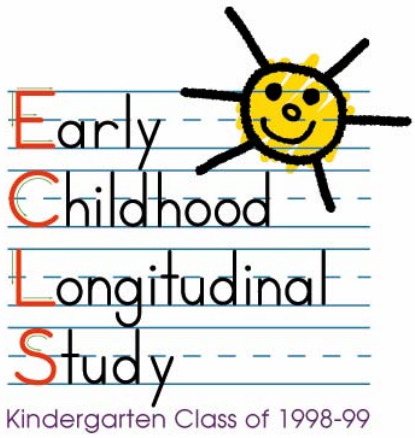
NATIONAL CENTER FOR EDUCATION STATISTICS



USER'S MANUAL FOR THE ECLS-K THIRD GRADE PUBLIC-USE DATA FILE AND ELECTRONIC CODE BOOK

NCES 2004-001

Prepared by

Westat
Rockville, Maryland
Karen Tourangeau
Mike Brick
Thanh Le
Siu Wan
Margaret Weant
Christine Nord
Nancy Vaden-Kiernan
Mary Hagedorn
Elizabeth Bissett
Richard Dulaney
Jean Fowler

Educational Testing Service
Princeton, New Jersey
Judith Pollack
Donald Rock
Michael J. Weiss

University of Toledo
Toledo, Ohio
Sally Atkins-Burnett

The National Center for Education Statistics
U.S. Department of Education
Washington, DC
Elvira Germino Hausken
Jerry West

Education Statistics Services Institute
Washington, DC
Amy Rathbun
Jill Walston

U.S. Department of Education
Institute of Education Sciences

NCES 2004-001

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

List of Tables

TABLE OF CONTENTS (continued)

List of Tables (continued)

List of Tables (continued)

TABLE OF CONTENTS (continued)

List of Tables (continued)

TABLE OF CONTENTS (continued)

List of Tables (continued)

TABLE OF CONTENTS (continued)

List of Tables (continued)

TABLE OF CONTENTS (continued)

List of Tables (continued)

TABLE OF CONTENTS (continued)

List of Tables (continued)

TABLE OF CONTENTS (continued)

List of Exhibits

List of Exhibits (continued)

List of Exhibits (continued)

List of Exhibits (continued)

List of Exhibits (continued)

GETTING STARTED

This document highlights key information you will need to work with the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) data and points you to the appropriate sections of the User's Manual so that you can get started quickly. To read more about any particular topic, go to the indicated section of the User's Manual. In this document, major differences between the third grade data collection and previous rounds are summarized; cautions and caveats about using the data are provided; and basic information about using the Electronic Code Book (ECB) is summarized.

You are working with the public-use file of the third grade data collection. In preparing the public-use file, the National Center for Education Statistics (NCES) takes steps to minimize the likelihood that an individual school, teacher, parent, or child participating in the study can be identified. Every effort is made to protect the identity of individual respondents. Some modifications to the data contained in the restricted-use file have been made to ensure confidentiality. The modifications that are implemented do not affect the overall data quality and most researchers should be able to find all that they need in the public-use files. Chapter 1, section 1.4.1, provides a general description of the differences between the public-use and restricted-use files. Table 7-13 in chapter 7 contains a list of the variables that have been modified. Section 7.8 contains additional information about the "masking" process.

Major Differences in the Third Grade Data Collection

- **Sample is not representative of third grade students, classrooms, or schools.** The ECLS-K base year sample is a representative sample of children attending kindergarten during the 1998–99 school year, of schools with kindergartens, and of kindergarten teachers. Because the first grade sample was freshened with students who had not attended kindergarten in the United States in the previous year, the first grade sample is representative of children attending first grades in the United States during the 1999–2000 school year. However, it is not representative of schools with first grades or of first grade teachers. The third grade sample is not representative of third grade students, third grade teachers, or schools with third grades. Children who started their schooling in the U.S. in second or third grade are not represented in the sample. The data should not be used to make statements about third grade students, schools with third grades, or third grade teachers. See chapter 4, section 4.6 for more details on this point.

- **Children rate their perceptions of social skills and interest in school subjects**. In previous rounds of the ECLS-K, parents and teachers reported about children's social skills. For the first time in the ECLS-K, the children provided information about

themselves by completing a short self-description questionnaire (SDQ). See sections 2.1 and 3.4 for additional information on the SDQ.

- **Social Rating Scale (SRS) is not collected from parents**. In the base year and in spring-first grade, parents and teachers completed the Social Rating Scale, which measures children's approaches to learning, self-control, interpersonal skills, and peer relations. In spring-third grade, only teachers completed this scale. Sections 2.3.2 and 3.3 provide information about the SRS.

- **Science is a separate assessment domain**. In previous years, the direct cognitive assessment included a general knowledge assessment that measured children's knowledge of the social and physical worlds. In third grade, children's knowledge of the world is more categorized into science and social studies domains. With limited time available for the direct assessment, the third grade assessment included only the science domain. Sections 2.1 and 3.1 provide information on the direct cognitive assessments.

Cautions and Caveats

Users of previous rounds of the ECLS-K data have repeatedly asked certain questions. NCES has developed a set of responses to users' most common questions. Please see the NCES web site for commonly asked questions and responses: http://nces.ed.gov/ecls.

In addition to the frequently asked questions and responses, there are other aspects of working with the data that are important to know, including the following:

- **Not all sample children are in third grade**. The third grade data file includes children who were in third grade in spring 2002, and others who were either back or ahead a year or more. Users need to be aware of this fact when using the data and interpreting the findings. Most children in the sample have been in school for at least four years (K-3) and some more (those who were repeating K in the base year). A very small number may have been in school less than four years (some part of the freshened sample added in first grade).

- **Student mobility and its consequences**. A random subsample of students who transferred from their base year schools was flagged to be followed in fall-first grade and in subsequent rounds of data collection. Sections 4.3.1 and 4.4 describe the subsampling of movers. There are a number of variables on the file that can be used to determine if a child moved to a different school between rounds or moved to a different school during the third grade data collection. Section 7.6 describes these variables. Student mobility has a number of consequences for the ECLS-K. It results in a reduction in sample size, fewer children per school, and more missing school and

teacher questionnaire data for movers. See section 5.6.1, tables 5-12 and 5-13 for more information on the response rates for movers and nonmovers.

- **Pay attention to missing data**. Users should always be sure to recode any missing data properly before conducting analyses. If analyzing data over time, it is especially important to check that all skip patterns are the same across years because some changed between rounds of data collection. There are 5 different possible missing data codes on the file. See section 7.2 for a discussion of the different missing values codes and the circumstances when they are used.

- **There may be no perfect weight**. The third grade data file contains 3 sets of cross-sectional weights and 8 longitudinal (panel) weights. Although there are a variety of weights on the file, there are scenarios for which there may not be a perfect weight. For a discussion of the weights and guidance in selecting an appropriate one, refer to sections 4.6.1 and 9.4.1.

- **Defining special populations**. The ECLS-K includes a number of analytic groups of interest that can be identified and studied separately. For example, the third grade file contains variables that identify children who have a disability diagnosed by a professional (P5DISABL), children receiving nonparental child care (P5CARNOW), and those who live in households with incomes below the poverty threshold (W3POVRTY). With variables from earlier rounds of data collection, it is possible to identify children who participated in Head Start in the year prior to kindergarten (HSATTEND from the base year and P4HSBEFK asked of new respondents in spring-first grade) and language minority children (WKLANST), as well as other subgroups. These variables are not contained on the third grade cross-sectional data file, but will be available on the K–3 longitudinal data file to be released in late 2003. Users who desire to study a specific subpopulation should search the ECB using the "NARROW" feature of the ECB to list variables that might help them identify their population of interest. See section 8.3.1 for a description of this feature.

- **Examining school and classroom effects**. When studying the effects of school and classrooms, it is important to restrict the analytic sample to children in the same classroom and/or same schools. Each type of respondent (child, parent, regular teacher, special education teacher, and school) has a unique ID number. These ID numbers can be used to identify children in the same classrooms and schools. Section 7.1 describes the available identification variables.

- **Date of assessments and elapsed times between assessments are not the same for all children**. The ECB contains variables that indicate the month, day, and year in which the direct assessment was administered. The ECB also contains composite variables for children's age at assessment for each sampled child. See the NCES web site http://nces.ed.gov/ecls for information on how to calculate elapsed time period between two assessments.

- **Measuring achievement gains**. One of the major strengths of the ECLS-K is the ability to measure children's achievement gains as they progress from kindergarten through the early elementary grades. There are several different approaches to

measuring gains. See section 3.1.6 for a discussion of measuring gains with the ECLS-K.

ECB Reference Guide

- **Electronic Code Book (ECB).** The ECB is designed to run under Windows 95®, Windows 98®, Windows 2000®, Windows XP®, or Windows NT® on a Pentium-class or higher PC. The PC should have a minimum of 20 megabytes (MB) of available disk space. The ECB offers the most convenient way to access the data because it enables users to search the names and labels of variables, to examine question wording and response categories for individual items, and to generate SAS, SPPS for Windows, or Stata programs for extracting selected variables (see section 8.1.2 for a description of the several features of the ECB). Section 8.2 of the User's Manual contains detailed instructions on how to install and open the ECB. The ECB allows users to easily examine the variables in the ECLS-K ECB data set. The data user can create SAS, SPSS for Windows, and Stata programs that will generate an extract data file from the text (ASCII) data file on the CD-ROM. This text data file is referred to as the "child catalog" and is named child3p.dat in the CD-ROM root directory. For more information about the data file, see section 8.7 of the User's Manual. The ECB CD-ROM also contains Portable Document Format (PDF) files of the associated questionnaires and of the User's Manual.

- **Data File.** The third grade child catalog contains one record for each of 15,305 responding students in spring-third grade. Data collected from teachers and schools are stored in the child catalog. The file, named child3p.dat, is stored in the root directory of the CD-ROM as an ASCII file. It is strongly recommended, however, that users access the data using the ECB software available on the CD-ROM rather than access the ASCII file directly. Appendix B on the CD-ROM contains the record layout for the child catalog.

- **Identification Variables**. The third grade data file contains a child identification variable (CHILDID) that uniquely identifies each record. The same ID is used across each round of the survey. Teachers on the child records are identified with ID variables T5_ID; schools are identified by the ID variables S5_ID. See sections 7.1 and 7.6 in the User's Manual for further information on these identification variables.

- **Instruments**. For the ECLS-K third grade data collection, data were collected using computer-assisted interviewing for parent interviews and child assessments. Self-administered questionnaires were used to collect information from teachers and school administrators or their designees. Chapter 2 of the User's Manual provides an overview of the instruments. To help you decide what variables to use in your analyses, you should always review the actual instruments. Seeing the specific wording of the questions and the context in which they are asked is useful in understanding the results of your analyses and can help minimize errors. Appendix A on the ECLS-K ECB CD-ROM contains, with some exceptions, the third grade instruments. The exceptions are measures that contain copyright-protected materials

and instruments covered by agreements with the test publishers that restrict distribution.

- **Composite Variables.** Numerous composites have been constructed for the ECLS-K data to make it easier for users to use the data set. Most composite variables were created using two or more variables that are on the data file or using information from other sources. Others are recodes of single variables. Composites based on the child assessment include height, weight, and Body Mass Index (BMI). Composites based on the teacher data include class size, percentage of limited-English-proficient children in the class, and student grade level. Composites based on the school data include the percentage of minority students, school type, and school instructional level. Composites based on the parent data include parent education, poverty status, socioeconomic status. See section 7.4 and table 7-12 of the User's Manual for details on all the composites contained on the third grade public-use data file. It is strongly recommended that users give serious consideration to using the composite variables because these variables represent the best data the study has and some include sources not available on the data file.

- **Assessment Scales.** A key feature of the ELCS-K data is the assessments conducted on each child. These assessments included direct and indirect cognitive assessments and measures of children's social development. Chapter 2 provides a general description of the survey instruments, including the direct and indirect assessments. The third grade direct cognitive assessment contained items in reading, mathematics, and science. See section 3.1 of the User's Manual for details on the direct cognitive assessment and the scores that are available for analysis. Section 3.1.5 of the User's Manual discusses choosing the appropriate score for analysis. Section 3.1.6 discusses approaches to measuring student gains in achievement. The indirect cognitive assessment consisted of the Academic Rating Scale (ARS), which was developed for the ECLS-K to measure teachers' evaluations of students' academic achievement in four domains: language and literacy (reading and writing), science, social studies, and mathematical thinking. See section 3.2 of the User's Manual for more information on the ARS.

  The measures of children's social development consisted of the Teacher Social Rating Scale (SRS), which asked third grade teachers to tell how often students exhibited certain social skills and behaviors, and a Self-Description Questionnaire (SDQ) in which the students rated their own perceptions of competence and interest in reading, mathematics, and all school subjects. Children also rated their competence and popularity with peers and reported on problem behaviors. See sections 3.3 and 3.4 for more information on the SRS and SDQ and the scores that are available for analysis.

- **Sample Design and Weights.** The ECLS-K employs a complex sample design. See chapter 4 for a description of the sample design. In order to obtain accurate estimates, you will need to select the appropriate weights. Section 4.6.1 describes the cross-sectional weights and provides advice for which weight to use for a given type of analysis. See exhibit 4-1 for a summary of the cross-sectional weights available for analysis. A description of the longitudinal weights is provided in chapter 9. Section 9.4.1 describes the K–3 longitudinal (panel) weights and provides advice for which

panel weight to use for a given type of analysis. See exhibit 9-1 for a summary of the K–3 longitudinal (panel) weights.

- **Creating a Longitudinal File**. It is possible to combine the third grade data with data from kindergarten and first grade. Instructions on how to create such a file are provided in chapter 9. Most users, however, will probably want to wait for the release of the public-use longitudinal data set. This data set will be available in 2004.

1. INTRODUCTION

This manual provides guidance and documentation for users of the third grade data[1] of the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K). It begins with an overview of the ECLS-K study. Subsequent chapters provide details on the instruments and measures used, the sample design, weighting procedures, response rates, data collection and processing procedures, and the structure of the data file.

The ECLS-K focuses on children's early school experiences beginning with kindergarten. It is a multisource, multimethod study that includes interviews with parents, the collection of data from principals and teachers, and student records abstracts, as well as direct child assessments. The ECLS-K has been developed under the sponsorship of the U.S. Department of Education, National Center for Education Statistics (NCES). Westat is conducting this study with assistance provided by Educational Testing Service (ETS) in Princeton, New Jersey.

The ECLS-K is following a nationally representative cohort of children from kindergarten through fifth grade. The base year data were collected in the fall and spring of the 1998–99 school year when the sampled children were in kindergarten. A total of 21,260 kindergartners throughout the nation participated.

