

Slide 1 of 22

Title Slide: Getting Started with SASS Data

Slide 2 of 22

The Schools and Staffing Survey (or SASS) is a set of inter-related surveys. This module introduces users to the data collected across the 2011-12 SASS. It describes the process of obtaining previously tabulated data and generating custom tables from PowerStats and the resources that are available to learn more about the surveys, the data, and the data files.

This module describes the contents of each of the 2011-12 SASS data files, their variables, and variable naming conventions. Information presented in this module will be helpful for understanding some of the more detailed information presented in subsequent SASS modules. For this reason, users who are planning to proceed through the modules and use SASS data for analytic purposes are strongly encouraged to complete this module first.

Slide 3 of 22

Information in SASS was collected from multiple respondents at each sampled school but all respondents provided data for the same 2011-12 school year. Principals were asked to designate a school respondent who, if possible, could serve as the school coordinator. A coordinator's duties included facilitating data collection by passing out questionnaires to the appropriate staff, reminding the staff to complete them, and collecting the questionnaires to return to the U.S. Census Bureau. The school respondent/coordinator could have been the principal, assistant or vice principal, or any other administrative employee. All SASS data collection was by paper questionnaires collected by the U.S. Census Bureau, aside from the Teacher Listing form (or TLF). The Teacher Listing form was an electronic spreadsheet for collecting the entire school's teacher roster. Principals were asked to submit the Teacher Listing Form online, which can be accessed by clicking on the corresponding underlined screen text.

Slide 4 of 22

For traditional public schools, the district associated with the sample school received the District questionnaire unless the sampled school was the only school in the district. Traditional public school principals and public charter school principals received the School Principal questionnaire. Both traditional public school principal and public charter principal data are on the Public Principal data file.

Public charter schools and schools in one-school districts received the Public School questionnaire with district items. All of the public school data, however, after data processing, imputation, and weighting, were combined into the Public School data file. The district items from the public charter schools and schools in one-school districts were added to the District file.

Getting Started with SASS Data

Data from all data collection components are available in the SASS data files. More specifically, the data files include variables providing information obtained from the questionnaires. All of these data collection instruments are available online and can be accessed by clicking on the underlined screen text, 'questionnaires.' The next module in this series – titled 'Data Collected Through the SASS' – describes the range of topics covered in SASS in more detail and provides information about the frame and derived variables that are available for analysis.

Slide 5 of 22

Public-use descriptive statistics are survey results that have been generated by NCES. Public-use data files allow users to manipulate the data but have been altered to prevent inadvertent disclosure of respondent identity. No public-use data files have been generated for SASS after 1999-2000 or for the Teacher Follow-up Survey (or TFS) after 2000-2001. The widespread availability of online information about public and private schools makes it much more difficult to protect the anonymity of respondents to SASS or TFS.

The most easily accessed SASS data results – that is, public-use descriptive statistics - are only available in SASS reports, the SASS/TFS Table Library, or by generating tables within PowerStats. Each of these sources of SASS data results are accessible by clicking on the corresponding underlined screen text. Data from 1999-2000 SASS and 2000-2001 TFS are available as public-use data sets within eDAT. eDAT, or the Education Data Analysis Tool, can be accessed by clicking on the corresponding underlined screen text.

Restricted-use datasets allow researchers to analyze survey results with full access to the data supplied by respondents. As discussed in the DLDT Common Module titled, "Acquiring micro-level NCES datasets," restricted-use data are only available to researchers who apply for and are granted a restricted-use license to use them. The restricted-use files for a given round of data collection contain all of the original survey responses and variables, while PowerStats or published reports may not cover all of the variables or response categories of interest. It is important to note that the 2011-12 SASS is not available in downloadable public-use data files, only within the PowerStats data tool and restricted-use micro-level data files. In these training modules, any references to 2011-12 SASS data files are to restricted-use data files only.

Researchers who are uncertain of which data file is right for their analytic purposes should first examine the public-use descriptive statistics provided in PowerStats.

Slide 6 of 22

Reports are available for the 2011-12 SASS District, School, Principal, Teacher, and School Library Media Center data. These reports can be accessed via the NCES website by clicking on the underlined title of interest. Each of these reports features a few data elements from that particular SASS component, but can only cover a fraction of the topics collected in SASS. The first part of each report presents a page of

Getting Started with SASS Data

“selected findings” or highlights of data results, plus a number of data tables with more detailed results. Following the data tables and standard error tables, there is a technical summary of SASS data collection, data processing, response rates, and imputation and weighting methods.

