

Slide 1 of 16

Title Slide: Getting Started with the ECLS-K Data

Slide 2 of 16

This module provides an overview of the data collected in the ECLS-K that are available for researchers to analyze. The overview is intended to help users begin to develop a conceptual understanding of the micro-level data files containing the data produced from the study. This module also describes the resources that are available to learn more about the study, the data, and the data files. Information presented in this module will be helpful in understanding some of the more detailed information presented in subsequent ECLS-K modules. For this reason, users who are planning to proceed through the subsequent ECLS-K modules and use ECLS-K data for analytic purposes are strongly encouraged to complete this module first.

Slide 3 of 16

Information pertaining to the ECLS-K children was collected at and about multiple time points in the children's lives -- from the fall of kindergarten through the spring of eighth grade. Information was collected from multiple sources using a variety of data collection methods.

Data from all of these time points and respondents are available in the ECLS-K data file. More specifically, the data file includes variables providing information obtained from the vast majority of questions asked in the parent interviews, the self-administered hard-copy teacher questionnaires, and the self-administered hard-copy school administrator questionnaires. All of these data collection instruments are available on-line and can be accessed by clicking the underlined screen text on the slide.

Information about the schools children attended was obtained from NCES's school universe files, the Common Core of Data and the Private School Survey, and is included in the data files as well.

Due to copyright restrictions, the direct cognitive and fine and gross motor child assessments are not publicly available for viewing and use. Also, item-level data for the direct cognitive assessments are not available on the main public- and restricted-use files, so assessment scores representing children's performance on the assessments were calculated and provided on the data files for researchers. Additionally, some of the indirect assessments, such as those measuring children's socioemotional development, are copyright protected and, therefore, are not publicly available. Summary scores, rather than item-level data, from these copyrighted indirect assessments are included in the ECLS-K data file. The module titled 'Data Collected through the ECLS-K' describes the assessments that were fielded in the ECLS-K in more detail and provides information about the assessment scores that are available for analysis.

Slide 4 of 16

In addition to variables containing information that was obtained directly from respondents in the study, there are several other categories of variables available on the data file. These categories are administrative variables, composite variables, and weighting variables.

There is a wealth of information about study respondents and different aspects of data collection that was compiled and maintained in administrative records throughout the course of the study. This information typically was not collected from respondents through the data collection instruments. Some administrative information that may be useful for analyses is provided in the data files. Examples of such administrative variables include the ZIP code of the child's school, the child's age at the time of assessment, and teacher and school identification numbers.

Composite variables are variables that were created using information from two or more variables, two or more sources, or both. Sometimes the composite variables were derived using information that is not available to researchers in the data files, for example sensitive administrative data that could not be provided for data confidentiality reasons. Detailed descriptions of all the composite variables can be found in the data file user's manuals that are accessible by clicking the underlined screen text on this slide and expanding the heading "Technical/Methodological." The user's manuals also are provided on the discs with the ECLS-K data.

Lastly, the data files contain many variables that are used in the process of weighting and standard error adjustment. Weighting and standard error calculation procedures are discussed at a general level in the common modules and within the ECLS-K module titled, 'Sample Design, Weights, Variance, and Missing Data.'

Slide 5 of 16

The ECLS-K data are available in both public- and restricted-use formats. Public-use data can be accessed by anyone. As discussed in the 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.

Most of the available public-use ECLS-K data are included in one data file, the Kindergarten Through Eighth Grade Full Sample Public-Use Data File. Note that there is an error in one set of child assessment scores on this file described within the errata that can be accessed by clicking on the underlined screen text on this slide.

Restricted-use data are available in round-specific cross-sectional files and supplementary data files. All of the restricted-use files produced for the ECLS-K are listed later in this module.

Getting Started with the ECLS-K Data

The restricted-use file for a given round of data collection contains more data and a wider range of data values than are included for that round in the public-use files. For most users, the public-use files provide all the data they will need for most analyses, though some users may find that only the restricted files have the specific data they need. It is recommended that researchers who are uncertain of which data file to use first examine the public-use data file to ascertain whether their specific analytical objectives can be met using that data file. It is easy to identify for which variables data are available in the restricted files only; all values for all cases for these variables are set to -2 in the public-use file. Also, each data file user's manual has a table in chapter 7 that lists all the variables that have been altered in some way for the public-use file.

Both public- and restricted-use data come on a CD or DVD containing the data and data documentation, as well as Electronic Codebook software.