Two more waves of data were collected in the fall and spring of the 1999–2000 school year when most, but not all, of the base year children were in first grade.[2] The fall-first grade data collection was limited to a 30 percent subsample of schools[3] (see exhibit 1-1). It was a design enhancement to enable researchers to measure the extent of summer learning loss and the factors that contribute to such loss and to better disentangle school and home effects on children's learning. The spring-first grade data collection, on the full sample, was part of the original study design and can be used to measure annual school progress and to describe the first grade learning environment of children in the study. All children assessed during the base year were eligible to be assessed in the spring-first grade data collection regardless of whether they repeated kindergarten, were promoted to first grade, or were promoted to second grade. In addition, children who were not in kindergarten in the United States during the 1998–99

[1] The term "third grade" is used throughout this document to refer to the data collections that took place in the 2001–02 school year, at which time most of the sampled children—but not all of them—were in third grade.
[2] Though the majority of base year children were in first grade during the 1999–2000 school year, about 5 percent of the sampled children were retained in kindergarten and a handful of others were in second grade during the 1999–2000 school year.
[3] Approximately 27 percent of the base year students who were eligible to participate in year 2 attended the 30 percent subsample of schools.

school year and, therefore, did not have a chance to be selected to participate in the base year of the ECLS-K were added to the spring-first grade sample.[4] Such children include immigrants to the United States who arrived after fall 1998 sampling, children living abroad during the 1998–99 school year, children who were in first grade in 1998–99 and repeated it in 1999–2000, and children who did not attend kindergarten. Their addition allows researchers to make estimates for all first graders in the United States rather than just for those who attended kindergarten in the United States in the previous year.

A fifth wave of data was collected in the spring of the 2001–02 school year when most, but not all, of the sampled children were in third grade.[5] In addition to the school, teacher, parent, and child assessment data collection components, children were asked to complete a short self-description questionnaire, which asked them how they thought and felt about themselves both socially and academically. The spring-third grade data collection can be used to measure school progress and to describe the third grade learning environment of children in the study.

Exhibit 1-1. ECLS-K waves of data collection: Years 1998–2004

Data collection	Date of collection	Sample
Fall-kindergarten	Fall 1998	Full sample
Spring-kindergarten	Spring 1999	Full sample
Fall-first grade	Fall 1999	30 percent subsample[1]
Spring-first grade	Spring 2000	Full sample
Spring-third grade	Spring 2002	Full sample
Spring-fifth grade	Spring 2004	Full sample

[1] Fall data collection consisted of a 30 percent sample of schools containing approximately 27 percent of the base year students eligible to participate in year 2.
NOTE: See section 1.3 for a description of the study components.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99

The sample of children in the third grade round of data collection of the ECLS-K represents the cohort of children who were in kindergarten in 1998–99 or in first grade in 1999–2000. Since the sample of children fielded in 2001–02 was not freshened with third graders who did not have a chance to be sampled in kindergarten or first grade (for example, because they were out of the country during their kindergarten and first grade year), this sample of children does not represent all third graders in 2001–02.

[4] Their addition is referred to as "freshening" the sample. See chapter 4, section 4.3.2 for more detail on the freshening process.
[5] Approximately 89 percent of the children interviewed were in third grade during the 2001–02 school year, 9 percent were in second grade, and less than 1 percent were in fourth grade or higher.

The vast majority of children in third grade in the 2001–02 school year are members of the cohort. However, third graders who repeated second or third grade and recent immigrants are not covered. Data were collected from teachers and schools to provide important contextual information about the environment for the sampled children. The teachers and schools are not representative of third grade teachers and schools in 2001–02. For this reason, the only weights produced from the study are for making statements about children, including statements about the teachers and schools of those children.

The final wave of data collection that is currently planned is scheduled for spring 2004 when most of the study children will be in the fifth grade.

The ECLS-K has several major objectives and numerous potential applications. The ECLS-K combines (1) a study of achievement in the elementary years; (2) an assessment of the developmental status of children in the United States at the start of their formal schooling and at key points during the elementary school years; (3) cross-sectional studies of the nature and quality of kindergarten programs in the United States; and (4) a study of the relationship of family, preschool, and school experiences to children's developmental status at school entry and their progress during the kindergarten and early elementary school years.

The ECLS-K is part of a longitudinal studies program comprising two cohorts—a kindergarten cohort and a birth cohort. The birth cohort (ECLS-B) is following a national sample of children born in the year 2001 from birth through first grade. The ECLS-B focuses on the characteristics of children and their families that influence children's first experiences with the demands of formal school, as well as children's early health care and in- and out-of-home experiences. Together these cohorts will provide the depth and breadth of data required to more fully describe and understand children's health and early learning, development, and education experiences.

The ECLS-K has both descriptive and analytic purposes. It provides descriptive data on children's status at school entry, their transition into school, and their progress through fifth grade. The ECLS-K also provides a rich data set that enables researchers to analyze how a wide range of family, school, community, and individual variables affect children's early success in school; to explore school readiness and the relationship between the kindergarten experience and later elementary school performance; and to record children's cognitive and academic growth as they move through elementary school.

1.1 Background

National policymakers and the public at large have increasingly recognized that the prosperity of the United States depends on the successful functioning of the American education system. There is also growing awareness that school reform efforts cannot focus solely on the secondary and postsecondary years but must pay attention to the elementary and preschool years as well. Increased policy interest in the early grades and the early childhood period is reflected in President Bush's No Child Left Behind Act (http://www.ed.gov.nclb) and in his Good Start, Grow Smart initiative (http://www.whitehouse.gov/infocus/earlychildhood).

Efforts to expand and improve early education will benefit from insights gained through analyses of data from the large-scale, nationally representative ECLS-K data and the study's longitudinal design. The ECLS-K database contains information about the types of school programs in which children participate, the services they receive, and repeated measures of the children's cognitive skills and knowledge. The ECLS-K database also contains measures of children's physical health and growth, social development, and emotional well-being, along with information on family background and the educational quality of their home environments.

As a study of early achievement, the ECLS-K allows researchers to examine how children's progress is affected by such factors as placement in high or low ability groups, receipt of special services or remedial instruction, grade retention, and frequent changes in schools attended because of family moves. Data on these early school experiences are collected as they occur, with the exception of their experiences before kindergarten, which are collected retrospectively. Collecting this information as it occurs produces a more accurate measurement of these antecedent factors and enables stronger causal inferences to be made about their relationship to later academic progress.

The ECLS-K enables educational researchers and policy analysts to use a variety of perspectives on early childhood education, using techniques such as multilevel modeling to study how school and classroom factors affect the progress of individual children. The data collected will enable analysts to examine how children's status at school entry and performance in school are jointly determined by an interaction of child characteristics and school and family environments.

Data collected during the kindergarten year serve as baseline measures to examine how schooling shapes later individual development and achievement. The longitudinal nature of the study

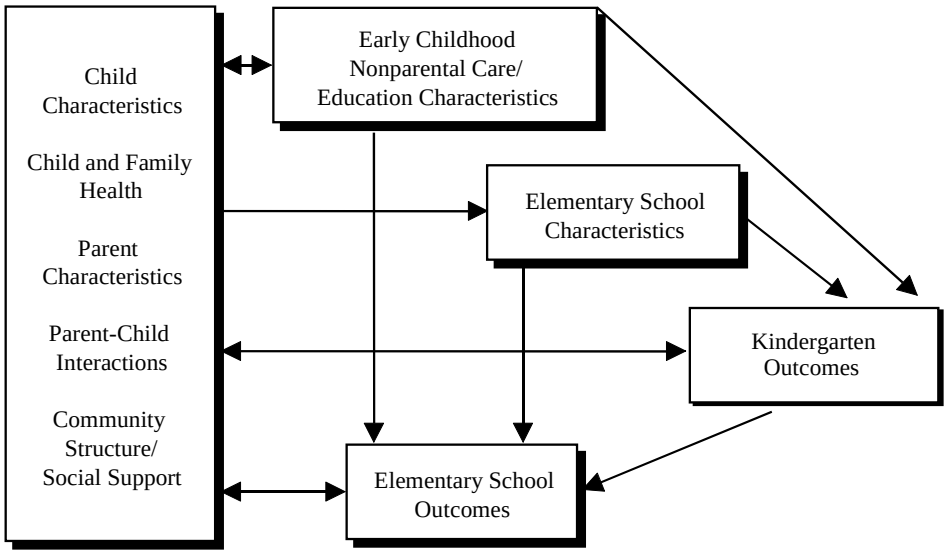
enables researchers to study children's cognitive, social, and emotional growth and to relate trajectories of change to variations in children's experiences in kindergarten and the early grades.

The spring-third grade data collection can be used to describe the diversity of the study children and the classrooms and schools they attend. It can also be used to study children's academic gains in the years following kindergarten. The ECLS-K sample includes substantial numbers of children from various minority groups. Thus, the ECLS-K data present many possibilities for studying cultural and ethnic differences in the educational preferences and literacy practices of families, the developmental patterns and learning styles of children, and the educational resources and opportunities that different groups are afforded in the United States.

1.2 Conceptual Model

The design of the ECLS-K has been guided by a framework of children's development and schooling that emphasizes the interrelationships between the child and family, the child and school, the family and school, and the family, school, and community. The ECLS-K recognizes the importance of factors that represent the child's health status and socioemotional and intellectual development and incorporates factors from the child's family, community, and school-classroom environments. The conceptual model is presented in exhibit 1-2. The study has paid particular attention to the role that parents and families play in helping children adjust to formal school and in supporting their education through the primary grades. It has also gathered information on how schools prepare for and respond to the diverse backgrounds and experiences of the children and families they serve.

Exhibit 1-2. ECLS-K conceptual model



SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99.

1.3 Study Components

The emphasis placed on measuring children's environments and development broadly has critical implications for the design of the ECLS-K. The design of the study includes the collection of data from the child, the child's parents/guardians, teachers, and schools.

- **Children** participate in various activities to measure the extent to which they exhibit those abilities and skills deemed important to success in school. They are asked to participate in activities designed to measure important cognitive (i.e., general knowledge, literacy, and quantitative) and noncognitive (i.e., fine motor and gross motor coordination and socioemotional) skills and knowledge. Most measures of a child's cognitive skills are obtained through an untimed one-on-one assessment of the child. Beginning with the third grade data collection, children report on their own perceptions of their abilities and achievement as well as their interest and enjoyment of reading, math, and other school subjects. Children are assessed in each round of data collection.

- **Parents/guardians** are an important source of information about the families of the children selected for the study and about themselves. Parents provide information about children's development at school entry and their experiences both with family members and others. Information is collected from parents each time children are assessed using computer-assisted interviews (CAIs). Information is collected from parents/guardians in each round of data collection.

- **Teachers**, like parents, represent a valuable source of information on themselves, the children in their classrooms, and the children's learning environment (i.e., the classroom). Teachers are not only asked to provide information about their own backgrounds, teaching practices, and experience, they are also called on to provide information on the classroom setting for the sampled children they teach and to evaluate each sampled child on a number of critical cognitive and noncognitive dimensions. With the exception of the fall-first grade data collection, teachers complete self-administered questionnaires each time children are assessed.

- **School administrators**, or their designees, are asked to provide information on the physical, organizational, and fiscal characteristics of their schools, and on the schools' learning environment and programs. Special attention is paid to the instructional philosophy of the school and its expectations for students. Information is collected from school administrators via self-administered questionnaires during each spring data collection.

- **School office staff** are asked to complete a student records abstract form and a school fact sheet. The student records abstract form includes questions about an individual child's enrollment and attendance at the school, transfer to another school (if applicable), and verifies whether the child has an individualized education plan (IEP) on record. A student records abstract form is completed for each child in the study during each spring data collection.

  During the third grade data collection, school office staff were also asked to complete a school fact sheet. This form supplements the school administrator questionnaire with basic information about the school, including grade level, school type (public or private), length of school year, and attendance recordkeeping practices. This school fact sheet is only filled out once for each school in the study. Prior to the third grade data collection, the questions were part of the school administrator questionnaire.

1.4 ECLS-K Data Files

The ECLS-K data are released in restricted-use and public-use versions. A brief overview of the differences between the restricted-use and public-use files is provided here, followed by a description of the data files that are currently available.

1.4.1 Differences Between ECLS-K Restricted-Use and Public-Use Files

In preparing the public-use files, NCES takes steps to minimize the likelihood that an individual school, teacher, parent, or child participating in the study can be identified. Every effort is made to protect the identity of individual respondents. This is in compliance with the Privacy Act of

1974, as amended, the E-Government Act of 2002, the Education Sciences Reform Act of 2002, and the USA Patriot Act of 2001, which mandate the protection of confidentiality of NCES data that contain individually identifiable information. The process begins with a formal disclosure risk analysis. Variables identified as posing the greatest disclosure risk are altered, and in some instances, entirely suppressed. After modifying individual records that have the greatest risk of disclosure, the disclosure risk analysis is repeated to verify that the risk of disclosure has been reduced to acceptable levels.

The following data modifications account for the differences between the public-use and restricted-use data files:

- Outlier values are top- or bottom-coded;[6]
- Individual cases for which a particular variable poses an especially high risk of disclosure have the value of that variable altered (usually by no more than 5 to 10 percent) to reduce the risk;
- Some continuous variables are modified into categorical variables, and categories of certain categorical variables are collapsed;

[6] To understand top- and bottom-coding, consider a fictitious variable with the following frequency distribution:

Variable X frequency distribution

Value	Count	Percent
Total	4,641	100.00
0	45	1.97
1	193	4.16
2	2,846	61.32
3	1,318	28.40
4	220	4.74
5	18	0.39
6	1	0.02

The outliner values are 0, 1, 4, 5, and 6. Values 0 and 1 are bottom-coded and values 4, 5, and 6 are top-coded. The resulting masked variable has the following frequency:

Masked variable X frequency distribution

Value	Count	Percent
Total	4,641	100.00
≤ 1	238	6.13
2	2,846	61.32
3	1,318	28.40
≥ 4	239	5.15

- A small number of variables with too few cases and a sparse distribution are suppressed altogether, rather than modified; and
- A small number of variables are further masked to enhance confidentiality.

The modifications that are implemented to avoid identification of schools, teachers, parents, and children do not affect the overall data quality and most researchers should be able to find all that they need in the public-use files. While very few of the variables are suppressed, there are a few users who might require the restricted files. Those researchers examining certain rare subpopulations such as the disabled, or children with specific non-English home languages or countries of birth, for example, will find that the restricted-use files contain a few more variables. However, in many instances even though the detailed information on the restricted-use files may be of interest, the sample sizes will be too small to support these analyses. NCES recommends that researchers who are uncertain of which data release to use first examine the public-use files to ascertain whether their specific analytic objectives can be met using those data files.

1.4.2 Overview of Available Data Files

A variety of ECLS-K data files are available for use by analysts. These are described below beginning with the third grade data files.

- **ECLS-K third grade restricted- and public-use data files**. The third grade data are available only as a child-level file. The file includes all data collected from or about the children and their schools including data from the child assessments and from their parents, teachers, and schools. No third grade teacher or school files are released because the sample of teachers and schools is not nationally representative of third grade teachers and schools with third grades. Analysts who wish to examine children's experiences in third grade and the influence of their classroom or school characteristics on their third grade experiences should use the third grade file.

