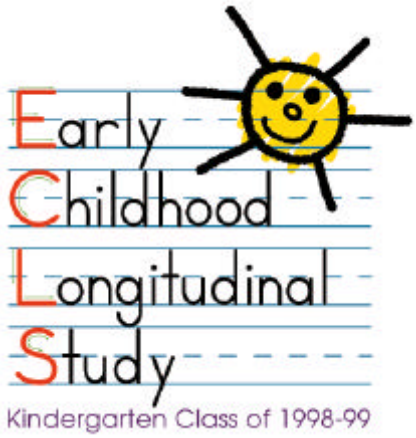
NATIONAL CENTER FOR EDUCATION STATISTICS



USER'S GUIDE TO THE LONGITUDINAL KINDERGARTEN-FIRST GRADE PUBLIC-USE DATA FILE
NCES 2002-149

Prepared by:

Westat
Rockville, Maryland
Karen Tourangeau
Christine Nord
Thanh Le
Siu Wan

The National Center for Education Statistics
U.S. Department of Education
Washington, DC
Jonaki Bose
Jerry West

U.S. Department of Education
Office of Educational Research and Improvement **NCES 2002-149**

U.S. Department of Education
Rod Paige
Secretary

Office of Educational Research and Improvement
Grover J. Whitehurst
Assistant Secretary

National Center for Education Statistics
Gary W. Phillips
Deputy Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, N.W.
Washington, DC 20208-5574

March 2002

The NCES World Wide Web Home Page is: ***http://nces.ed.gov***
The NCES World Wide Web Electronic Catalog is: ***http://nces.ed.gov/pubsearch/index.asp***

TABLE OF CONTENTS

List of Tables

USER'S GUIDE TO THE LONGITUDINAL KINDERGARTEN-FIRST GRADE PUBLIC-USE DATA FILE

1. Introduction

For the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K), the longitudinal kindergarten-first grade public-use data file, referred to hereafter as the K-1 longitudinal data file, combines data from the base (kindergarten) and first grade years. It was created so that analysts can easily examine children's growth and development between kindergarten and first grade without having to go through the process of merging several different data files. Special K-1 longitudinal weights have been developed and are included on the file. This file can be used to study children's learning across school years or to study the extent of summer learning or loss between kindergarten and the fall of the following school year.

There is not a separate user's manual for the K-1 longitudinal data file. To use this file, users should refer to this user's guide and to both the base year and first grade user's manuals (ECLS-K Base Year Public-Use User's Manual [NCES 2001-029], February 2001, and User's Manual for the ECLS-K First Grade Public-Use Data Files and Electronic Code Book [NCES 2002-135]). These manuals are included with the data file on the CD-ROM.

Users will obtain basic information about the K-1 longitudinal data file within this guide. It begins with a description of who is included on the file. It then provides an overview of the content of the data file, the K-1 longitudinal weights, and the data file structure. The last section provides an overview of the electronic code book (ECB). For information on the background of the ECLS-K, the data collection instruments, and the assessment and rating scale scores used in the ECLS-K, users should refer to the base year and first grade user's manuals. These manuals also describe the sample design and implementation and provide detailed information about the data file content and composite variables. For detailed information about the ECB and how to use it, users should refer to the PDF document "ECB User's Guide Chapter 8" found in the root directory of the Longitudinal Kindergarten-First Grade Public-Use ECB. The same information can also be found in chapter 8 of either the base year or first grade user's manuals. Table 1 provides an overview of the content of these manuals (they both follow the same format). This table will help users locate specific information about the ECLS-K.