Slide 6 of 16

The Electronic Codebook software, also called an ECB, helps users create customized data files with only the variables needed for analysis, which may be easier to work with and analyze than the full data files. Users can view descriptions of all the variables in the data file, along with the variable categories and frequency distributions, in the electronic codebook. Functions within the ECB allow users to identify which variables they want to include in their customized data files and then create a syntax file to be used within a statistical software package to generate their data file. Syntax files can be generated for SAS, SPSS, or Stata. Note that the ECB does not produce a data file. More information on the ECB, including a guide to using it, can be found in the user's manuals.

One word of caution: the ASCII data file, which ends with a .dat extension, contains the raw data for each child in the study. Do not alter this file in any way, or you will be altering the actual data.

Public-use data can also be downloaded as full data files from the data products page on the ECLS-K website. Additionally, an online data tool called the EDAT, which functions like the ECB, can be used to help researchers create customized data files containing public-use data.

Slide 7 of 16

Within the general categories of public- and restricted-use data, there are many files containing ECLS-K data that are available for analysis. Over the lifetime of the study, different cross-sectional and longitudinal public-use data files were released as new data became available. At the conclusion of the study, one final public-use data file, the kindergarten through eighth-grade full sample data file, was developed. All of the data that were ever released at the child-level in previous public-use files for the study are included in this data file. It contains data from every round of data collection, so it is appropriate for analyses of data from any time point. Previously released public-use data files do not contain revised data or the most recent and, therefore most precise,

Getting Started with the ECLS-K Data

assessment data, so it is recommended that researchers no longer use those earlier public-use files. Separate files with base-year data for teachers and schools are available for researchers who are interested in doing teacher- or school-level analyses. As previously mentioned, all three of these files are available on DVD or in the EDAT and can also be downloaded as full data files from the ECLS-K website.

Supplementary public-use data files that contain item-level data for the Self-Description Questionnaire, or SDQ, and Approaches to Learning items are also available as full data files on the ECLS-K website. More detail on the SDQ and Approaches to Learning items is provided in the module titled, 'Data Collected through the ECLS-K.'

Slide 8 of 16

While the kindergarten through eighth-grade public-use file is longitudinal, the restricted-use files are cross-sectional; each file only includes data for one grade of the study. There is one file that contains data collected in the fall and spring of kindergarten, one file that contains data collected in the fall and spring of first grade, one file that contains data collected in the spring of third grade, and so on.

In addition to the grade-specific, cross-sectional, restricted-use files, there are other restricted-use files with data related to specific topics. The data from the Head Start verification study and the kindergarten teacher and school administrator salary and benefits data that were collected in the base year are contained in their own restricted-use files. For the base year only, the special education teacher and student record abstract data were included in their own restricted-use files. After the base year, these data were included in the grade-specific, cross-sectional, restricted-use files.

The supplementary geocode data file, which is discussed in more detail in the module titled, 'Considerations for Analysis of ECLS-K Data,' contains geographic identifiers and associated Census-derived data for the kindergarten through third-grade data collections. Finally, there is a separate restricted-use file that includes parent and teacher item-level Social Rating Scale, or SRS, data for the kindergarten through fifth-grade rounds. The SRS is described in more detail in the module titled, 'Data collected through the ECLS-K.'

Slide 9 of 16

The variables in the ECLS-K data file are named using a standard set of conventions.

The first character of a variable's name generally indicates the source of the data. For example, variables that pertain to questions in the parent interview start with the letter P, while those that pertain to questions in the school administrator questionnaire start with the letter S.

The second character of the variable name generally indicates the round of data collection, with the numbers 1 through 7 corresponding to a particular round, as shown here. There is one exception to this convention among variables for which the second character is a number: the second character of composite variables beginning with W

Getting Started with the ECLS-K Data

indicates grade, not round. Composites beginning with a W are derived from information collected from parents and should be conceptualized as describing children's parents or household in a particular grade, rather than a particular round. For example, the second character is a K for variables that describe the children's parents or household during the child's kindergarten year, the second character is a 1 for variables that describe the children's parents or household during the first-grade rounds of data collection, and so on.

Slide 10 of 16

Look at the first example of a variable name shown on the screen, P2MEETTM. The first two characters of the variable indicate that the variable pertains to a question asked in the round 2, or spring kindergarten, parent interview.

The first two characters of the variable S6GFTPRG indicate that the variable pertains to the round 6, or spring fifth-grade, school administrator questionnaire.

At the bottom of the screen, there is a link to a handout that provides a comprehensive list of variable naming conventions for all rounds of collection. You may find it helpful to print this handout and use it as a reference as you proceed through the other modules.

Slide 11 of 16

In addition to using variable naming conventions, the ECLS-K followed conventions for variable labels.

As with the variable naming conventions, the first character of a variable label indicates the source of the data. The second character of the variable label indicates the round of data collection or, in the case of composite variables beginning with W, the grade.