  The third grade data file not only can be used to analyze data collected in the third grade but it also provides weights and variables that can be used in longitudinal data analysis of kindergarten, first grade, and third grade. In addition to the cross-sectional weights, cross-year (kindergarten–third grade) weights have been added to the third grade data file for those analysts who wish to examine children's learning across school years. Instructions on how to create a longitudinal file using the base year, first grade, and third grade data are provided in chapter 9. For more information on the third grade restricted-use data file, please see the *User's Manual for the ECLS-K Third Grade Restricted-Use Data File and Electronic Code Book* (NCES 2003–003). A longitudinal public-use file, however, is available that combines the base year, first

grade, and third grade data (see next bullet). Most analysts will find it more convenient to use the already created longitudinal file described below.

- **Longitudinal kindergarten–third grade (K–third grade) public-use data file**. This public-use data file combines data from the base, first grade, and third grade years. It contains cross-year weights so that analysts can examine children's growth and development between kindergarten and third grade. In order to streamline the file, the household roster that lists all household members, their relationship to the sampled child, and selected other characteristics, is not included on the file. Instead, composite variables describing the children's family structure and selected characteristics of the family members have been added to the file. Analysts who wish to study children's learning across school years, but who do not require the detailed household roster information, should use the longitudinal file. For information about this file, see the *User's Manual for the ECLS-K Longitudinal Kindergarten–Third Grade Public-Use Data Files and Electronic Code Book* (NCES, forthcoming).

- **ECLS-K first grade restricted- and public-use data files**. The first grade data (fall and spring) are available only as a child-level file. The file includes all data collected from or about the children and their schools including data from the child assessments and from their parents, teacher, and schools. No first grade teacher or school files are released because the sample of teachers and schools is not nationally representative of first grade teachers and schools with first grades. Analysts who wish to examine children's experiences in first grade and the influence of their classroom or school characteristics on their first grade experiences should use the first grade file.

  The first grade data file not only can be used to analyze data collected in the first grade but also provides weights and variables that can be used in longitudinal data analysis of both kindergarten and first grade. In addition to the cross-sectional weights, cross-year (kindergarten–first grade) weights have been added to the first grade data file for those analysts who wish to examine children's learning across school years. A longitudinal public-use file, however, is available that combines the base year and first grade data (see next bullet). Most analysts will find it more convenient to use the already created longitudinal file described below. For more information about the first grade file, see the *User's Manual for the ECLS-K First Grade Public-Use Data Files and Electronic Code Book* (NCES 2002–135) or the *User's Manual for the ECLS-K First Grade Restricted-Use Data Files and Electronic Code Book* (NCES 2002–128).

- **Longitudinal kindergarten–first grade (K-first grade) public-use data file**. This public-use data file combines data from the base and first grade years. It contains cross-year weights so that analysts can examine children's growth and development between kindergarten and first grade. In order to streamline the file, the household roster that lists all household members, their relationship to the sampled child, and selected other characteristics is not included on the file. Instead, composite variables describing the children's family structure and selected characteristics of the family members have been added to the file. Analysts who wish to study children's learning across school years or to study the extent of summer learning loss between kindergarten and the fall of the following school year, but who do not require the detailed household roster information, should use the longitudinal file. For

information about this file, see the *User's Manual for the ECLS-K Longitudinal Kindergarten–First Grade Public-Use Data Files and Electronic Code Book* (NCES 2002–149).

- **ECLS-K base year data files**. There are three main and four supplementary files available for the base year. The three main files are the child-level file, the teacher-level file, and the school-level file. The supplementary files are the teacher salary and benefits file, the special education file, the student record abstract file, and the Head Start Verification Study file.

  The child file contains all the data collected from or about the children, including data from the child assessments, and from their teachers, parents, and schools. Analysts who wish to obtain descriptive information about U.S. kindergarten students or their families, or who want to examine relationships involving children and families, children and teachers, or children and schools, should make use of the child file. Analysts wishing to obtain descriptive information about the population of kindergarten teachers in the United States, or to study relationships involving teachers as the principal focus of attention, should use the teacher file. Analysts who want to obtain descriptive information about public and private schools that contain kindergarten classes, or who want to examine relationships among school characteristics, should make use of the school file. These child-, teacher-, and school-level files are available in public-use and restricted-use versions. For more information on these files, refer to the *ECLS-K Base Year Public-Use Data Files and Electronic Code Book: User's Manual* (NCES 2001–029) or the *ECLS-K Restricted-Use Base Year: Child File, Teacher File, and School File* (NCES 2000–097).

- The **salary and benefits file** is collected at the school level and contains information on the base salary, merit pay, and benefit pay of teachers and principals. The salary and benefits data, when combined with other ECLS-K data, can be used to examine, for example, the relationship between student outcomes and school resource allocation and use. This file is only available as a restricted-use file. For more information about this file, see the *ECLS-K Base Year Restricted-Use Salary and Benefits File* (NCES 2001–014).

- The **special education file** is a child-based file that contains information on 784 children identified as receiving special education or related services in kindergarten. Special education teachers were asked to complete two questionnaires designed to collect information about their professional background and experience and about the nature of the special education program and special education services provided to each of the sampled children receiving services. It is only available as a restricted-use file. For more information about this file, see the *ECLS-K Base Year Restricted-Use Special Education Child File* (NCES 2001–015).

- The **student record abstract file** contains information from school records about children's school enrollment and attendance; Individualized Education Plan (IEP) and disability status; and home and school language. The student record abstract form was completed by school staff after the end of the school year. This file is useful in providing additional predictors and correlates of children's transitions to kindergarten and later progress in school. This file is only available as a restricted-use file. For

more information about this file, see the *ECLS-K Base Year Restricted-Use Student Record Abstract File* (NCES 2001–016).

- The **Head Start Verification file** contains information from Head Start program providers. The purpose of the **Head Start Verification Study** was twofold: (1) to identify which of the children reported by either their parents or their schools as having attended Head Start the year prior to kindergarten did indeed attend a Head Start program and (2) to evaluate the process of identifying Head Start participation through parent and school reports and provide further information on the actual process of verifying these reports. This file is a restricted-use file. For more information about this file, see the *ECLS-K Base Year Restricted-Use Head Start File* (NCES 2001–025). The outcomes of the verification process are also included as data items on the ECLS-K first grade and kindergarten–first grade longitudinal files.

1.5 Contents of Manual

This manual provides documentation for users of the third grade public-use data of the ECLS-K. The manual contains information about the data collection instruments (chapter 2) and the psychometric properties of these instruments (chapter 3). It describes the ECLS-K sample design and weighting procedures (chapter 4); data collection procedures and response rates (chapter 5); and data processing procedures (chapter 6). In addition, this manual shows how the public-use third grade data file is structured; provides definitions of composite variables (chapter 7); describes how to install and use the Electronic Code Book (chapter 8); and describes how to use and merge the base year, first grade, and third grade files (chapter 9). The Electronic Code Book contains unweighted frequencies for all variables. Because this manual focuses on the third grade data collection, minimal information is provided about the base year or first grade data. Users who wish to learn more about these data collections should refer to the *ECLS-K Base Year Public-Use Data Files and Electronic Code Book: User's Manual* (NCES 2001–029); the *ECLS-K Restricted-Use Base Year: Child File, Teacher File, and School File* (NCES 2000–097); the *User's Manual for the ECLS-K First Grade Public-Use Data Files and Electronic Code Book* (NCES 2002–135); or the *User's Manual for the ECLS-K First Grade Restricted-Use Data Files and Electronic Code Book* (NCES 2002–128). Additional information about the ECLS program can be found on the World Wide Web at http://nces.ed.gov/ecls.

2. DESCRIPTION OF DATA COLLECTION INSTRUMENTS

This chapter describes the survey instruments used during the third grade data collection of the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K). Exhibit 2-1 lists all the instruments used during the third grade data collection. The instrumentation for the base year and first grade data collections are also shown. Similarities and differences between the third grade instruments and those used in the previous rounds are highlighted throughout this chapter.

The ECLS-K third grade data collection occurred in the spring of the 2001–02 school year. Data were collected using computer-assisted interviewing (CAI) for parent interviews and child assessments. Self-administered questionnaires were used to collect information from teachers (teacher questionnaires and special education teacher questionnaires) and school administrators or their designees (school administrator questionnaire, school fact sheet, and student records abstract). Field staff completed the school facilities check list. In addition, children completed a short self-description questionnaire on their own as part of the direct child assessments.

The third grade data collection instruments, with some exceptions, are available on the CD-ROM as appendix A. The exceptions are the direct child assessment, the Social Rating Scale (SRS)[1] in the teacher questionnaire, and the Self-Description Questionnaire (SDQ).[2] These latter measures contain copyright-protected materials and agreements with the test publishers that restrict their distribution.

[1] Adapted with permission from *Social Skills Rating System, Elementary Scale A ("How Often?")* (Gresham and Elliott, 1990).
[2] Adapted with permission from *Self-Description Questionnaire I* (Marsh, 1990).

Exhibit 2-1. Instruments used in the ECLS-K, by round of data collection: School years 1998–99, 1999–2000, and 2001–02

Instruments	1998–99 school year		1999–2000 school year		2001–02 school year
	Fall-kindergarten	Spring-kindergarten	Fall-first grade	Spring-first grade	Spring-third grade
Parent interview	X	X	X	X	X
Child assessments	X	X	X	X	X
Teacher questionnaire part A	X	X	X	X[2]	X
Teacher questionnaire part B	X	X	X	X[2]	X
Teacher questionnaire part C	X	X	X	X[2]	X
Special education teacher questionnaire part A		X		X	X
Special education teacher questionnaire part B		X		X	X
Adaptive Behavior Scale		X		X	
Self-Description Questionnaire					X
School administrator questionnaire		X		X[3]	X
Student record abstract		X		X	X
School fact sheet					X
School facilities checklist		X		X	X
Salary and benefits questionnaire[4]		X			
Head Start verification[5]		X			

X Round that included the instrument.

[1] The fall-first grade data collection consisted of a 30 percent subsample of the study schools. See the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) User's Manual for the ECLS-K First Grade Public-Use Data Files and Electronic Code Book* (NCES 2002–135; U.S. Department of Education, National Center for Education Statistics, 2002b) or the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) User's Manual for the ECLS-K First Grade Restricted-Use Data Files and Electronic Code Book* (NCES 2002–128; U.S. Department of Education, National Center for Education Statistics, 2002c) for information about the purposes and methods of the fall-first grade data collection.

[2] In spring-first grade, there were two sets of teacher questionnaires—one for the teachers of children who had made the transition to the first grade or any higher elementary school grade, and the second for teachers of children who were repeating or attending the second year of kindergarten.

[3] In spring-first grade, there were two different school administrator questionnaires—one for school administrators in schools new to the study and one for school administrators in schools that participated in the base year data collection.

[4] The salary and benefits questionnaire collected information on the base salary, merit pay, and health benefit pay of teachers and principals. It was completed by the school or district business administrator or by a private school administrator or headmaster.

[5] The Head Start Verification Study confirmed parent and school reports of children's Head Start participation by matching information on the name and location of the Head Start facilities the children were reported to have attended against a database of Head Start centers. For each match, the center was contacted to confirm that the child had attended the center in the year before kindergarten.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first grade, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

2.1 Direct Child Assessments

One-on-one direct child assessments were administered using both hard-copy instruments and computer-assisted interviewing (CAI) in the spring of the 2001–02 school year. The children were assessed regardless of whether they were retained in second grade, promoted to third grade, or moved ahead to fourth. The assessments took about 90 minutes to administer. Exhibit 2-2 displays the major domains measured during the direct child assessments from all five rounds of data collection. As in the previous rounds, the third grade assessments included cognitive and physical components. In addition, the third grade assessment contained a socioemotional component completed by the children. The spring-third grade cognitive assessment scores include measures that can be compared to the base year assessments conducted in the fall of 1998 and the spring of 1999 and to the first grade assessments conducted in the fall of 1999 and the spring of 2000 to study children's gains in reading and mathematics. Chapter 3 contains a detailed description of the scores and information on their use and interpretation.

Exhibit 2-2. ECLS-K direct child assessments, by domain and round of data collection: School years 1998–99, 1999–2000, and 2001–02

Direct child assessment domain	1998–99 school year		1999–2000 school year		2001–02 school year
	Fall-kindergarten	Spring-kindergarten	Fall-first grade	Spring-first grade	Spring-third grade
Language screener (Oral Language Development Scale [OLDS])[1]	X	/	/	/	
Reading (language and literacy)	X	X	X	X	X
Mathematical thinking	X	X	X	X	X
Socioemotional development					X
General knowledge (science and social studies)	X	X	X	X	
Science					X[2]
Psychomotor	X				
Height and weight	X	X	X	X	X

X Round that included the instrument.

/ OLDS was administered to language minority students who were new to the study in the spring or did not pass the cut score in the English version during the previous OLDS administration.

[1] The OLDS was given to children with a non-English language background to determine if the children understood English well enough to receive the direct child assessments in English. For further information on the OLDS, please refer to the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) Base Year Public-Use Data Files and Electronic Code Book: User's Manual* (NCES 2001–029; U.S. Department of Education, National Center for Education Statistics, 2000) or the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) Restricted-Use Base Year Child File, Teacher File, and School File* (NCES 2000–097; U.S. Department of Education, National Center for Education Statistics, 2001). The OLDS was not used in third grade because the vast majority of children passed it by spring-first grade.

[2] In spring-third grade, general knowledge assessment was replaced with a science assessment. Children received a science assessment that measured their understanding of science concepts and scientific investigation skills.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first grade, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

The third grade direct child assessment began by verifying the child's name and administering a short set of warm-up exercises similar in form to the items used in the SDQ (see below). The assessor then administered the SDQ followed by the reading, math, and science assessments, and then by the physical measurements.

2.1.1 Socioemotional Development

To measure children's socioemotional development, the ECLS-K assessors administered the SDQ, which is used to determine how children think and feel about themselves both socially and academically. The SDQ consists of 42 statements. Children rated their perceptions of competence and their interest in reading, mathematics, and "all school subjects." They also rated their perceptions of competence and popularity with peers and reported on problem behaviors with which they might struggle. Each behavior was rated in relation to their perception of themselves on a one to four response scale: "not at all true," "a little bit true," "mostly true," or "very true." The 42 items factored into six scales:

- **SDQ Reading** scale includes eight items about reading grades, the difficulty of reading work, and their interest in and enjoyment of reading.
- **SDQ Mathematics** scale includes eight items about mathematics grades, the difficulty of mathematics work, and their interest in and enjoyment of mathematics.
- **SDQ School** scale includes seven items about how well they do in "all school subjects" and their enjoyment of "all school subjects."
- **SDQ Peer** scale includes six items about how easily they make friends and get along with children as well as their perception of their popularity.
- **SDQ Anger/Distractibility** scale includes six items about externalizing problem behaviors such as fighting and arguing "with other kids," talking and disturbing others, and problems with distractibility.
- **SDQ Sad/Lonely/Anxious** scale includes seven items about internalizing problem behaviors such as feeling "sad a lot of the time," feeling lonely, feeling ashamed of mistakes, and worrying about school and friendships.