Table 1.—Contents of the base year and first grade user's manuals

Chapter	Content
Chapter 1 Introduction	Provides information about the background of the ECLS-K, the conceptual model underlying the study, the study components, and the ECLS-K data files.
Chapter 2 Description of Data Collection Instruments	Describes the different data collection instruments, including the direct child assessments, the parent interview, the teacher questionnaires, and the school administrator questionnaire.
Chapter 3 Assessment and Rating Scale Scores Used in the ECLS-K	Provides detailed information on the assessment and rating scale scores developed for the ECLS-K and how to use them.
Chapter 4 Sample Design and Implementation	Provides information on sampling procedures, calculation and use of sample weights, variance estimation, and design effects. **Note:** Information specific to the creation of cross-round (i.e., kindergarten-first grade) weights included in the K-1 longitudinal file is not described in this chapter. Details about the cross-round weights are described in chapter 9 of the first grade user's manual.
Chapter 5 Data Collection Methods and Response Rates	Provides information about the data collection methods, field staff training, response rates, and data collection quality control.
Chapter 6 Data Preparation	Provides information on data preparation, including coding and editing specifications for computer-assisted interviewing and hard-copy questionnaires.
Chapter 7 Data File Content and Composite Variables	Provides detailed information on the identification variables, missing variables, and composite variables that have been created. **Note:** Section 7.7 in the first grade user's manual contains details on how to merge base year school-level data with the first grade child-level data. This is not relevant to the K-1 longitudinal data file. These data have already been merged for users of the K-1 longitudinal data file.
Chapter 8 Electronic Code Book	Provides instructions on the hardware and software requirements of the electronic code book (ECB), its features, how to install, start, and exit the ECB, working with taglists (i.e., lists of commonly used variables), and extracting data from the ECB to create an analysis data file.
Chapter 9 Creating a Longitudinal File (First grade user's manual only)	This chapter was included in the first grade public-use and restricted-use cross-sectional user's manuals for users who wanted to create their own longitudinal file. It provides information on the types of research questions that can be answered with a longitudinal file and provides detailed information about the development and use of the K-1 longitudinal (i.e., cross-round) weights. **Note:** The instructions on merging the files in section 9.3 are not relevant to the K-1 longitudinal data file as the data have already been merged.

2. Individuals Included on the K-1 Longitudinal File

To be included on the K-1 longitudinal file, a parent interview or child assessment must have been conducted in the spring of kindergarten (spring 1999) and in at least one point in time during the first grade year (fall 1999 or spring 2000). Thus, children are included if there were either a parent or child assessment conducted during the spring of kindergarten **and** a parent interview or child assessment was completed in any of the following combinations of time points:

- Fall-first grade or
- Fall-kindergarten and fall-first grade or
- Spring-first grade or
- Fall-kindergarten and spring-first grade or
- Fall-kindergarten and fall-first grade and spring-first grade.

This does not mean that analyses using other combinations of K-1 data cannot be conducted. Even though customized weights were not created for the other K-1 data combinations, existing weights can be used for analyses of these combinations. See the section on K-1 longitudinal weights for details.

The K-1 longitudinal data file is a child-level file. All parent, teacher, and school information collected for any particular child from each round of data collection has been attached to that child's record (a more detailed description of the record layout follows). In all, the K-1 longitudinal data file has 17,212 child records. For detailed information about response rates in each round of data collection, see chapter 5 of the base year and first grade user's manuals.

3. Content

With a few exceptions, the K-1 longitudinal data file contains all the data that were collected from parents, children, teachers, or schools in the base year (fall and spring) and in the first grade (fall and spring) data collections. To streamline the file, however, the data from the household rosters that list all household members, their relationship to the sampled child, and selected other characteristics are not included on the file. The composite variables describing critical household roster-based information such as the children's family structure and selected characteristics of the family members have been retained

on the file (see chapter 7 of the base year and first grade public-use user's manuals for a description of these and other composite variables).

In addition, cross-sectional weights (associated with a single wave of data collection) and within-grade longitudinal weights (for within-kindergarten or within-first grade longitudinal analysis) are not included on the K-1 longitudinal data file, which contains just the K-1 longitudinal weights (for analysis of both kindergarten and first grade data). There are no cross-sectional or within-grade longitudinal weights included on the file because the K-1 longitudinal data file should not be used to examine just the kindergarten data or just the first grade data. The reason it should not be used in this way is that not all children interviewed in kindergarten were interviewed in first grade. Similarly, not all children interviewed in first grade were interviewed in kindergarten. Thus, the population of base year respondents contained in the K-1 longitudinal data file is a subset of those who were interviewed during the base year. Similarly, the population of first grade children on the K-1 longitudinal data file is a subset of those who were interviewed during the first grade data collection.

Similar to the first grade files, the K-1 longitudinal data file contains a few base year variables that were not in the base year files. They fall into three categories: (1) base year recalibrated assessment scores, (2) base year recalibrated academic rating scale (ARS) scores and (3) new and corrected base year composites. The direct child assessment scores were recalibrated to obtain gain scores that could be compared across four waves of data. The ARS scores were recalibrated because an error was identified in the base year ARS scores. Specifically, the fall and spring base year ARS scores used slightly different metrics. These scores were recalibrated using a combined calibration of fall and spring kindergarten ratings. Therefore, the unit for the corrected fall- and spring-kindergarten scores is the same, though comparisons between fall- and spring-kindergarten scores are not recommended. Although the item stems are similar across grades, the actual items include performance criteria that increase from one grade to the next. Moreover, the kindergarten and first grade ARS scores are placed on different metrics. Therefore, change scores should *not* be used between kindergarten and first grade.