Slide 7 of 22

The SASS/TFS Table Library is a set of tables with data from 2011-12 SASS as well as some previous rounds of SASS and TFS. These tables are organized by survey year and by type of respondent. The first way to search is to browse all of the tables for a selected SASS data collection year. That is the drop-down box to the left of the page – “Tables by Year.” If you click on Search, you will get a list of all the tables for that survey year and respondent.

Another way to search for data tables is to use the drop-down boxes to the right of the Table Library page. First, click on the “Search a year” box. Then click on the “Data regarding” box. This allows you to pick the type of respondent you are seeking—districts, schools, principals, teachers, or library media centers. You also have the option to restrict your search to Public or Private sector data. The last drop-down box, Topic, is for the level of data – national, state, or affiliation.

It is not necessary to use all of the drop-down boxes in order to run a search, but recall that the more specific the terms for a search, the fewer, more specific results you will get.

The same choices can be made for TFS, with one change. First you will have to change the “Search a Survey” button from SASS to TFS. Doing this will change your search options. You will be able to “Search a year”; explore data regarding Movers, Stayers, or Leavers; and search by “Sector”. Remember that the TFS data files are by category of respondent (that is Stayers, Movers, or Leavers) rather than by Public or Private sector as in SASS.

Slide 8 of 22

Users can generate customized data results or tables from the public-use SASS data by using PowerStats. Each time you visit PowerStats you will be asked to log in and agree to the terms in the NCES Data Usage Agreement.

Once logged in, PowerStats can be used to search published tables and create custom tables and analyses.

On the SASS PowerStats home screen, you will see tabs which can be clicked for more information about SASS, the SASS Questionnaires, the SASS Codebooks, and any available PowerStats updates.

Visit the Datalab to create custom tables and analyses in PowerStats by clicking on the corresponding underlined screen text.

Slide 9 of 22

After logging in to PowerStats, you will have the option to create a table or regression, as well as access informative PowerStats training modules. These modules will help you get to know the PowerStats system and the analyses that are possible using SASS data.

From this screen, you can also access tables or regressions that you have recently created, in the “Recent Work” section.

Slide 10 of 22

The basic PowerStats analysis workspace is shown here. You will notice that the SASS variables run along the left-side of the screen, and the table creation occurs in the center of the screen. The PowerStats training modules contain step-by-step information for table and regression creation in PowerStats.

For more information on PowerStats and to access the PowerStats training modules, click on the underlined screen text.

Slide 11 of 22

The decision whether to use the PowerStats public-use data tool or to seek a restricted-use license agreement with NCES depends on whether the basic data tables provided satisfy your research interest. Exploring PowerStats for the data you seek from SASS may give you that answer relatively quickly. While there are many data tables that can be generated from PowerStats, there is no ability to construct your own variables or to use continuous data in the form you prefer. With the restricted-use data files, the user can manipulate response categories and combine items into new variables.

It is also important to note that TFS (or the Teacher Follow-up Survey), PFS (or the Principal Follow-up Survey), and BTLS (or the Beginning Teacher Longitudinal Study) data are not available in PowerStats.

Slide 12 of 22

If you determine that only restricted-use SASS data will address your research question, the Restricted-Use Data Procedures Manual should be downloaded from the NCES website. The manual can be accessed by clicking on the underlined screen text, ‘Restricted-Use Data Licenses’. The manual provides step by step instructions on the process of applying online for the restricted-use license. It is not possible to download restricted-use data files from any part of the NCES website.

Slide 13 of 22

If you have access to restricted-use SASS data, you will find several important resources available within the restricted-use DVD. You should always start with the most important resource on the DVD which is the README file. This file contains information about the DVD contents and structure. It also provides you with installation information.

The data files are in ASCII format on the DVD. ASCII allows “raw” data to be read into any type of statistical software program without imposing any formatting. This allows importation of the raw data into any statistical software program of your choice.

Additionally, syntax code to support importation into SAS, Stata, and SPSS is also provided. Excel does not allow the user to weight the data and is not the appropriate choice for research with SASS data.

The DVD also contains PDFs of all of the SASS questionnaires. The questionnaires provide the source code variable name for each questionnaire item, along with the exact item wording and the response categories.