The items on the first four scales were adapted with permission from the *Self-Description Questionnaire I* (Marsh, 1990). The items in the two problem behavior scales were developed specifically for the ECLS-K.

Because children of this age have different levels of reading ability, assessors read the SDQ questions to each child even if a child said that he or she could read them. In this way, children's responses were not affected by their reading ability. Children were given a few seconds after each statement was read to mark their response in the SDQ questionnaire. Assessors were trained to maintain a brisk pace so that the children were not tempted to move ahead. The assessors were also trained not to look at the children's answers so that the children would not be tempted to answer in a more positive way then they would have otherwise. The entire questionnaire took about 5 minutes to administer. Assessors put the SDQ away after the child had completed it and entered the answers into the computer after the child had completed the remaining assessments and had left the room.

2.1.2 Cognitive Components

The direct cognitive assessments were individually administered at all five time points. A two-stage cognitive assessment approach was used to maximize the accuracy of measurement and reduce administration time by using the children's responses from a brief first stage routing test to select a second stage form of the appropriate level of difficulty.[3] The kindergarten-first grade (K-1) cognitive assessment focused on three general content areas: (1) reading; (2) mathematics; and (3) knowledge of the social and physical world, referred to as "general knowledge." The K-1 assessment did not ask the children to write anything or to explain their reasoning; rather, children pointed to their answers or responded orally to complete the tasks. The assessment battery was administered using small easels with the items printed on one side and administration instructions for the assessor on the other side. Assessors entered children's responses on a laptop computer.

The third grade direct cognitive assessments, as in previous years, included reading and mathematics domains. By third grade, however, children's knowledge of the world is more categorized into science and social studies domains. With limited time available for direct assessment, the third grade assessment included only the science domain. The third grade assessments also utilized a two-stage design. Easels were used to administer items in mathematics and science. The students also completed workbooks with open-ended mathematics questions. The reading passages and questions were in a

[3] For details on the two-stage assessment design, see the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) Base Year Public-Use Data Files and Electronic Code Book: User's Manual* (NCES 2001–029; U.S. Department of Education, National Center for Education Statistics, 2000) or the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) Restricted-Use Base Year Child File, Teacher File, and School File* (NCES 2000–097; U.S. Department of Education, National Center for Education Statistics, 2001) or the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Psychometric Report for Kindergarten Through First Grade* (NCES 2002–05; U.S. Department of Education, National Center for Education Statistics, 2002a).

booklet format to allow the student to refer back to the story when answering the questions. All questions were read by the assessor. In mathematics and science, all available response options were read to the child. However, the child read the response options in the reading assessment.

The ECLS-K third grade direct cognitive assessment battery was designed to assess children's academic achievement in spring of third grade, and to provide a means of measuring growth since kindergarten entry. Child development and primary education experts consulted on the design and development of the assessment instruments. They recommended that the knowledge and skills assessed by the ECLS-K third grade assessments should represent the typical and important cognitive goals of elementary schools' curricula. The subject matter domains of language use and literacy skills (reading), mathematics, and science were selected. This focus on the main academic subjects of the elementary grades was made because of the central nature of these skills as antecedents of individuals' later educational outcomes.

Pools of test items in each of the content domains were developed by a team of elementary education specialists. Items were chosen to extend the longitudinal scales initiated in kindergarten and first grade, but there were grade-appropriate changes in content and format. Test items were reviewed by elementary school curriculum specialists for appropriateness of content and difficulty, and for relevance to the test framework. In addition items were reviewed for sensitivity issues related to minority concerns. Items that passed these content, construct, and sensitivity screenings were field tested in spring 2000. The content validity of the ECLS-K item pools was established by comparing the results of the ECLS-K with scores on the Woodcock-McGrew-Werder Mini-Battery of Achievement (MBA; Woodcock, McGrew, and Werder, 1994) that was also administered during the field test. Additional information about the development of the third grade cognitive assessment battery can be found in the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Psychometric Report for the Third Grade* (U.S. Department of Education, National Center for Education Statistics, forthcoming[a]).

2.1.2.1 Reading

The K-1 reading (language and literacy) assessment included questions designed to measure basic skills (print familiarity, letter recognition, beginning and ending sounds, rhyming sounds, "sight" word recognition), vocabulary (receptive vocabulary), and comprehension (listening comprehension,

words in context). Comprehension items were targeted to measure skills in initial understanding, developing interpretation, personal reflection, and demonstrating critical stance.

The K-1 reading assessment contained five proficiency levels. These five levels reflect a progression of skills and knowledge. Children are thought to master a level if they pass the items within a level. If a child had mastered one of the higher proficiency levels, he or she was very likely to have passed the items that made up the earlier levels as well. The five levels were as follows: (1) identifying upper- and lower-case letters of the alphabet by name; (2) associating letters with sounds at the beginning of words; (3) associating letters with sounds at the end of words; (4) recognizing common "sight" words; and (5) reading words in context.

The third grade reading assessment included items that were designed to measure phonemic awareness, single word decoding, vocabulary (reading), and passage comprehension. The comprehension items measured skills in initial understanding, developing interpretation, personal reflection, and demonstrating a critical stance. The passage reading section examined sentence, paragraph, and story comprehension and comprised a variety of literary genres including poetry, letters, informational text, and narrative text. The test items marking the highest two K-1 proficiency levels, recognizing common "sight" words and reading words in context, were retained in the third grade assessment. Three higher proficiency levels were added at the third grade level: literal inference, extrapolation, and evaluation.

Thus the third grade reading assessment contained five proficiency levels. These five levels reflected a progression of skills and knowledge: if a child had mastered one of the higher levels, he or she was very likely to have passed the items from the earlier levels as well. The third grade proficiency levels were as follows: (1) recognizing common "sight" words; (2) reading words in context; (3) making inferences using cues that were directly stated with key words in text (literal inference); (4) identifying clues used to make inferences (extrapolation), and using personal background knowledge combined with cues in a sentence to understand use of homonyms; and (5) demonstrating understanding of author's craft and making connections between a problem in the narrative and similar life problems (evaluation).

2.1.2.2 Mathematical Thinking

The K-1 mathematics assessment was designed to measure skills in conceptual knowledge, procedural knowledge, and problem solving. Approximately one-half of the mathematics assessment

consisted of questions on number sense and number properties and operations. The remainder of the assessment included questions in measurement; geometry and spatial sense; data analysis, statistics, and probability; and patterns, algebra, and functions. The mathematics assessment contained several items for which manipulatives were available for children to use in solving the problems. Paper and pencil were also offered to the children to use for the appropriate parts of the assessment.

The items in the K-1 mathematics assessment could also be grouped into five proficiency levels, though the math clusters were less homogeneous in content than the reading clusters. The clusters of math items included the following: (1) identifying some one-digit numerals, recognizing geometric shapes, and one-to-one counting up to ten objects; (2) reading all one-digit numerals, counting beyond ten, recognizing a sequence of patterns, and using nonstandard units of length to compare the size of objects; (3) reading two-digit numerals, recognizing the next number in a sequence, identifying the ordinal position of an object, and solving a simple word problem; (4) solving simple addition and subtraction problems; and (5) solving simple multiplication and division problems and recognizing more complex number patterns.

The third grade mathematics assessment addressed the following content strands: number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and pattern, algebra, and functions. The cognitive processes (conceptual, procedural, and problem solving) are assessed in each of the strands. Some of the items draw upon knowledge from more than one strand. For example, an item might require that a child apply knowledge about geometry, measurement, and number operations to answer the question correctly. Proficiency levels defined in the third grade assessment included levels 4 and 5 retained from the earlier test forms, plus two new levels: place value, and rate and measurement.

Thus the items in the third grade mathematics assessment could be grouped into four proficiency levels. The clusters of third grade mathematics items included the following: (1) solving simple addition and subtraction problems; (2) solving simple multiplication and division problems and recognizing more complex number patterns; (3) demonstrating understanding of place value in integers to hundreds place; and (4) using knowledge of measurement and rate to solve word problems.

2.1.2.3 Science

The K-1 assessment battery differed from the third grade battery. The K-1 battery included a measure of general knowledge whereas the third grade included a measure of science. The K-1 general knowledge assessment battery consisted of items that measured knowledge in the natural sciences and social studies in a single scale. The science subdomain measured two broad classes of science competencies: (1) conceptual understanding of scientific facts and (2) skills and abilities to form questions about the natural world, to answer such questions on the basis of the tools and the evidence collected, to communicate answers, and to explain how the answers were obtained. The social studies subdomain included questions that measured children's knowledge in a wide range of disciplines such as history, government, culture, geography, economics, and law. The science subdomain included questions from the fields of life, earth, space, and physical sciences. The assessment items drew on children's experiences with their environment, and many questions related to more than one of the categories. The items captured information on children's conception and understanding of the social, physical, and natural world and of their ability to draw inferences and comprehend implications. The skills children need to establish relationships between and among objects, events, or people and to make inferences and to comprehend the implications of verbal and pictorial concepts were measured.

The subject matter content of the K-1 general knowledge assessment domain was too diverse and the items insufficiently ranked or graded to permit the formation of a set of proficiency levels. It was also not possible to develop separate scores for science and social studies. Instead, a single score was calculated to represent each child's breadth and depth of understanding and knowledge of the world around them.

As noted previously, the third grade battery addressed the science domain. Equal emphasis was placed on life science, earth and space science, and physical science. Similar to the K-1 assessment of general knowledge, children needed to demonstrate understanding of the physical and natural world, draw inferences, and comprehend relationships. In addition, third-graders needed to interpret scientific data, formulate hypotheses, and identify the best plan to investigate a given question. As with the K-1 general knowledge assessment, no set of proficiency levels was developed.

2.1.3 Physical Components

In the fall of the base year there were two parts to the physical component of the child assessment, psychomotor and anthropometric. The psychomotor component (fine and gross motor) was not administered beyond fall kindergarten. The anthropometric component consisted of recording the children's height and weight in order to measure their physical growth and development. A Shorr Board (for measuring height) and a digital bathroom scale were used to obtain the height and weight measurements, which were recorded on a height and weight recording form and entered into a laptop computer by field staff. Each height and weight was taken twice. For additional detail on the procedures used to collect height and weight, see the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Third Grade Methodology Report* (U.S. Department of Education, National Center for Education Statistics, forthcoming[b]).

2.2 Parent Interview

The third grade parent interview was conducted using a computer-assisted interview (CAI). The parent interview was conducted primarily in English, but provisions were made to interview parents who spoke other languages with bilingual English-Spanish interviewers or interpreters for other languages. Most of the interviews were conducted by telephone, but a small percentage (2 percent) were conducted in person.

The parent interview for the spring-third grade data collection lasted on average 62 minutes and asked approximately 500 questions covering third grade school experiences, child care, parent characteristics, and child health. Exhibit 2-3 provides an overview of the topics covered in the third grade and in the previous rounds of data collection. As can be seen in the table, key topics such as family structure, parental involvement in school, and the child's home environment and cognitive stimulation are covered in most rounds. Other topics, such as parent income, employment, and education, are measured at least once in each school year. The general content areas are similar across the questionnaires, though some topics were added and a few were dropped. For example, in spring-third grade, among the questions added were ones on reading resources in the home (e.g., regular receipt of a newspaper or magazine or the availability of a dictionary or encyclopedia), the respondent's reading practices, and if there was a place set aside for the child to do homework. Topics that were dropped included the parent's report of the Social Rating Scale and attendance at religious services.

Exhibit 2-3. ECLS-K parent interview, by major content topics and round of data collection: School years 1998–99, 1999–2000, and 2001–02

Parent questionnaire topic	1998–99 school year		1999–2000 school year		2001–02 school year
	Fall-kindergarten	Spring-kindergarten	Fall-first grade	Spring-first grade	Spring-third grade
Family structure	X	X	X	X	X
Demographics	X	X	X	X	X
Household roster	X	X	X	X	X
Marital status	X	X	X	X	X
Immigration status		X		X	X[1]
Primary language(s) spoken in home	X	/	/	/	/
Parent's involvement with child's school		X	X	X	X
Child care	X		X	X	X
Current arrangements with relatives	X		X	X	X
Current arrangements with nonrelatives	X		X	X	X
Current arrangements with centers	X		X	X	X
Head Start attendance year before kindergarten	X	/	/	/	
Child care arrangements year before kindergarten	X	/	/	/	
Child's health and well-being	X	X		/	
Birth weight	X	/	/	/	
Physical functioning	X	/	/	X	X
Services for children with special needs	X	/	/	X	X
Social skills rating	X	X		X	
Home environment and cognitive stimulation	X	X	X	X	X
Frequency of literacy activities	X	X	X	X	X
Computer use		X	X	X	X
Television viewing		X	X	X	X
Homework					X
Family routines					X
Summer activities and time use			X		

See notes at end of exhibit.

Exhibit 2-3. ECLS-K parent interview, by major content topics and round of data collection: School years 1998–99, 1999–2000, and 2001–02—Continued

Parent questionnaire topic	1998–99 school year		1999–2000 school year		2001–02 school year
	Fall-kindergarten	Spring-kindergarte	Fall-first grade	Spring-first grade	Spring-third grade
Parental educational expectations for child	X		X	X	X
Neighborhood		X	X	X	X
Safety		X		X	X
Resources (e.g., community center, library)			X		
Parent education	X	/	/	X[2]	X[2]
Parent employment	X			X[2]	X[2]
Parent income		X		X	X
Welfare and other public assistance use	X	X		X	X
Parent/child interaction		X		X	X
Parent discipline		X			X
Parent health and emotional well-being		X			X
Relationships and social support	X	X			X
Marital satisfaction		X			X
Background data	X	X		X	
Mother's age at first birth	X				
Mother's age at child's birth				/	
WIC[3] benefits during pregnancy	X	/	/	/	/
Whether mother worked for pay between when child was born and time child entered kindergarten	X	/	/	/	
Nonresident Parent					
Contact with child	X	X		X	X
School involvement		X			X
Paternity		X		X	X
Child Support		X		X	X

X Round that included the topic.

/ Content areas asked only of new parent respondents in each round.

[1] Asked if new person added to roster or an existing person has missing information on this item.

[2] Updated if changed from previous round.

[3] Supplemental Food Program for Women, Infants, and Children administered by the Food and Nutrition Service, U.S. Department of Agriculture.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first grade, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

The order of preference for the respondent to the parent interview was the same as in previous rounds: (1) the respondent from the previous round (if there was one), (2) the child's mother, (3) another parent or guardian, or (4) some other adult household member. In a majority of the cases (92 percent), the grade 3 respondent was the same as the respondent from the previous round. The child's mother was the respondent in 87 percent of the cases and the child's father in 9 percent.