The specifics of the ARS and composite problems are described in the first grade public-use user's manual in the chapter titled Base Year Errata and Composites. The other errors listed in that section have either been corrected (errata number 1 through 7) or are not pertinent to the K-1 longitudinal data file (erratum number 8). For example, the base year poverty and urbanicity composites were detected to have errors and were recreated and included with the first grade data file (appendix D) and in the K-1 longitudinal data file. Specifically, WKPOV_R replaces WKPOVRTY and KURBAN_R replaces

KURBAN. Similarly, the imputation flag IF_INC_R replaces IF_INC. Errata numbers 3, 6, and 7 were corrected but did not require replacing existing variables.

4. K-1 Longitudinal Weights

There are several sets of K-1 longitudinal weights computed for children with complete data from different combinations of rounds. All K-1 longitudinal weights are child-level weights. There are no K-1 longitudinal weights at the school or teacher level since school- and teacher-level weights are not computed for the first grade year due to lack of representativeness.

The K-1 longitudinal weights, available on the K-1 longitudinal data file ECB, are described in table 2. This table is designed to help users choose the appropriate weight for their analysis.

Table 2.— ECLS-K: K-1 longitudinal weights

K-1 longitudinal (panel) weights	
Weight	to be used for analysis of ...
C23CW0	child direct assessment data from BOTH spring-kindergarten and fall-first grade, alone or in conjunction with any combination of a limited set of child characteristics (e.g., age, sex, race-ethnicity).
C23PW0	parent interview data from BOTH spring-kindergarten and fall-first grade.
C123CW0	child direct assessment data from fall- AND spring-kindergarten AND fall-first grade, alone or in conjunction with any combination of a limited set of child characteristics (e.g., age, sex, race-ethnicity).
C123PW0	parent interview data from fall- AND spring-kindergarten AND fall-first grade.
C24CW0	child direct assessment data from BOTH spring-kindergarten and spring-first grade, alone or in conjunction with any combination of a limited set of child characteristics (e.g., age, sex, race-ethnicity).
C24PW0	parent interview data from BOTH spring-kindergarten and spring-first grade.
C124CW0	child direct assessment data from fall-kindergarten AND spring-kindergarten AND spring-first grade, alone or in conjunction with any combination of a limited set of child characteristics (e.g., age, sex, race-ethnicity).
C124PW0	parent interview data from fall-kindergarten AND spring-kindergarten AND spring-first grade.

Table 2.— ECLS-K: K-1 longitudinal weights (continued)

K-1 longitudinal (panel) weights	
Weight	to be used for analysis of ...
C1_4CW0	child direct assessment data from all four rounds of data collection, alone or in conjunction with any combination of a limited set of child characteristics (e.g., age, sex, and race-ethnicity).
C1_4PW0	parent interview data from all four rounds of data collection.
Y2COMW0	child direct assessment data from fall-kindergarten AND spring-kindergarten AND spring-first grade, in conjunction with parent and/or teacher data from spring-first grade, AND one or more base year rounds of parent and/or teacher data.

First, decide which two or more points in time are the focus of the analysis. The analysis could pertain to two points in time (e.g., spring-kindergarten and fall-first grade, or spring kindergarten and spring-first grade), three points in time (any three of fall-kindergarten, spring-kindergarten, fall-first grade, and spring-first grade), or four points in time (all four rounds of data). For example, if the analysis uses spring-kindergarten and fall-first grade data, then the appropriate weight would be one that begins with C23 (denoting child-level data from round 2 AND round 3). Second, consider the source of the data, which also affects the choice of the weight. In table 2, details under "to be used in the analysis of …" provide guidance based on whether the data were collected through the child assessments, parent interviews, or teacher questionnaires A or B. For the same example noted earlier, the two weights available are C23CW0 and C23PW0. If parent data from spring-kindergarten and fall-first grade are needed for the analysis, then C23PW0 should be used, otherwise C23CW0 can be used.

K-1 longitudinal weights are used to produce estimates of differences between two or more rounds of data collection spanning across both kindergarten and first grade. Simple examples involving two rounds of data collection are as follows: (1) estimating the differences in children's mean assessment scores between spring kindergarten and spring first grade using C24CW0 and (2) estimating the difference in social rating scores as reported by parents using C24PW0 (social rating scores as reported by teachers and parents are not available for fall-first grade). K-1 longitudinal weights are also used to study the characteristics of children who were assessed in two or more rounds of data collection. For example, one can study the characteristics of kindergarten children who went to summer schools and the effect of summer school attendance on their assessment scores in fall-first grade for children who were assessed in both spring-kindergarten and fall-first grade. If the analysis includes data collected from the parents in rounds 2 and 3, then C23PW0 can be used in the analysis. However, if the analysis involves just the key characteristics (e.g., race) available for most children and the child assessment data from rounds 2 and 3

