aid File** contains financial aid records from postsecondary institutions respondents reported attending and federal records of the Guaranteed Student Loan Program and of the Pell Grant program.

The **HS&B HEGIS and PSVD File** contains the postsecondary school codes for schools HS&B respondents reported attending in the first and second follow-ups. In addition, the file provides data on institutional characteristics, such as type of institution, highest degree offered, enrollment, admissions requirements, tuition, and so forth. This file permits analysts to link HS&B questionnaire data with institutional data for postsecondary schools attended by respondents.

NLS-72 Files

The **NLS-72 Base Year Through Fourth Follow-Up (1979) File** contains data from the base year through fourth follow-up for over 23,000 respondents. Data include school experiences and test results during the base year and subsequent activities related to work, postsecondary schooling, military service, family formation, and goals and aspirations.

The **NLS-72 Fifth Follow-Up File** consists of the results of the fifth follow-up survey, carried out in 1986, when sample members were about thirty-two years old. Data include work experience going back to 1979, postsecondary schooling, extensive family formation history, periods of unemployment, goals and aspirations, and selected attitudes. Records in this file can be linked through student ID to those in the *NLS-72 Base Year Through Fourth Follow-Up (1979).*

The **NLS-72 Teacher Supplement File** contains the responses of the portion of the fifth follow-up NLS-72 sample who had obtained teacher certification and/or had teaching experience. Data include certification history, subjects taught, years of experience, attitudes toward teaching as a career, and subsequent work experiences of those who had left teaching. These data can be linked through the respondent ID to the *NLS-72 Fifth Follow-Up File* and to the *NLS-72 Base Year Through Fourth Follow-Up File.*

The **Postsecondary Education Transcript Study of the NLS-72 Sample** contains transcript data on dates of attendance, fields of study, degrees earned, and the titles, grades, and credits of every course attempted at each school attended, coded into hierarchical files with the student as the highest level of aggregation. Although no survey forms were used, detailed procedures were developed for extracting and processing information from the postsecondary school transcripts that were collected in 1984 for all members of the NLS-72 cohort who reported attending any form of postsecondary schooling in any of the first through fourth follow-up surveys. (Over 14,000 individuals reported over 24,000 instances of school attendance).

Appendix F

Guidelines for Using SAS with NELS:88 Student Data

The files provided on the public release tape include SAS cards and a SAS system file.

The SAS system file includes:

1) Base Year Questionnaire Data
2) Base Year Flags, Weight, and Composites

NCES and NORC strongly suggest that all SAS users be aware of the potential problem areas when using student data files via SAS.

1. SAS users should use the '(KEEP=...)' and '(DROP=...)' options in the 'SET...;' statement and/or in the 'DATA...;' statement when creating working data files so that unwanted variables are not included in the files. It is faster (but not essential) for variables in the '(KEEP=...)' statement to be listed in the same order as they occur in the main system file. Remember also that the '(KEEP=...)' option does not reorder the variables in the new data set.

2. You may have to delete at least one third of the label cards given in this file because of SAS system limitations, which are present at many computer installations.

3. The large number of **VALUE** cards in the **PROC FORMAT** section requires that a special **DD** statement be placed just after the **//EXEC SAS** card to increase the capacity of the format library during a SAS run:

```
//LIBRARY DD SPACE=(TRK,(25,25,60))
```

This may not be possible at some computer installations, so it may be necessary to delete some values.

4. When working with large files, it may be necessary to override the default work space with the following DD card:

```
//WORK DD UNIT=SYSCR,SPACE=(CYL,(40,40))
```

Place the **//WORK DD** card just after the **//EXEC SAS** card (or after the **//LIBRARY DD** card, if that is included as well).

5. The formats given in the **PROC FORMAT** step here are not permanently associated with each variable. Whenever they are needed for a procedure, it is necessary to include them in this **PROC FORMAT** step before the procedure that will use them, as shown in the following example:

```
//EXEC SAS,OPTIONS='NOGRAPHICS',REGION=1280K
//LIBRARY DD SPACE=(TRK,(25,25,60))
//WORK DD UNIT=SYSCR,SPACE=(TRK,(1000,1000))
//IN01 DD DSN=ACT.STUDENT.SASLIB,
//        UNIT=SYSDA,
//        DISP=SHR
//SYSIN DD *
```

```
OPTIONS DQUOTE;

PROC FORMAT;

VALUE FSEXV     1 = "MALE"
                2 = "FEMALE"
                6 = "MULTIPLE RESPONSE"
                7 = "REFUSAL"
                8 = "MISSING"
                9 = "LEGITIMATE SKIP";
 VALUE FBYS46V 1 = "VERY SURE WILL GRAD"
                2 = "PROBABLY GRADUATE"
                3 = "PROBABLY WON'T GRADUATE"
                4 = "VERY SURE WON'T GRADUATE"
                6 = "MULTIPLE RESPONSE"
                7 = "REFUSAL"
                8 = "MISSING"
                9 = "LEGITIMATE SKIP";

PROC FREQ DATA=IN01.STQ;
FORMAT     SEX    FSEXV.
           BYS46  FBYS46V.;

 TABLES SEX*BYS46;
 TITLE "SEX OF RESPONDENT BY CERTAINTY OF HIGH SCHOOL GRADUATION";
```

At the end of the formats given in this file, there is a frequency procedure and a means procedure (in comment form), which together contain **FORMAT...;** statements for every variable for which there is a format. These **FORMAT...;** statements will save users a lot of time because they can be used in any SAS procedure.

When users create their own formats they should keep in mind that a format for a character variable must have a format name beginning with '$', and that format names must not end in a digit.

6. For very large files, the user may encounter problems when sorting. Various options may be added to the **//EXEC SAS** card to circumvent these problems. A suggested example is given below (consult the SAS manual for descriptions of these options):

```
// EXEC SAS,OPTIONS='NODYNALLOC',REGION=1280K,SORT=30
```

7. It is suggested that the user include the **LENGTH** statement when creating new variables, in order to save space and computer memory.

8. For many tabulations, **PROC TABULATE** produces the most readable output. The SAS user may use the format statements (provided) for classification variables to produce the row values of tabulate tables.

9. Output from SAS can be downloaded to personal computers for production of final reports. NCES has available a program for taking into account the sample design when computing standard errors. The program, known as CTAB, is a Taylor series based routine that uses an ASCII file to compute standard errors for crossclassifications. The program also produces labeled tabular output suitable for use in publications. CTAB is available for use on microcomputers, and can be obtained through NCES.

10. Use the NCES- and NORC- defined composite and classification variables whenever possible to simplify programming. These classification variables were carefully constructed and, for many of them, sources of data from outside the student questionnaire were merged into the student data to construct the variables.

11. SAS and SPSS-X system files can now be converted at many computer installations. Contact your own facility to obtain the information necessary to create an SPSS-X file from SAS and vice versa.