2.3 Teacher Questionnaires

During the spring-third grade data collection, each child's teacher received a self-administered questionnaire consisting of three distinct parts. The first section, part A, asked about the teacher's classroom and the characteristics of the students, instructional activities and curricular focus, instructional practices in different subject areas (language arts, mathematics, science, and social studies), and student evaluation methods. The teacher was also asked about parent involvement. Only teachers of sampled children completed part A, unlike the base year when all kindergarten teachers in the school, regardless of whether they taught a sampled child, completed it. Part B asked questions on school and staff activities and the teacher's views on teaching, the school environment, and overall school climate. Background questions about the teacher were also included in this section. Teachers were asked to complete one copy of part C for each of the sampled children in their classrooms; in this part, teachers were asked to respond to 39 questions about the child's academic performance. The Academic Rating Scale (ARS) gathered data on each sampled child's skills in areas of language and literacy, mathematical thinking, science, and social studies. Part C also included questions from the Social Rating Scale (SRS) that collected data on five areas of children's social skills. The ARS and SRS are described in more detail in sections 2.3.1 and 2.3.2, respectively. The same teacher questionnaires were completed by the teacher of the sampled child regardless of the child's grade level.

In addition to the teacher questionnaire described above, the ECLS-K also included special education teacher questionnaires described in section 2.4.

Exhibit 2-4 shows the distribution of topics covered in the spring-third grade teacher questionnaires and previous rounds of data collection.

Exhibit 2-4. Teacher questionnaires, by major content topics and round of data collection: School years 1998–99, 1999–2000, and 2001–02

Content topic	1998–99 school year		1999–2000 school year		2001–02 school year
	Fall-kindergarten	Spring-kindergarten	Spring-first grade (First grade teacher)	Spring-first grade (Kindergarten teacher)	Spring-third grade (3rd grade teacher)
Parts A and B					
Description of class—age, race/ethnicity, and sex distribution	X[1]		X[1]	X[1]	X[1]
Class organization					
Activities/interest areas	X[2]	/	X[1]	X[2]	X[1]
Types of materials/ resources		X[1]	X[1]	X[1]	X[1]
Instructional time in different subjects		X[1]	X[1]	X[1]	X[1]
Child vs. teacher-initiated activities	X[2]	X[1]	X[1]	X[2]	X[1]
Homework time in different subjects					X[1]
Time in reading and math achievement groups		X	X	X	X[1]
Classroom characteristics					
Children with special needs		X[1]	X[1]	X[1]	X[1]
Classroom aides		X[1]	X[1]	X[1]	X[1]
Class assignment and grouping		X[1]	X[1]	X[1]	X[1]
Behavior of children in classroom	X[1]	X[1]	X[1]	X[1]	X[1]
Instructional Information					X[1]
Language arts		X	X	X	X[1]
Mathematics		X	X	X	X[1]
Science		X	X	X	X[1]
Social studies		X	X	X	X[1]
Parent involvement		X[1]	X[1]	X[1]	X[1]
Share progress information with parents		X[1]	X[1]	X[1]	
Professional development		X[1]	X[2]	X[1]	X[2]

See notes at end of exhibit.

Exhibit 2-4. Teacher questionnaires, by major content topics and round of data collection: School years 1998–99, 1999–2000, and 2001–02—Continued

Content topic	1998–99 school year		1999–2000 school year		2001–02 school year
	Fall-kindergarten	Spring-kindergarten	Spring-first grade (First grade teacher)	Spring-first grade (Kindergarten teacher)	Spring-third grade (3rd grade teacher)
Teachers' evaluation and grading practices	X[2]	/	X[1]	X[2]	X[1]
Teachers' views on school readiness	X[2]	/	X[2]	X[2]	
Perceptions about school climate	X[2]	/	X[2]	X[2]	X[2]
Perception of personal influence on policies and classroom planning	X[2]	/	X[2]	X[2]	X[2]
Teacher demographic information	X[2]	/	X[2]	X[2]	X[2]
Teacher experience and education	X[2]	/	X[2]	X[2]	X[2]
Job satisfaction	X[2]	/	X[2]	X[2]	X[2]
Transition to school activities	X[2]	/	X[2]	X[2]	
Part C					
Indirect child cognitive evaluation by teacher (ARS)	X	X	X	X	X
Language and literacy, mathematics, general knowledge (science and social studies)	X	X	X	X	X
Social skills (SRS)	X	X	X	X	X
Additional information on sampled child		X	X	X	X
Participation in special services and programs		X	X	X	X

See notes at end of exhibit.

Exhibit 2-4. Teacher questionnaires, by major contact topics and round of data collection: School years 1998–99, 1999–2000, and 2001–02—Continued

Content topic	1998–99 school year		1999–2000 school year		2001–02 school year
	Fall-kindergarten	Spring-kindergarten	Spring-first grade (First grade teacher)	Spring-first grade (Kindergarten teacher)	Spring-third grade (3rd grade teacher)
Overall academic skills and physical activity levels		X	X	X	X
Reading group participation		X	X	X	X
Parental involvement		X	X	X	X

X Round that included the construct.
/ Content areas asked only of new teacher participants.
[1] Topic is in teacher questionnaire part A.
[2] Topic is in teacher questionnaire part B.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first grade, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

2.3.1 Academic Rating Scale

The kindergarten and first grade ARS contained three scales: language and literacy, mathematics, and general knowledge. There are four scales of the third grade ARS: language and literacy, mathematical thinking, science, and social studies. The areas measured in the ARS overlap and augment what is measured in the direct cognitive assessment. The items were designed to ascertain the current skill levels, knowledge, and behaviors of the child in third grade based on the teacher's past observation and experience with the child. In the third grade, the teacher most knowledgeable of each sampled child's skills and knowledge in each of the content areas was asked to complete the ratings. Thus, each sampled child's primary or homeroom classroom teacher was asked to forward the questionnaire to the appropriate content area teacher to complete.

Although the topics covered in the ARS are similar across years, the skills that children exhibit for a particular topic, such as reads fluently, increase by grade. Teachers were provided with examples that helped them establish the level of difficulty of a particular item. For example, reading

fluency is covered in both first and third grade, but the third grade item sets a higher difficulty level, as seen below:

- **Spring-first grade: Reads first grade books fluently—**for example, easily reads words in meaningful phrases rather than reading word by word.
- **Spring-third grade: Reads fluently—**for example, easily reads words as part of meaningful phrases rather than word by word, including words with three or more syllables, such as rambunctious, residential, genuinely, and pneumonia.

Similarly, in mathematics the item about demonstrating understanding of place value has a similar stem in both grades, but the third grade item sets a higher level of difficulty:

- **Spring-first grade: Demonstrates an understanding of place value—**for example, by explaining that fourteen is ten plus four, or using two stacks of ten and five single cubes to represent 25.
- **Spring-third grade: Shows understanding of place value with whole numbers—**for example, correctly orders the numbers 19,321, 14,999, 9,900, and 20,101 from least to greatest, or correctly regroups when adding and subtracting.

Below is a description of the content of the third grade ARS:

- The **Language and Literacy** section of the ARS consists of eight items. Teachers are asked to rate each child's proficiency in expressing ideas, use of strategies to gain information, reading on grade level, and writing.
- In the **Mathematical Thinking** section, teachers rate each child on nine items that tap the following skills: number concepts (place value, fractions, and estimation), data analysis, measurement, operations (division), geometry, application of mathematical strategies, and creating and extending patterns.
- The **Science** section of the ARS consists of seven items. Teachers are asked to rate each child's ability to make predictions, form explanations and conclusions based on observation and investigation, communicate scientific information, apply scientific principles, and demonstrate understanding of life science, earth and space, and physical science.
- The **Social Studies** section of the ARS consists of six items. Teachers are asked to rate each child's knowledge and understanding of cultural differences, economics, geography (map skills and the interaction between humans and the environment), history, and government.

See chapter 3, section 3.2 for scale scores, value ranges, means, and standard deviations for the ARS.

2.3.2 Teacher Social Rating Scale

Teachers rated individual students' social development on part C of the teacher questionnaire. These items are intended to measure approaches to learning, self-control, and interpersonal skills. The items were rated on a scale of one (never) to four (very often). The same five scales defined for the K-1 assessments are formed from these items. Three of the scales capture positive aspects of children's development and two represent problem behaviors. In third grade, examination of the responses suggested a different perception of student's self-control and interpersonal social abilities. The self-control scale includes items on control of attention as well as control of emotions and behavior in interactions. Third grade students who were rated higher on self-control were also rated higher on interpersonal skills that involved peers. Thus the file includes a peer relations score that combines responses on both the interpersonal items and self-control items that relate to peers, as well as these scales reported separately to facilitate comparison with earlier rounds of data collection. See chapter 3, section 3.3 for variable names, ranges, means, and standard deviations for these scales.

- The **Approaches to Learning** scale (Teacher SRS) measures behaviors that affect the ease with which children can benefit from the learning environment. It includes six items that rate the child's attentiveness, task persistence, eagerness to learn, learning independence, flexibility, and organization. In the third grade administration, an item "child follows classroom rules" was added to the SRS to increase variance in the self-control scale.

- The **Self-Control** scale (Teacher SRS) has four items that indicate the child's ability to control behavior by respecting the property rights of others, controlling temper, accepting peer ideas for group activities, and responding appropriately to pressure from peers.

- The **Interpersonal Skills** scale (Teacher SRS) has five items that rate the child's skill in forming and maintaining friendships; getting along with people who are different; comforting or helping other children; expressing feelings, ideas, and opinions in positive ways; and showing sensitivity to the feelings of others.

- The **Peer Relations** scale (grade three Teacher SRS) has nine items. The scale is a combination of the items from the **interpersonal skills** and **self-control scales**. In the third grade, the teacher ratings indicated that self-control and interpersonal skills are so strongly related that they form a single scale that represents the child's skill in establishing and maintaining peer relationships.

The two problem behavior scales reflect behaviors that may interfere with the learning process and the child's ability to interact positively in the classroom.

- **Externalizing Problem Behaviors** scale (Teacher SRS) includes acting out behaviors. The kindergarten and first grade forms have five items on this scale that rate the frequency with which a child argues, fights, gets angry, acts impulsively, and disturbs ongoing activities. To increase the variance on this scale, an item was added in third grade asking about the frequency with which a child talks during quiet study time.

- The **Internalizing Problem Behavior** scale (Teacher SRS) asks about the apparent presence of anxiety, loneliness, low self-esteem, and sadness. This scale comprises four items.

These measures are adapted with permission from the instrument *Social Skills Rating Scale: Elementary Scale A ("How Often?")* (Gresham and Elliott, 1990).

2.4 Special Education Teacher Questionnaires

In the spring-third grade data collection, ECLS-K supervisors reviewed accommodation and inclusion information for children who received special education services. During the preassessment phone call with the school coordinator, the field supervisors asked for the names of sampled children receiving special education services, and the names of the teachers providing these services. The supervisor then listed special education staff working with each child (e.g., speech pathologists, reading instructors, and audiologists). Questionnaires were given to these special education teachers and related services providers. If a child received special education services from more than one special education teacher/provider, a field supervisor determined the child's primary special education teacher/service provider. The primary special education teacher/service provider was defined as

- The teacher who managed the child's individualized education plan (IEP);

- The teacher who spent the most amount of time providing special education services to the child; or

- The teacher who was most knowledgeable about the child's special needs and use of assistive technologies.

The spring-third grade special education teacher questionnaires were very similar to the ones used in previous rounds. The only differences were that questions on transition to school were not asked and a few new questions were added. Exhibit 2-5 provides a summary of the content areas addressed in the special education teacher questionnaires in spring-third grade and in the previous rounds. The questionnaires addressed topics such as the child's disability, IEP goals, the amount and type of services used by sampled students, and communication with parents and general education teachers.

Exhibit 2-5. Special education teacher questionnaires, by major content topics and round of data collection: School years 1998–99, 1999–2000, and 2001–02

Content topic	1998–99 school year spring-kindergarten	1999–2000 school year spring-first grade	2001–02 school year spring-third grade
Part A (Teacher Level)			
Teacher's sex	X	X	X
Teacher's age	X	X	X
Teacher's race/ethnicity	X	X	X
Teaching experience	X	X	X
Educational background	X	X	X
Special education teacher background	X	X	X
Location of service provision	X	X	X
Student load per week	X	X	X
Part B (Child Level)			
Disability category	X	X	X
IEP goals for the school year	X	X	X
Extent of services	X	X	X
Types of services provided for the year	X	X	X
Primary placement	X	X	X
Teaching practices, methods, and materials	X	X	X
Assistive technologies used by child	X	X	X
General education goals, expectations, and assessments	X	X	X
Collaboration/communication with child's general education teacher	X	X	X
Frequency of communicating with child's parents	X	X	X

See notes at end of exhibit.

Exhibit 2-5. Special education teacher questionnaires, by major content topics and round of data collection: School years 1998–99, 1999–2000, and 2001–02—Continued

Content topic	1998–99 school year spring-kindergarten	1999–2000 school year spring-first grade	2001–02 school year spring-third grade
Receipt of formal evaluations in the past year	X	X	X
When child first had an individualized education plan (IEP)			X
Likelihood child will have an IEP next school year			X
Percentage of IEP goals that have been met this school year			X

X Round that included the topic.
NOTE: Data were collected only in the spring of each school year.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first grade, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

Part A of the special education teacher questionnaire was designed to collect information about the special education teacher's professional background and experience. Part B asked about the special education services provided to the child and the nature of the child's special education curriculum. The special education teacher of a sampled child or children was asked to complete a copy of part B for each sampled child she or he was responsible for overseeing.

2.5 School Administrator Questionnaire

The principal, administrator, or headmaster at the school attended by the sampled child was asked to complete the school administrator questionnaire in the spring of 2002. This self-administered questionnaire was intended to gather information about the school, student body, teachers, school policies, and administrator characteristics. The questionnaire was divided into nine sections. The first seven sections of the school administrator questionnaire requested mainly factual information about each school and the programs offered at the school. Either a principal or a designee who was able to provide the requested information could complete these sections. The school's principal was asked to complete the remaining two sections concerning his or her background and evaluations of the school climate. Exhibit 2-6 summarizes the content areas addressed in this questionnaire in spring-third grade and previous rounds.