The remainder of the variable label provides some description of the information contained in the variable. Labels for variables associated with specific questions asked directly of parent interview respondents, teachers, and school administrators include the question number. Labels for parent interview variables also have identifiers for the section of the interview in which the question was asked. The labels also include some text related to the actual question wording.

Slide 12 of 16

For example, the label for the variable P4EVALUA is P4 CHQ040 CHD LEARNING ABILITY EVALUATED. The first two characters of the label indicate that the variable pertains to a question asked in the round 4, or spring first-grade, parent interview. The characters CHQ indicate that this question is from the Child Health section of the parent interview. The characters 040 indicate the question number within that parent interview's CHQ section. Lastly, the text CHD LEARNING ABILITY EVALUATED is a paraphrasing of the actual question wording, which asked parents if their child had ever been "evaluated by a professional in response to {his/her} ability to pay attention or learn."

Getting Started with the ECLS-K Data

Due to software limitations placed on the number of characters that could be used for labels, it was not possible to include all of the question text in a label. However, the label conventions were developed to make it easy for researchers to map from a variable in the data file to the corresponding question within the study instrumentation so that they can see the full question text. The label for P4EVALUA tells a researcher to look at question 040 in the child health section of the spring first-grade parent interview to see the question wording.

Slide 13 of 16

The descriptions used in labels for variables that are not associated with specific questions asked directly of respondents are varied in their format. For example, labels for administrative variables, such as the one shown here for variable F2SPECS, contain characters indicating the data source and round like all other variable labels do. However, they do not contain identifiers for a specific survey instrument or a question number because the variables are not derived from one survey item. The description in the label for F2SPECS identifies this as a variable indicating whether the child was receiving special education services as recorded in the Field Management System, or FMS. The wording in the label is abbreviated due to limitations on the number of characters.

Slide 14 of 16

The data file is organized into blocks of data corresponding to specific content areas, as shown here, which include: identification numbers for sampled children and their parents, schools, and teachers; administrative data from the Field Management System; school characteristics from NCES school universe data files, basic child information; full sample weights; child assessment scores; composite variables; child assessment item-level data and child questionnaire data; parent interview data; classroom teacher data; school fact sheet data; school administrator questionnaire data; school facilities checklist data; special education related data; student record abstract data; replicate weights; and Taylor Series sampling strata and primary sampling unit variables.

The variables in the ECLS-K are stored in a certain order, beginning with identifier variables, such as child ID, and ending with the replicate and Taylor Series weights.

Slide 15 of 16

The ECLS-K training modules have been designed to help you understand and analyze ECLS-K data. While completing these modules will provide you with a solid foundation of knowledge, the modules alone do not provide sufficient detail to ensure successful analyses that reflect a comprehensive understanding of the data and what they mean.

Anyone interested in using the data should read through the user's manuals and other technical reports, which provide a wealth of information about how the different measures used in the study were developed, what they indicate, and how data from the measures are provided in the data file.

Getting Started with the ECLS-K Data

In particular each user's manual has a section called data considerations, caveats, anomalies, and errata that users should familiarize themselves with once they know which variables they are interested in using. All data files undergo a thorough review before they are released to the public. During this process, NCES finds issues in the data that users should know about, for example problems due to administration errors or unusual patterns in the data that may or may not be explainable.

Users may want to read through the substantive reports that use ECLS-K data to get a sense of what kind of information is available and what kinds of analyses can be conducted with the data. For example, for each round of collection, NCES releases an initial report intended to provide a first look at data collected at that round. In addition, these release reports contain information about the study design for that round.

All of the publicly available user's manuals, technical reports, and substantive reports published by NCES for the ECLS-K can be found by clicking the underlined screen text on the slide. Additionally, a bibliography listing government reports, research studies published in peer-reviewed journals, and dissertations that use ECLS-K data is available on the [NCES Bibliography Search Tool](#). Note that this NCES Bibliography Search Tool is no longer updated by NCES but remains available as a resource.

It is also important for users to review the data collection instruments that are available online so that they know exactly how questions were asked and understand the groups of people to or about whom questions were asked. As you review the instruments you should pay special attention to the paths through the questionnaire and the skip patterns.

Lastly, it is always a good idea to check the ECLS website occasionally for information about new data releases or issues that have been found in the data of which people should be aware.

Slide 16 of 16

This module has provided you with an overview of the data collected through the ECLS-K that are available for researchers to analyze. This module also described the resources that are available to learn more about the study, the data, and the data files. The module's objectives and the resources provided throughout the module are summarized here for your reference.

You may now proceed to the next module in the series, or click the exit button to return to the landing page.