then C23CW0 can be used to estimate the difference in assessment scores between spring-kindergarten and fall-first grade. An example where data from more than two rounds are used is as follows: examining whether gains children make in their reading knowledge and skills during the kindergarten year and from the end of kindergarten to the end of first grade are related to parents' and teachers' beliefs about kindergarten readiness and parental educational expectations. In this case, the weight Y2COMW0 would be appropriate. As noted in the first grade user's manual, any longitudinal analysis that uses data from fall-first grade will be limited to a 27 percent subsample of children.[*]

There may be combinations of data for which no weights were developed. For example, there is no specific weight to study changes in children's classroom environments as they move from kindergarten to first grade if child assessment or parent data are not used in the analysis. In this example, the data come from part A of the teacher's questionnaire (TQA). The preferred weight for this analysis would be C24CW0, which is the weight for child direct assessment data from both spring-kindergarten and spring-first grade. Of the child records in the K-1 longitudinal file, 86 percent have TQA data from both spring-kindergarten and spring-first grade. In this group, 99 percent have nonzero C24CW0, compared with 86 percent with nonzero Y2COMW0 and 91 percent with nonzero C124CW0, the other two longitudinal weights available for analyses of child data. The preferred weight is the one that will yield the largest number of records for analysis, which is C24CW0 in this case. Analytically, it can be argued that since the direct assessments are conducted in schools, this weight comes closest to capturing the children in participating schools and thus to capturing the children with relevant school environment data. Similarly, if data from the school administrator's questionnaire (SAQ) are used in the analysis of the K-1 longitudinal data, then the same arguments can be used to select the weight. In this case, 73 percent of children have SAQ data from both kindergarten and first grade; of these, 99 percent have nonzero C24CW0 compared with 85 percent with nonzero Y2COMW0 and 90 percent with nonzero C124CW0. Therefore, the preferred weight is also C24CW0. For further advice on which weights to use when analyzing a complex combination of data, contact NCES at *ECLS@ed.gov.*

[*] As described in the first grade user's manual, fall-first grade was a design enhancement whose goal was to enable researchers to study the extent of summer learning losses and gains and the factors associated with them. The fall data collection was limited to students in a 30 percent subsample of schools.

5. Characteristics of Longitudinal Weights

The statistical characteristics of the longitudinal weights are presented in table 3. For each weight, the number of cases with nonzero values is presented together with the mean weight, the standard deviation, the coefficient variation (i.e., the standard deviation as a percentage of the mean weight), the minimum value of the weight, the maximum value of the weight, the skewness, the kurtosis, and the sum of weights.

The difference in the estimate of the population of students (sum of weights) between the different panels of students and types of weights is due to a combination of factors, among them: (1) the number of base year respondents who became ineligible (due to death, leaving the country, or being a nonsampled mover) after the base year, and (2) the adjustment of the weights for the children of unknown eligibility.

Table 3.—Characteristics of child-level K-1 longitudinal weights

Variable name	Number of cases	Mean	Standard deviation	CV (× 100)	Minimum	Maximum	Skewness	Kurtosis	Sum
C23CW0	5,216	739.84	587.55	79.42	68.23	7,182.37	3.98	21.56	3,858,997
C23PW0	4,861	793.83	515.75	64.97	84.26	5,853.21	2.97	13.04	3,858,805
C123CW0	4,729	815.99	646.25	79.20	76.08	7,696.79	3.89	21.55	3,858,824
C123PW0	4,295	898.37	597.89	66.55	95.35	6,421.30	3.05	14.20	3,858,492
C24CW0	16,371	234.81	200.69	85.47	1.78	3,272.40	4.22	31.65	3,844,009
C24PW0	14,938	257.25	198.94	77.34	1.93	2,580.41	3.30	19.64	3,842,784
C124CW0	15,001	256.28	228.52	89.17	1.54	3,877.43	3.71	24.60	3,844,472
C124PW0	13,413	286.40	214.80	75.00	2.06	3,275.79	3.84	26.53	3,841,463
C1_4CW0	4,542	847.78	639.83	75.47	77.56	7,528.68	3.49	18.68	3,850,619
C1_4PW0	4,012	959.07	617.93	64.43	108.75	6,780.92	2.86	13.48	3,847,785
Y2COMW0	13,983	274.83	241.55	87.89	2.03	3,803.82	4.26	29.97	3,842,961

For information about the development of the longitudinal weights, see chapter 9 of the first grade user's manual.