Exhibit 2-6. School administrator questionnaire, by major content topics and round of data collection: School years 1998–99, 1999–2000, and 2001–02

Content topic	1998–99 school year	1999–2000 school year		2001–02 school year
		Spring-first grade		
	Spring-kindergarten	Returning schools	New schools	Spring-third grade
School characteristics	X	/	X	
School type	X		X	
Admission requirements	X			
School size	X	X	X	X
Average daily attendance				X
Student characteristics	X	X	X	X
Race/ethnicity of students	X	X	X	X
Children eligible for special services	X	X	X	X
Types of kindergarten programs	X			
School facilities and resources	X	/	X	X
Computer equipment	X	X	X	X
Community characteristics and school safety	X	X	X	X
Teaching and other school staff characteristics	X	X	X	X
Range of salary paid to teachers	X		X	
Race/ethnicity of staff	X	X	X	X
Full- and part-time staff in different specialties				X
School policies and programs	X	/	X	X
Assessments, testing, and retention	X	X	X	X
School-family-community connections	X	/	X	X
Programs and activities for families	X		X	X
Parent involvement and participation	X	X	X	
Programs for special populations	X	X	X	X
ESL[1] and bilingual education	X	X	X	X
Special education	X	/	X	X
Gifted and talented	X		X	X
Principal characteristics	X	X	X	X
Sex, race/ethnicity, age of principal	X	X	X	X
Experience and education	X	X	X	X
School governance and climate	X	X	X	X
Goals and objectives for teachers	X	X	X	X
School functioning and decisionmaking	X	X	X	X

/ Fewer details on the topic were collected than for new schools.

[1] English as a second language.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first grade, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

For nonresponding and late-responding schools, interviewers were trained to visit the school and encourage the school administrators to complete the questionnaire. If necessary, the interviewers were to sit down with the administrators to help them fill out the questionnaire. However, if the school administrators were still reluctant to complete the full questionnaire, the interviewers were instructed to obtain key information. This key information covered such topics as the school environment, particularly the safety of the school; school policies and practices; school programs for special populations; staffing and teacher characteristics; and principal characteristics.

2.6 School Fact Sheet

The school fact sheet collects basic information about the school including the grades taught in the school, school sector and focus, the length of the school year, and whether the school keeps student attendance records. Some of this information had been included in the school administrator questionnaire and student record abstract in previous rounds. A separate school fact sheet was developed for the spring-third grade round for ease of administration.

2.7 School Facilities Checklist

ECLS-K supervisors completed the facilities checklist during their visits to the school in the spring of third grade. The facilities checklist collects information about the (1) number of portable classrooms on school grounds, (2) presence of security measures, (3) presence of environmental factors that may affect the learning environment, and (4) overall learning climate of the school.

2.8 Student Records Abstract Form

School staff completed the student records abstract form for each sampled child in the spring of kindergarten, first grade, and third grade. This instrument was used to obtain information about the child's attendance record, presence of and details on a child's IEP, and the type of language or English proficiency screening that the school used. A copy of each child's report card was also obtained. The spring-third grade version of the student records abstract form differed from the spring-kindergarten version in two ways: First, no data were collected on the pre-kindergarten Head Start status of children in

the third grade. Second, two questions on the form were modified to enable the school to provide more comprehensive answers to the question of the status of the child in the previous school year (1998–99) and whether a student had an IEP. See chapter 5, section 5.4 for more detail on the collection of these forms.

REFERENCES

Gresham, F.M. and Elliott, S.N. (1990). *Social Skills Rating Scale: Elementary Scale A ("How Often?").* Circle Pines, MN: American Guidance Services, Inc.

Marsh, H.W. (1990). *Self-Description Questionnaire I.* Campbelltown, New South Wales, Australia: University of Western Sydney, Macarthur.

U.S. Department of Education, National Center for Education Statistics. (2000). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) Base Year Public-Use Data Files and Electronic Code Book: User's Manual* (NCES 2001–029). Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics. (2001). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) Restricted-Use Base Year Child File, Teacher File, and School File* (NCES 2000–097). Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics (2002a). *Early Childhood Longitudinal Study–Kindergarten Class of 1998–99 (ECLS-K), Psychometric Report for Kindergarten Through First Grade* (NCES 2002–05). Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics. (2002b). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) User's Manual for the ECLS-K First Grade Public-Use Data Files and Electronic Code Book* (NCES 2002–135). Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics. (2002c). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) User's Manual for the ECLS-K First Grade Restricted-Use Data Files and Electronic Code Book* (NCES 2002–128). Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics (forthcoming [a]). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Psychometric Report for the Third Grade*. Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics (forthcoming [b]). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Third Grade Methodology Report.* Washington, DC: U.S. Department of Education.

Woodcock, R.W., McGrew, K.S., and Werder, J.K. (1994). *Woodcock-McGrew-Werder Mini-Battery of Achievement.* Itasca, IL: Riverside Publishing.

This page intentionally left blank.

3. ASSESSMENT AND RATING SCALE SCORES USED IN THE ECLS-K

Several types of scores were used in the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) to describe children's cognitive and social development during kindergarten through third grade. These scores were for the direct cognitive assessment, the Academic Rating Scale (ARS), the Social Rating Scale (SRS), and the Self-Description Questionnaire (SDQ). Descriptions of the scores for each assessment or scale follow, along with variable names, variable descriptions, and descriptive statistics from the ECLS-K data files.[1] Guidelines for when and how to use each cognitive assessment score are also provided in this chapter.

3.1 Direct Cognitive Assessment

The third grade direct cognitive assessment contained items in reading, mathematics, and science. In each subject area, children received a 15- to 17-item routing test. Performance on the routing items guided the selection and administration of one of three second-stage forms. The second-stage form contained items of appropriate difficulty for the level of ability indicated by the routing items.[2]

The third grade direct cognitive assessment built on the framework established in the kindergarten and first grade rounds of data collection, but differed in several important respects:

- **No English language screening:** In kindergarten and first grade, children who were identified as coming from a language minority background were administered a language-screening assessment, the Oral Language Development Scale (OLDS), prior to administration of the direct cognitive assessments. Once they achieved a score sufficient for assessment in English, the OLDS was not administered in subsequent rounds of data collection. At kindergarten entry, about 15 percent of the ECLS-K participants were found to need screening for English proficiency. By spring of first grade, less than 6 percent of the sample was screened, and nearly two-thirds of them achieved the score required to go on to the rest of the assessment. The number of sampled children who might still lack English proficiency two years later, in third grade, was assumed to be so small that the language screening assessment was unnecessary. Therefore, the OLDS was not administered in the third grade data collection.

[1] This user's manual is applicable to the data gathered during the 2001–02 school year; information contained in this manual about data gathered during the 1998–99 school year (base year of the study) and 1999–01 school year (first grade) is provided primarily for background and comparison purposes.

[2] See chapter 2, section 2.1, for additional information on the two-stage process for the direct cognitive assessments.

- **New assessment instruments:** The four rounds of data collection in kindergarten and first grade used the same set of assessment instruments in reading, mathematics, and general knowledge. Children were routed to different levels of difficulty within each assessment domain depending on their performance on a short routing test in each subject area. Because children's academic skills in third grade could be expected to have advanced beyond the levels covered by the kindergarten/first grade (K-1) assessments, a new set of assessment instruments was developed for the third grade. Some of the K-1 assessment items were retained in the third grade forms to support development of a longitudinal score scale.

- **Science assessment:** The K-1 general knowledge assessment included basic natural science concepts as well as concepts in social studies. For third grade, a science assessment replaced the general knowledge assessment. There was no longitudinal scale for measuring gains in science through third grade, because the third grade science assessment was not comparable to the K-1 general knowledge assessment.

- **Assessment format:** The format of the third grade assessment was similar to that of prior rounds, with some changes to accommodate the more advanced level of the questions. As before, a survey administrator presented the questions to the child and entered responses into a computer for each individually administered assessment. A workbook of one to seven questions that required computations or written responses was added to the third grade mathematics assessment. The reading assessment in third grade was administered in booklet format instead of on an easel to accommodate the length of the reading passages used in the assessment.

- **Item cluster scores:** The K-1 assessment scores included a count of the number right on three questions related to familiarity with conventions of print. Additional cluster scores, based on small numbers of reading and science items, are reported for the third grade assessment and are described in detail below.

One of the critical goals of the ECLS-K was to measure children's growth in cognitive achievement across the early elementary school years. Due to budgetary constraints, data were not collected in 2000–01, when most of the sampled children were in second grade. The absence of second grade data presented a challenge for establishing longitudinal scales to link the first grade to third grade scores. Very few children answered the most difficult items in the spring-first grade data collection correctly. Third grade field-test assessment results indicated that these same items would be too easy for the vast majority of third graders. The ability levels of first graders overlapped with those of third graders only in the tails of the distributions. Without any second grade data, it would have been difficult to place the items reliably along the difficulty scale, making it difficult to accurately estimate cognitive gains from first to third grade. In order to bridge this gap, reading and math assessments were administered to a sample of approximately 900 second graders in 43 schools. While the bridge sample was a convenience sample and was not designed to be nationally representative, efforts were made to include a diverse sample of children and schools. About 77 percent of the bridge sample children were White and 23

percent minority; 30 public schools and 13 religious or other private schools participated; and attention was given to recruiting schools spanning a wide range of socioeconomic (SES) levels. However, because the bridge sample participants did not constitute a nationally representative sample of second graders, and were not part of the ECLS-K longitudinal sample, no scores were reported for this group. The purpose of the bridge sample was to obtain data on the performance of the assessment items, rather than to track the progress of the children themselves, in order that reliable gain scores could be estimated for the first-to-third graders in the ECLS-K sample. The longitudinal scores necessary for measuring gain over time were estimated by pooling the four rounds of kindergarten/first-grade data with the data from the ECLS-K third graders and the second grade bridge sample. Details of the scoring procedures will be described in the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99* (*ECLS-K) Third Grade Methodology Report* (U.S. Department of Education, National Center for Education Statistics forthcoming).

The scores used to describe children's performance on the direct cognitive assessment included broad-based measures that reported performance in each domain as a whole, as well as targeted scores reflecting knowledge of selected content or mastery within a set of hierarchical skill levels. Some of the scores were simple counts of correct answers, while others were based on Item Response Theory (IRT), which uses patterns of correct and incorrect answers to obtain estimates that are comparable across different assessment forms. The different types of scores that were used to describe children's performance on the direct cognitive assessment are described in detail in this chapter. Number-right scores and IRT scale scores measured children's performance on a set of questions with a broad range of difficulty. Standardized scores (T-scores) reported children's performance relative to their peers. Criterion-referenced proficiency scores and item cluster scores evaluated children's performance with respect to subsets of items that mark specific skills.

Tables 3-1 through 3-9 show the types of scores, variable names, descriptions, and summary statistics for the direct cognitive assessment. The name and description for each variable in the tables begin with a "C," indicating that it is a child variable, and a data collection round number, either 1 (fall-kindergarten), 2 (spring-kindergarten), 3 (fall-first grade), 4 (spring-first grade), or 5 (spring-third grade). Weighted means use weight C1_5SC0, the round 1-2-3-4-5 panel weight, for the four kindergarten and first grade rounds, and the round 5 cross-sectional weight, C5CW0, to represent population estimates for third grade. Kindergarten and first grade IRT scores, T-scores, and proficiency probability scores in this data base differed slightly from the corresponding scores in the previously released data files because they were re-estimated along with the bridge and third grade scores. In addition, all kindergarten and first

grade score statistics presented here differ from previous estimates because the panel weight used restricted estimates to children who participated in all five rounds of data collection.

3.1.1 Number-Right Scores

Number-right scores are counts of the raw number of items a child answered correctly. These scores are useful for descriptive purposes only for assessments that are the same for all children. When these scores are for assessments that differ in difficulty, they are not comparable to each other. For example, a student who took the middle difficulty mathematics second-stage form would probably have answered more questions correctly if he or she had taken the easier low form and fewer if the more difficult high form had been administered. For this reason, raw number-right scores were reported in the database only for the first-stage (routing) tests, which were the same for all children being assessed in that round of data collection. The routing test in each subject area consisted of sets of items spanning a wide range of skills. For example, the K-1 reading routing test emphasized pre-reading skills, while the routing test in third grade contained easy and difficult decoding words, selecting the best word to complete a sentence, and a series of questions based on a reading passage. An analyst might use the routing test number-right scores to report actual performance on these particular sets of tasks. Note that because the same routing test was used for the fall-kindergarten through spring-first grade data collections, rounds 1 through 4, score comparisons may be made among these rounds. However, scores on the third grade routing test, which contained more difficult items, are *not* comparable with the kindergarten or first grade number-right scores. The third grade routing test number-right scores should *not* be compared with the kindergarten or first grade routing test number-right scores.

See table 3-1 for the variable names, descriptions, ranges, weighted means, and standard deviations for the routing test number-right scores for the kindergarten and first grade surveys. Table 3-2 has the same information for the third grade routing test.

Table 3-1. Direct cognitive assessment: Routing test number-right, kindergarten/first grade (K-1) assessments: School years 1998–99 and 1999–2000

Variable name	Description	Range of values	Weighted mean	Standard deviation
C1R2RNOR	C1 RC2 Reading Routing #Right - K-1 Assmt	0 - 20	5.9	4.0
C2R2RNOR	C2 RC2 Reading Routing #Right - K-1 Assmt	0 - 20	10.0	4.1
C3R2RNOR	C3 RC2 Reading Routing #Right - K-1 Assmt	0 - 20	11.7	4.2
C4R2RNOR	C4 RC2 Reading Routing #Right - K-1 Assmt	0 - 20	16.3	3.7
C1R2MNOR	C1 RC2 Math Routing #Right - K-1 Assmt	0 - 16	4.5	3.0
C2R2MNOR	C2 RC2 Math Routing #Right - K-1 Assmt	0 - 16	7.3	3.4
C3R2MNOR	C3 RC2 Math Routing #Right - K-1 Assmt	0 - 16	8.9	3.4
C4R2MNOR	C4 RC2 Math Routing #Right - K-1 Assmt	0 - 16	11.8	3.0

NOTE: Table estimates based on C1_5SC0 weight. Table estimates may differ from those reported in earlier user's manuals and the *ECLS-K Psychometric Report for Kindergarten Through First Grade* (NCES 2002–05) because of sample attrition. See chapter 7, section 7.3 for variable naming conventions.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 3-2. Direct cognitive assessment: Routing test number-right, third grade assessment: School year 2001–02

Variable name	Description	Range of values	Weighted mean	Standard deviation
C5R2RNR3	C5 RC2 Reading Routing #Right - Gr3 Assmt	0 - 15	9.9	2.8
C5R2MNR3	C5 RC2 Math Routing #Right - Gr3 Assmt	0 - 17	8.8	4.4
C5SROUNR	C5 Science Routing Test - Number Right	0 - 15	8.2	3.4

NOTE: Table estimates based on C5CW0 weight. See chapter 7, section 7.3 for variable naming conventions.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

3.1.2 Item Response Theory (IRT) Scale Scores; Standardized Scores (T-Scores)

Broad-based scores using the full set of assessment items in reading, mathematics and science were calculated using IRT procedures. The IRT scale scores estimated children's performance on the whole set of assessment questions, while standardized scores (T-scores) reported children's performance relative to their peers on the content domains. IRT makes it possible to calculate scores that can be compared regardless of which second-stage form a child takes. IRT uses the pattern of right, wrong, and omitted responses to the items actually administered in an assessment and the difficulty, discriminating ability, and "guess-ability" of each item to place each child on a continuous ability scale. The items in the routing tests, plus a core set of items shared among the different second-stage forms and

different rounds of data collection, made it possible to establish a common scale. It is then possible to estimate the score the child would have achieved if all of the items in all of the assessment forms had been administered.