Much more complete detail on how the survey instruments were designed, how the survey was conducted, how the data were processed, and how the data files can be merged, is contained in the Data Documentation. In earlier rounds of SASS, the same report is titled Data File User’s Manual.

The Restricted-Use DVD is reissued when TFS and PFS data are released. The reissued DVD contains all the SASS resources listed here, as well as the TFS and PFS data files and their supporting documentation.

Slide 14 of 22

The questionnaire variables in the SASS data files are named using a standard set of conventions. The naming conventions for other variables such as imputation flags, weights, and frame or derived variables will be discussed later.

For questionnaire variables only, the first character of the questionnaire variable name is a prefix letter for the data file type: “A” for Administrator (or Principal); “D” for District; “S” for School; “T” for Teacher” and “M” for Library Media Center.

The rest of the variable name indicates the numeric source code that is found next to the questionnaire item on the survey instrument. Source codes are four digits and are not necessarily in numeric order on the questionnaire but are put in numeric order on the data files. Each source code is assigned to a particular questionnaire item response. There may be one or more source codes for any given questionnaire item. This is why downloading the PDF of the questionnaire is so important – you need to see the exact survey item wording, what the item response categories are and obtain source codes for the variables of interest to you.

Getting Started with SASS Data

An example is item 6 on the Teacher Questionnaire (In what school year did you begin teaching at THIS school?). Note that the small number to the left of the item wording is source code 0030. Adding the data file prefix, this is variable T0030 in the Public Teacher data file.

Slide 15 of 22

Variable labels offer more description than the variable names. Key words from the item are incorporated in an abbreviated form into the variable label.

For example, let's consider the variable 'T0030' that is associated with item 6 from the teacher questionnaire. The 'T' in the first character position indicates that this variable is from a teacher questionnaire, and the number is the source code for that item, item 6 on the public teacher or 4A questionnaire, shown here. The variable label associated with 'T0030' is 'Tchfrstyr' which is a shortened version of "Teacher first year."

Slide 16 of 22

This slide provides additional examples of the variable naming conventions used in SASS. In the case of Total Income, or TOTINCOME (spelled T-O-T-Income), the lack of a source code number in the variable name indicates that it is a non-questionnaire variable. Derived variables are variables in which the survey items are recoded or reconfigured (such as adding several income items together).

In the case of S0250, the S indicates that the variable is from the School questionnaire. The 0250 indicates the source code from the questionnaire. The way to identify the variable from the source code is to look it up on the survey questionnaire, in this case the School questionnaire.

It is important to note that variable names are listed in source code order. The other variables shown here come from other components including the Principal questionnaire, as indicated by the first character A (for Administrator); the district questionnaire, indicated by the letter D in the first character position; and the Teacher and Library Media Center questionnaires, as indicated by the letters T and M.

Lastly, the variable TFNLWGT includes an abbreviated "Final Weight" in the variable name including a prefix of T that it is a weight variable for the Teacher file. Replicate weights are all provided with the data file prefix followed by REPWT (spelled R-E-P-W-T) and the numbers 1 through 88 because there are 88 replicate weights on each data file. The example shown here includes a prefix of "A" for the principal file, followed by REPWT1 to denote a principal file replicate weight.

Slide 17 of 22

Frame variables come from the sampling frame, either the Common Core of Data (CCD) for public schools and districts or the Private School Universe Survey (PSS) for private schools. ENRKG12 is the total enrollment from the CCD, where ENR stands for enrollment and KG12 is the grade range Kindergarten through grade 12. This variable has been renamed from the CCD variable name, but the data remain the same as on the CCD school file. LOWGRADE is the CCD variable for the lowest grade offered by the school. Frame variables are usually alphabetic, however some do have numbers (but not source codes).

Derived variables are recoded variables or variables based on more than one questionnaire item. These variable names are entirely alphabetic, with no source code numbers. They were created for users using information from two or more variables, two or more sources, or both. Derived variables have been created for reporting purposes, and to help analysts save time and effort by not having to recreate commonly-used variables themselves. Detailed descriptions of all the derived variables can be found in the data file documentation.

For example, URBANIC is a recode of the 12-level community size variables from CCD (LOCALE). This variable collapses 12 levels which include (Large city, Midsize city, Small city, Large suburb, Midsize suburb, Small suburb, Fringe Town, Distant Town, Remote Town, Fringe Rural, Distant Rural, and Remote Rural) into 4 levels (City, Suburb, Town, and Rural), which can be more useful for reporting purposes.