6. Data File Structure

The K-1 longitudinal data file is arranged in blocks of related variables. Table 4 provides an overview of the order and content of these blocks. A more detailed listing of the block arrangements with a description of the topical variable grouping and associated field IDs can be found in the on-line help.

Table 4.— Organization of the K-1 longitudinal data file

Block	Description of types of variables within the block
ID	Identification variables for child, parent, teachers, and schools (e.g., child ID, teacher ID)
Field Management System	Variables taken from the field management system (e.g., whether child received special education services)
Demographic	Demographic variables used in drawing the sample (e.g., the type of school, census region, round the child joined the study, the child's age, and the child's race).
Change flags	Indicators of whether the student changed teachers or schools between the different rounds
Full sample weights	The 11 longitudinal weights
Status flags	Flags indicating the presence or absence of a component for each round of data collection (e.g., presence of fall kindergarten direct child assessment, parent interview, teacher questionnaire A, B, or C, school administrator questionnaire, school facilities checklist, or student record abstract).
Edits and error flags	
Child assessment scores	Re-calibrated child assessment scores, social rating scale scores
Composites	Composite variables constructed for users (e.g., socioeconomic status, family structure)
Data from the instruments	Variables from all the instruments. The data are ordered as follows: direct child assessment data, parent interview data, teacher questionnaire data, school administrator data, school facilities checklist data, and data collected from the student record abstract. Each segment begins with round 1 data (fall-kindergarten) and ends with round 4 data (spring-first grade).
Taylor series identifiers	Taylor series primary sampling unit and strata identifiers for the K-1 longitudinal weights
Replicate weights	Replicate weights (40 replicates for weights based on fall first grade data, 90 replicates in all other cases)

7. Electronic Code Book

The ECLS-K K-1 longitudinal data file CD-ROM contains an ECB that allows users to easily examine the variables in the K-1 longitudinal data file.

The ECLS-K-1 longitudinal data file ECB allows a user to perform the following:

- Search the names and labels of variables in the database (called catalog) to select variables for analysis (see section 8.3, Variable List from the base year or first grade user's manuals).
- Examine the question wording, response categories, and response frequencies for variables the user selects (see section 8.4.9, Viewing Code Book and Variable Information from the base year or first grade user's manuals).
- Create a list of variables (called "a taglist") to be extracted from the catalog, save the list for later use, print the list as a code book, or use a predefined list on the ECB (see section 8.4, Working Taglist, from the base year or first grade user's manuals).
- Automatically generate SAS, SPSS for Windows, or STATA programs to extract selected variables from the whole data set or for a subset of the cases that are defined by the user (see section 8.5, Extracting Data from the ECB from the base year or first grade user's manuals).

The data user can create SAS, SPSS for Windows, and STATA programs that will generate an extract data file from the data file on the CD-ROM. The ECLS-K-1 longitudinal data file ECB does not create a SAS, SPSS for Windows, or STATA data file. It will prepare the statements that can be used with the user's own SAS, SPSS for Windows, or STATA software to create a file. As noted earlier, the CD-ROM contains both ASCII data sets that the ECB uses to extract specific subdata files. The CD-ROM must be in the drive for the data to be extracted.

For detailed information about using the ECB, see the on-line help. Users can also refer to chapter 8 of the base year or first grade user's manuals for information about using ECBs. Chapters 8 in the user's manuals use examples from the base year or the first grade data files. The functionality of the ECB is essentially the same across data sets; only the examples used to illustrate the different ECB features change.

8. K-1 Longitudinal Data File Limiting Fields

The limiting fields for the K-1 longitudinal data file include the following: (1) child sampled in round 3 (R3SAMPLE), (2) child's school type changed between rounds 2 and 4 (R4R2SCHG), (3) first-time kindergartner status (P1FIRKDG), and (4) grade level of the child in the spring-first grade data collection (T4GVL). These limiting fields allow cases belonging to different categories within each variable to be included or excluded from the extraction. The selection indicator will be either a "Yes" or a "No" to specify whether each value (category) for the variable should be included or excluded, respectively. The default setting for every category (value) for the limiting variables is "Yes," meaning that all records will be present in the extract file. To exclude all records that have data for that particular category, the user can change the "Yes" to "No" by double-clicking on it. For example, the user can select "No" for grade levels other than first grade and "Yes" for first grade to create a data file that only includes children in first grade. At least one of the codes must be selected as "Yes" or no records will be extracted for analysis. See chapter 8, section 8.6 of the user's manual for the ECLS-K First Grade Public-Use Data Files and Electronic Code Book (NCES 2002-135) or the on-line help for further details on limiting fields.