IRT has several other advantages over raw number-right scoring. By using the overall pattern of right and wrong responses and the characteristics of each item to estimate ability, IRT can compensate for the possibility of a low-ability student guessing several hard items correctly. If answers on several easy items are wrong, the probability of a correct answer on a difficult item would be quite low. Omitted items are also less likely to cause distortion of scores, as long as enough items have been answered right and wrong to establish a consistent pattern. Unlike raw scoring, which treats omitted items as if they had been answered incorrectly, IRT procedures use the pattern of responses to estimate the probability of correct responses for all assessment questions. Finally, IRT scoring makes possible longitudinal measurement of gain in achievement over time, even though the assessments that are administered are not identical at each point. The common items present in the routing test and in overlapping second-stage forms allow the scores to be placed on the same scale, even as the two-stage design adapts to children's growth over time. As noted earlier, kindergarten and first grade responses were pooled with bridge sample and third grade data to stabilize the longitudinal estimates. In addition, the maximum values of the reading and mathematics scale scores have been extended to include the more difficult items administered in the third grade assessments. As a result, the re-estimated K-1 IRT scores in this database differ from the IRT scores in the kindergarten and first grade files previously released.

The IRT scale scores in the database represent estimates of the number of items students would have answered correctly at each point in time if they had taken all of the 154 questions in all of the first- and second-stage reading forms, the 123 questions in all of the mathematics forms, and the 62 science items. These scores are not integers because they are probabilities of correct answers, summed over all items in the pools. Reading and mathematics gain scores may be obtained by subtracting the IRT scale scores at fall-kindergarten from the IRT scale scores at spring-first grade, spring-first grade from spring-third grade, and so forth. (Note that scores for different subject areas are not comparable to each other because they are based on different numbers of questions and content that is not necessarily equivalent in difficulty, i.e., it would not be correct to assume that a child is doing better in reading than in mathematics because his or her IRT scale score in reading is higher than in mathematics.) Gain scores are not available for science because the science assessment is new in third grade and is not comparable to the general knowledge assessment administered in the earlier rounds.

See table 3-3 for variable names, descriptions, ranges, weighted means, and standard deviations for the IRT scale scores.

Table 3-3. Direct cognitive assessment: Item Response Theory scale scores: School year 2001–02

Variable name	Description	Range of values	Weighted mean	Standard deviation
C1R2RSCL	C1 RC2 Reading IRT Scale Score	0 – 154	27.4	10.3
C2R2RSCL	C2 RC2 Reading IRT Scale Score	0 – 154	38.6	13.4
C3R2RSCL	C3 RC2 Reading IRT Scale Score	0 – 154	44.8	16.4
C4R2RSCL	C4 RC2 Reading IRT Scale Score	0 – 154	66.9	20.9
C5R2RSCL	C5 RC2 Reading IRT Scale Score	0 – 154	106.1	20.7
C1R2MSCL	C1 RC2 Math IRT Scale Score	0 – 123	21.2	8.9
C2R2MSCL	C2 RC2 Math IRT Scale Score	0 – 123	31.1	11.6
C3R2MSCL	C3 RC2 Math IRT Scale Score	0 – 123	38.0	13.3
C4R2MSCL	C4 RC2 Math IRT Scale Score	0 – 123	54.5	16.2
C5R2MSCL	C5 RC2 Math IRT Scale Score	0 – 123	83.3	18.3
C5SSCALE	C5 Science IRT Scale Score	0 – 62	33.5	10.0

NOTE: Table estimates for C1-C4 variables are based on C1_5SC0 weight; estimates for C5 variables are based on C5CW0 weight. Table estimates may differ from those reported in earlier user's manuals and the *ECLS-K Psychometric Report for Kindergarten Through First Grade* (NCES 2002-05) because of re-estimation of scores on a longitudinal scale that includes third grade, and because of sample attrition. See chapter 7, section 7.3 for variable naming conventions.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Standardized scores (T-scores) provide norm-referenced measurements of achievement, that is, estimates of achievement *relative to the population as a whole*. A high mean T-score for a particular subgroup indicates that the group's performance is high in comparison to other groups. It does not represent mastery of a particular set of skills; only that the subgroup's mastery level is greater than a comparison group. Similarly, a change in mean T-scores over time reflects a change in the group's status with respect to other groups. In other words, T-scores provide information on *status compared with children's peers*, while the IRT scale scores and proficiency scores represent *status with respect to achievement on a particular criterion set of assessment items*. The T-scores only provide an indicator of the extent to which an individual or a subgroup ranks higher or lower than the national average and how much this relative ranking changes over time.

The standardized scores reported in the database are transformations of the IRT theta (ability) estimates, rescaled to a mean of 50 and standard deviation of 10 using cross-sectional sample weights for each wave of data. For example, a fall-kindergarten reading T-score of 45 (C1R2RTSC) represents a reading achievement level that is one-half of a standard deviation lower than the mean for the

fall-kindergarten population represented by the assessed sample of ECLS-K participants. If the same child had a reading T-score of 50 in third grade (C5R2RTSC) this would indicate that the child has made up his or her initial deficit and is reading at a level comparable to the national average.

See table 3-4 for variable names, descriptions, and ranges for the standardized T-scores. Weighted means and standard deviations for the kindergarten and first grade scores in this table deviate slightly from the mean 50.0, standard deviation 10.0 metric because of sample attrition.

Table 3-4. Direct cognitive assessment: standardized scores: School year 2001–02

Variable name	Description	Range of values	Weighted mean	Standard deviation
C1R2RTSC	C1 RC2 Reading T-Score	0 - 96	50.6	10.1
C2R2RTSC	C2 RC2 Reading T-Score	0 - 96	50.6	9.9
C3R2RTSC	C3 RC2 Reading T-Score	0 - 96	50.4	9.8
C4R2RTSC	C4 RC2 Reading T-Score	0 - 96	50.2	9.9
C5R2RTSC	C5 RC2 Reading T-Score	0 - 96	50.0	10.0
C1R2MTSC	C1 RC2 Math T-Score	0 - 96	50.2	10.1
C2R2MTSC	C2 RC2 Math T-Score	0 - 96	50.3	10.1
C3R2MTSC	C3 RC2 Math T-Score	0 - 96	50.4	9.8
C4R2MTSC	C4 RC2 Math T-Score	0 - 96	50.5	9.8
C5R2MTSC	C5 RC2 Math T-Score	0 - 96	50.0	10.0
C5STSCOR	C5 Science T-Score	0 - 96	50.0	10.0

NOTE: Table estimates for C1-C4 variables are based on C1_5SC0 weight; estimates for C5 variables are based on C5CW0 weight. Table estimates may differ from those reported in earlier user's manuals and the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) Psychometric Report for Kindergarten Through First Grade* (NCES 2002-05; U.S. Department of Education, National Center for Education Statistics, 2002) because of re-estimation of scores on a longitudinal scale that includes third grade, and because of sample attrition. See chapter 7, section 7.3 for variable naming conventions.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

3.1.3 Item Cluster Scores

Several item cluster scores were reported for the reading and science assessments. These are simple counts of the number right on small subsets of items linked to particular skills. These clusters of items are also included in the broad-range scores described above. Because they are based on very few assessment items, their reliabilities are relatively low. See section 3.1.7 for reliability statistics.

3.1.3.1 Reading

The K-1 reading assessment contained three questions assessing children's familiarity with conventions of print. The score for these questions was obtained by counting the number of correct answers (zero to three) for the following three items, administered while the child was looking at an illustrated story:

- Indicating that reading goes from left to right;
- Going to the beginning of the next line after a line ends; and
- Finding the end of the story.

These items were part of the reading score calculations in the direct cognitive assessment but did not necessarily fit into a hierarchical pattern of skill mastery. For example, some children scored high on print familiarity but could not recognize letters, while others had the reverse pattern. These items were not included in the third grade reading forms because nearly all children had mastered them by the end of first grade.

A set of four relatively difficult decoding items was reported for the third grade assessment. These were words that were unlikely to be in most children's everyday vocabulary but could be sounded out phonetically. The print familiarity scores for the four kindergarten and first grade rounds are based on the same tasks and may be compared with each other; however, the grade three decoding score was an entirely new task, so comparisons with scores in the earlier rounds are not meaningful.

See table 3-5 for variable names, descriptions, ranges, weighted means, and standard deviations for the reading cluster scores: print familiarity and decoding score.

Table 3-5. Direct cognitive assessment: Reading cluster scores: School year 2001–02

Variable name	Description	Range of values	Weighted mean	Standard deviation
C1R2RPRN	C1 RC2 Print Familiarity	0 – 3	1.8	1.1
C2R2RPRN	C2 RC2 Print Familiarity	0 – 3	2.4	0.9
C3R2RPRN	C3 RC2 Print Familiarity	0 – 3	2.6	0.8
C4R2RPRN	C4 RC2 Print Familiarity	0 – 3	2.8	0.6
C5R2RDEC	C5 RC2 Decoding Score Gr 3	0 – 4	1.1	1.2

NOTE: Table estimates for C1-C4 variables are based on C1_5SC0 weight; estimates for C5 variables are based on C5CW0 weight. Table estimates may differ from those reported on earlier user files because of sample attrition. See chapter 7, section 7.3 for variable naming conventions.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

3.1.3.2 Science

The 15 routing test items of the third grade science assessment tapped a range of basic concepts, with five questions each in life science, physical science, and earth science:

- Life Science: a sample of concepts related to anatomy/health, animal characteristics/ behavior, and ecology;
- Physical Science: a sample of concepts related to states of matter, sound, physical characteristics, and the scientific method; and
- Earth Science: a sample of concepts related to the solar system, earth, soil, minerals, and weather.

Number-right scores for these item clusters are reported. The items were not selected to have comparable levels of difficulty within each set. For example, the mean of 3.0 for the life science cluster compared with 2.6 for earth science does not mean in any sense that children were doing better or learning more relative to the domain curriculum in life science compared with earth science. These clusters simply sample a small set of questions of varying difficulty and content within each domain.

See table 3-6 for variable names, descriptions, ranges, weighted means, and standard deviations for the science cluster scores.

Table 3-6. Direct cognitive assessment: Science cluster scores: School year 2001–02

Variable name	Description	Range of values	Weighted mean	Standard deviation
C5LIFESC	C5 Life Science Gr3	0 – 5	3.0	1.4
C5PHYSSC	C5 Physical Science Gr3	0 – 5	2.7	1.4
C5EARTSC	C5 Earth Science Gr3	0 – 5	2.6	1.3

NOTE: Table estimates based on C5CW0 weight. See chapter 7, section 7.3 for variable naming conventions.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

3.1.4 Proficiency Levels

Proficiency levels provide a means of distinguishing status or gain in specific skills within a content area from the overall achievement measured by the IRT scale scores and T-scores. Clusters of four assessment questions having similar content and difficulty were included at several points along the score scale of the reading and mathematics assessments. Clusters of four items provided a more reliable assessment of proficiency than did single items because of the possibility of guessing; it is very unlikely that a student who had not mastered a particular skill would be able to guess enough answers correctly to pass a four-item cluster. The following 8 reading and 7 mathematics proficiency levels (exhibits 3-1 and 3-2) were identified in the reading and mathematics assessments for kindergarten through third grade. No proficiency scores were computed for the science assessment because the questions did not follow a hierarchical pattern.

The proficiency levels were assumed to follow a Guttman model, that is, a student passing a particular skill level was expected to have mastered all lower levels; a failure should be consistent with nonmastery at higher levels. Only a very small percentage of students in kindergarten through third grade had response patterns that did not follow the Guttman model, that is, a failing score at a lower level followed by a pass on a more difficult item cluster. Overall, including all five rounds of data collection, less than 7 percent of reading response patterns, and less than 5 percent of math assessment results, failed to follow the expected hierarchical pattern. This does not necessarily indicate a different order of learning for these children; since most of the proficiency-level items were multiple choice, many of these reversals may be due to children guessing.

Exhibit 3-1. Reading proficiency levels, kindergarten through third grade: School years 1998–99, 1999–2000, and 2001–02

Reading proficiency level	Description
Level 1	Letter recognition: identifying upper- and lower-case letters by name
Level 2	Beginning sounds: associating letters with sounds at the beginning of words
Level 3	Ending sounds: associating letters with sounds at the end of words
Level 4	Sight words: recognizing common "sight" words
Level 5	Comprehension of words in context: reading words in context
Level 6	Literal inference: making inferences using cues that are directly stated with key words in text (for example, recognizing the comparison being made in a simile)
Level 7	Extrapolation: identifying clues used to make inferences, and using background knowledge combined with cues in a sentence to understand use of homonyms
Level 8	Evaluation: demonstrating understanding of author's craft (how does the author let you know…), and making connections between a problem in the narrative and similar life problems

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

Exhibit 3-2. Mathematics proficiency levels, kindergarten through third grade: School years 1998–99, 1999–2000, and 2001–02

Mathematics proficiency level	Description
Level 1	Number and shape: identifying some one-digit numerals, recognizing geometric shapes, and one-to-one counting of up to ten objects
Level 2	Relative size: reading all single-digit numerals, counting beyond ten, recognizing a sequence of patterns, and using nonstandard units of length to compare objects
Level 3	Ordinality, sequence: reading two-digit numerals, recognizing the next number in a sequence, identifying the ordinal position of an object, and solving a simple word problem
Level 4	Addition/subtraction; solving simple addition and subtraction problems
Level 5	Multiplication/division: solving simple multiplication and division problems and recognizing more complex number patterns
Level 6	Place value: demonstrating understanding of place value in integers to the hundreds place
Level 7	Rate and measurement: using knowledge of measurement and rate to solve word problems

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

Two types of scores are reported with respect to the proficiency levels: a single indicator of highest level mastered, and a set of IRT-based probability scores, one for each proficiency level. More information on each of these types of scores is provided below.

3.1.4.1 Highest Proficiency Level Mastered

Mastery of a proficiency level was defined as answering correctly at least 3 of the 4 questions in a cluster. This definition results in a very low probability of guessing enough right answers by chance, generally less than 2 percent. At least two incorrect or "don't know" responses indicated lack of mastery. Questions that were answered with an explicit "I don't know" were treated as wrong, while omitted items were not counted. Since the ECLS-K direct cognitive child assessment was a two-stage design (where not all children were administered all items), and since more advanced assessment instruments were administered in third grade, children's data did not include all of the assessment items necessary to determine pass/fail for every proficiency level at each round of data collection. The missing information was not missing at random; it depended in part on children being routed to second stage assessment forms of varying difficulty. In order to avoid bias due to the non-randomness of the missing proficiency level scores, imputation procedures were undertaken to fill in the missing information.