Slide 18 of 22

SASS data are fully imputed – this means all data records contain a response for each questionnaire item.

As a result, each item from the SASS questionnaires will be represented by three variables in the SASS data files: the fully imputed questionnaire variable, the edit flag variable, and the imputation flag variable. Flags are used to indicate, or flag, which edits or imputations were conducted for each variable individually. The edit flag and the imputation flag are set to zero when a valid response was provided to the actual questionnaire item, meaning no edits or imputations were necessary. Edit flag variables all start with “EF underscore” followed by the source code of the questionnaire variable, while imputation flag variables all start with “F underscore” followed by the source code. In other words, “EF_T0250” is the variable name for the edit flag corresponding to the teacher questionnaire item 0250. Following the same format, “F_T0250” is the imputation flag corresponding to the teacher questionnaire item 0250.

If the response provided on the questionnaire fell outside the acceptable valid value range, the edit flag was set to one of the following values:

- 1 – If the Item was edited during only the consistency edits.
- 2 – If the original value was ratio adjusted during the logic edit.

Getting Started with SASS Data

3 – If the data were added using data from other variables in the same questionnaire.

4 – If the data were added using data from another associated questionnaire (principal record, district record, school record or Teacher Listing Form).

5 – If the data were added using data from the sample file (CCD for public non-teachers, PSS for private non-teachers, or Teacher Listing Form).

12 – If the item was edited during the consistency edits, and item was ratio adjusted during the logic edit.

13 – If the item was edited during the consistency edits and data were added using data from other variables in the same questionnaire.

14 – If the item was edited during the consistency edits and data were added using data from another associated questionnaire (principal record, district record, school record or Teacher Listing Form).

And 15 – if the Item was edited during the consistency edits and data were added using data from the sample file (CCD for public non-teachers, PSS for private non-teachers, or Teacher Listing Form).

When an item was skipped, or no response was provided, the imputation flag value was set to a value between 7 and 9, depending on the method used to impute the value:

7 – If the Item was imputed using hot deck imputation. Hot deck imputation is selecting a value from a matched respondent. Hot deck methods impute missing values within a data matrix by using available values from the same matrix. The object, from which these available values are taken for imputation within another, is called the donor.

8 – If the Item was imputed using the mean or mode of similar respondents.

9 – If the Item was imputed based upon post-imputation review.

The Final Weight contains all of the sampling and nonresponse adjustments needed to weight the sample estimate up to the target population. DFNLWGT is the District file Final Weight. The replicate weights are a set of weights that are needed to calculate standard errors, which will be described in more detail in the module titled “SASS Sample Design, Weights, Variance, and Missing Data.” DREPWT1, spelled D-R-E-P-W-T-1, is the first replicate weight in this series. The prefix used with each data file, such as D for District, is also used for DREPWT variables 1 through 88.

Slide 19 of 22

Let's take a closer look at imputation flag and weighting variables. Imputation flag variables all start with the letter 'F' and the underscore character, then have the alphabetic prefix letter for the file name and the number of the source code for the particular variable. Therefore, the variable "F_T0030" is an imputation flag variable for the variable "T0030". "T0030" is item 6 on the 2011-12 SASS Teacher questionnaire.

Each data file has one and only one final weight variable. The variable name has the file alpha prefix letter (D, A, S, T, M) followed by FNLWGT. Instructions and examples for assigning weights when using SASS will be provided in detail in the module titled, 'Considerations for Analysis of SASS Data.'

Slide 20 of 22

In addition to the resources already discussed, there are two additional resources that may be of interest. The [Online Bibliography Search Tool](#) allows you to see what other researchers have published using SASS data. To use the Search Tool, at a minimum, you must enter in the Data Source (and SASS is one of the NCES datasets listed). You do not have to know the exact research article title or the author's name, as any field left blank widens the search results.

The SASS [Methods and Procedures](#) webpages gives a summary of the sampling methods and data collection procedures used in each round of SASS. The Data Documentation or Data File User's Manual for the particular survey year has more in-depth coverage of these topics.

Slide 21 of 22

This module has introduced users to the data collected across the 2011-12 SASS. It described the process of obtaining previously tabulated data and generating custom tables from PowerStats and the resources that are available to learn more about the surveys, the data, and the data files.

Slide 22 of 22

Additionally, important resources that have been provided throughout the module are summarized here along with the module's objectives for your reference.

You may now proceed to the next module in the series or exit the module.