Pass or fail for each proficiency level was based on actual counts of correct or incorrect responses, if they were present. If too few items were administered or answered to determine mastery of a level, a pass/fail score was assigned based on the remaining proficiency scores only if they indicated a pattern that was unambiguous. That is, a "fail" was inferred for a missing level if there were easier cluster(s) that had been failed and no higher cluster passed; or a "pass" was assumed if harder cluster(s) were passed and no easier one failed. In the case of ambiguous patterns (e.g., pass, missing, fail, where the missing level could legitimately be either a pass or a fail), an additional imputation step was undertaken that relied on information from the child's performance on all of the domain items answered in that round of data collection. IRT-based estimates of the probability of a correct answer were computed for each missing assessment item and used to assign an imputed right or wrong answer. These imputed responses were then aggregated in the same manner as actual responses to determine mastery at each of the missing levels. More than 80 percent of the "highest level" scores in both reading and mathematics were determined on the basis of item response data alone; the rest utilized IRT-based probabilities for some or all of the missing items. Scores were not imputed for missing levels that included a reversal (e.g., fail, blank, pass) because no resolution of the missing data could result in a consistent hierarchical pattern.

Scores in the data file represented the highest level of proficiency mastered by each child at each round of data collection, whether this determination was made by actual item responses, by imputation, or by a combination of methods. The highest proficiency level mastered implies that children demonstrated mastery of all lower levels and non-mastery of all higher levels. A zero score indicates non-mastery of the lowest proficiency level. Scores were excluded only if the actual or imputed mastery level data resulted in a reversal pattern as defined above. The highest proficiency level-mastered scores did not necessarily correspond to an interval scale, so in analyzing the data, they should be treated as ordinal.

See table 3-7 for variable names, descriptions, and weighted percentages for the highest proficiency level-mastered scores.

Table 3-7. Third grade direct cognitive assessment: highest proficiency level mastered, in percent: School year 2001–02

Variable name	Description	Below Level 1	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8
C1R2RPF	C1 RC2 Reading Highest Prof Lvl Mastered	35	36	14	13	1	1	0	0	0
C2R2RPF	C2 RC2 Reading Highest Prof Lvl Mastered	7	20	23	37	9	3	1	0	0
C3R2RPF	C3 RC2 Reading Highest Prof Lvl Mastered	4	12	19	40	15	6	2	1	0
C4R2RPF	C4 RC2 Reading Highest Prof Lvl Mastered	1	3	5	14	34	30	11	3	1
C5R2RPF	C5 RC2 Reading Highest Prof Lvl Mastered	0	0	0	2	5	21	26	25	21
C1R2MPF	C1 RC2 Math Highest Prof Lvl Mastered	8	36	36	17	3	0	0	0	†
C2R2MPF	C2 RC2 Math Highest Prof Lvl Mastered	2	14	30	37	15	2	0	0	†
C3R2MPF	C3 RC2 Math Highest Prof Lvl Mastered	2	7	20	39	26	5	0	0	†
C4R2MPF	C4 RC2 Math Highest Prof Lvl Mastered	0	2	6	21	48	20	3	0	†
C5R2MPF	C5 RC2 Math Highest Prof Lvl Mastered	0	0	0	5	20	31	28	15	†

†Not applicable. Eight proficiency levels were defined for reading, 7 for mathematics. See chapter 7, section 7.3 for variable naming conventions.
NOTE: Table estimates for C1-C4 variables based on C1_5SC0 weight; estimates for C5 variables based on C5CW0 weight. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

3.1.4.2 Proficiency Probability Scores

Proficiency probability scores were reported for each of the proficiency levels described above, at each round of data collection. The scores estimate the probability of mastery of each level, and can take on any value from zero to one. An IRT model was employed to calculate the proficiency probability scores, which indicated the probability that a child would have passed a proficiency level, based on the child's whole set of item responses in the content domain. The item clusters were treated as single items for the purpose of IRT calibration, in order to estimate students' probabilities of mastery of each set of skills. The hierarchical nature of the skill sets justified the use of the IRT model in this way.

The proficiency probability scores differed from the highest-level scores in that they could be used to measure gains over time, and from the IRT scale scores in that they targeted specific sets of skills. The proficiency probability scores can be averaged to produce estimates of mastery rates within population subgroups. These continuous measures can provide a close look at individuals' status and change over time. Gains in probability of mastery at each proficiency level allow researchers to study not only the amount of gain in total scale score points but also where along the score scale different children made their largest gains in achievement during a particular time interval. For example, subtracting the level 1 probability at time 1 from the level 1 probability at time 2 indicates whether a student advanced in mastery of the particular set of level 1 skills during this time interval. Thus, students' school experiences can be related to improvements in specific skills.

See tables 3-8 and 3-9 for variable names, descriptions, ranges, weighted means, and standard deviations for the proficiency probability scores in reading and mathematics.

Table 3-8. Third grade direct cognitive assessment: proficiency probability scores—reading: School year 2001–02

Variable name	Description	Range of values	Weighted mean	Standard deviation
C1R2RPB1	C1 RC2 Prob1 - Letter Recognition	0 - 1	0.70	0.36
C1R2RPB2	C1 RC2 Prob2 - Beginning Sounds	0 - 1	0.31	0.33
C1R2RPB3	C1 RC2 Prob3 - Ending Sounds	0 - 1	0.17	0.27
C1R2RPB4	C1 RC2 Prob4 - Sight Words	0 - 1	0.03	0.14
C1R2RPB5	C1 RC2 Prob5 - Word in Context	0 - 1	0.01	0.10
C1R2RPB6	C1 RC2 Prob6 - Literal Inference	0 - 1	0.00	0.04
C1R2RPB7	C1 RC2 Prob7 - Extrapolation	0 - 1	0.00	0.01
C1R2RPB8	C1 RC2 Prob8 - Evaluation	0 - 1	0.00	0.01
C2R2RPB1	C2 RC2 Prob1 - Letter Recognition	0 - 1	0.94	0.17
C2R2RPB2	C2 RC2 Prob2 - Beginning Sounds	0 - 1	0.70	0.32
C2R2RPB3	C2 RC2 Prob3 - Ending Sounds	0 - 1	0.51	0.34
C2R2RPB4	C2 RC2 Prob4 - Sight Words	0 - 1	0.15	0.27
C2R2RPB5	C2 RC2 Prob5 - Word in Context	0 - 1	0.04	0.17
C2R2RPB6	C2 RC2 Prob6 - Literal Inference	0 - 1	0.01	0.08
C2R2RPB7	C2 RC2 Prob7 - Extrapolation	0 - 1	0.00	0.02
C2R2RPB8	C2 RC2 Prob8 - Evaluation	0 - 1	0.00	0.01
C3R2RPB1	C3 RC2 Prob1 - Letter Recognition	0 - 1	0.97	0.13
C3R2RPB2	C3 RC2 Prob2 - Beginning Sounds	0 - 1	0.82	0.27
C3R2RPB3	C3 RC2 Prob3 - Ending Sounds	0 - 1	0.66	0.32
C3R2RPB4	C3 RC2 Prob4 - Sight Words	0 - 1	0.27	0.34
C3R2RPB5	C3 RC2 Prob5 - Word in Context	0 - 1	0.09	0.25
C3R2RPB6	C3 RC2 Prob6 - Literal Inference	0 - 1	0.03	0.14
C3R2RPB7	C3 RC2 Prob7 - Extrapolation	0 - 1	0.01	0.06
C3R2RPB8	C3 RC2 Prob8 - Evaluation	0 - 1	0.00	0.03
C4R2RPB1	C4 RC2 Prob1 - Letter Recognition	0 - 1	1.00	0.04
C4R2RPB2	C4 RC2 Prob2 - Beginning Sounds	0 - 1	0.96	0.13
C4R2RPB3	C4 RC2 Prob3 - Ending Sounds	0 - 1	0.92	0.19
C4R2RPB4	C4 RC2 Prob4 - Sight Words	0 - 1	0.74	0.34
C4R2RPB5	C4 RC2 Prob5 - Word in Context	0 - 1	0.42	0.41
C4R2RPB6	C4 RC2 Prob6 - Literal Inference	0 - 1	0.15	0.29
C4R2RPB7	C4 RC2 Prob7 - Extrapolation	0 - 1	0.04	0.13
C4R2RPB8	C4 RC2 Prob8 - Evaluation	0 - 1	0.02	0.07
C5R2RPB1	C5 RC2 Prob1 - Letter Recognition	0 - 1	1.00	0.00
C5R2RPB2	C5 RC2 Prob2 - Beginning Sounds	0 - 1	1.00	0.00
C5R2RPB3	C5 RC2 Prob3 - Ending Sounds	0 - 1	1.00	0.01
C5R2RPB4	C5 RC2 Prob4 - Sight Words	0 - 1	0.98	0.08
C5R2RPB5	C5 RC2 Prob5 - Word in Context	0 - 1	0.93	0.20
C5R2RPB6	C5 RC2 Prob6 - Literal Inference	0 - 1	0.74	0.35
C5R2RPB7	C5 RC2 Prob7 - Extrapolation	0 - 1	0.42	0.37
C5R2RPB8	C5 RC2 Prob8 - Evaluation	0 - 1	0.26	0.27

NOTE: Table estimates for C1-C4 variables are based on C1_5SC0 weight; estimates for C5 variables are based on C5CW0 weight. Table estimates may differ from those reported in earlier user's manuals and the *ECLS-K Psychometric Report for Kindergarten Through First Grade* (NCES 2002-05) because of re-estimation of scores on a longitudinal scale that includes third grade, and because of sample attrition. See chapter 7, section 7.3 for variable naming conventions.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 3-9. Third grade direct cognitive assessment: proficiency probability scores—mathematics: School year 2001–02

Variable name	Description	Range of values	Weighted mean	Standard deviation
C1R2MPB1	C1 RC2 Prob1 – Count, Number, Shape	0 - 1	0.91	0.19
C1R2MPB2	C1 RC2 Prob2 – Relative Size	0 - 1	0.54	0.35
C1R2MPB3	C1 RC2 Prob3 – Ordinality, Sequence	0 - 1	0.20	0.31
C1R2MPB4	C1 RC2 Prob4 – Add/Subtract	0 - 1	0.04	0.12
C1R2MPB5	C1 RC2 Prob5 – Multiply/Divide	0 - 1	0.00	0.04
C1R2MPB6	C1 RC2 Prob6 – Place Value	0 - 1	0.00	0.00
C1R2MPB7	C1 RC2 Prob7 – Rate & Measurement	0 - 1	0.00	0.00
C2R2MPB1	C2 RC2 Prob1 - Count, Number, Shape	0 - 1	0.99	0.06
C2R2MPB2	C2 RC2 Prob2 - Relative Size	0 - 1	0.83	0.25
C2R2MPB3	C2 RC2 Prob3 - Ordinality, Sequence	0 - 1	0.53	0.39
C2R2MPB4	C2 RC2 Prob4 - Add/Subtract	0 - 1	0.17	0.26
C2R2MPB5	C2 RC2 Prob5 - Multiply/Divide	0 - 1	0.02	0.08
C2R2MPB6	C2 RC2 Prob6 - Place Value	0 - 1	0.00	0.01
C2R2MPB7	C2 RC2 Prob7 - Rate & Measurement	0 - 1	0.00	0.00
C3R2MPB1	C3 RC2 Prob1 - Count, Number, Shape	0 - 1	0.99	0.04
C3R2MPB2	C3 RC2 Prob2 - Relative Size	0 - 1	0.92	0.18
C3R2MPB3	C3 RC2 Prob3 - Ordinality, Sequence	0 - 1	0.73	0.35
C3R2MPB4	C3 RC2 Prob4 - Add/Subtract	0 - 1	0.32	0.33
C3R2MPB5	C3 RC2 Prob5 - Multiply/Divide	0 - 1	0.05	0.14
C3R2MPB6	C3 RC2 Prob6 - Place Value	0 - 1	0.00	0.03
C3R2MPB7	C3 RC2 Prob7 - Rate & Measurement	0 - 1	0.00	0.00
C4R2MPB1	C4 RC2 Prob1 - Count, Number, Shape	0 - 1	1.00	0.01
C4R2MPB2	C4 RC2 Prob2 - Relative Size	0 - 1	0.98	0.08
C4R2MPB3	C4 RC2 Prob3 - Ordinality, Sequence	0 - 1	0.94	0.19
C4R2MPB4	C4 RC2 Prob4 - Add/Subtract	0 - 1	0.70	0.32
C4R2MPB5	C4 RC2 Prob5 - Multiply/Divide	0 - 1	0.23	0.30
C4R2MPB6	C4 RC2 Prob6 - Place Value	0 - 1	0.03	0.11
C4R2MPB7	C4 RC2 Prob7 - Rate & Measurement	0 - 1	0.00	0.02
C5R2MPB1	C5 RC2 Prob1 - Count, Number, Shape	0 - 1	1.00	0.00
C5R2MPB2	C5 RC2 Prob2 - Relative Size	0 - 1	1.00	0.00
C5R2MPB3	C5 RC2 Prob3 - Ordinality, Sequence	0 - 1	1.00	0.01
C5R2MPB4	C5 RC2 Prob4 - Add/Subtract	0 - 1	0.96	0.11
C5R2MPB5	C5 RC2 Prob5 - Multiply/Divide	0 - 1	0.75	0.32
C5R2MPB6	C5 RC2 Prob6 - Place Value	0 - 1	0.39	0.39
C5R2MPB7	C5 RC2 Prob7 - Rate & Measurement	0 - 1	0.14	0.28

NOTE: Table estimates for C1-C4 variables are based on C1_5SC0 weight; estimates for C5 variables are based on C5CW0 weight. Table estimates may differ from those reported in earlier user's manuals and the *ECLS-K Psychometric Report for Kindergarten Through First Grade* (NCES 2002-05) because of re-estimation of scores on a longitudinal scale that includes third grade, and because of sample attrition. See chapter 7, section 7.3 for variable naming conventions.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Some examples of interpretation and use of the proficiency probability scores whose means appear in table 3-8 are the following:

- At entry to kindergarten about 70 percent (mean probability = .70) of children were proficient at letter recognition (C1R2RPB1).
- The largest gains between spring-kindergarten and spring-first grade were made in reading simple sight words, with 15 percent of children having mastered this skill at the end of kindergarten (C2R2RPB4) compared to 74 percent a year later (C4R2RPB4).
- There were only small gains in letter recognition after spring-kindergarten, because most children, 94 percent, knew their letters by this time (C2R2RPB1).
- Children's skills in making inferences based on cues directly stated in text (literal inference) increased dramatically between first and third grade, from 15 percent (C4R2RPB6) to 74 percent (C5R2RPB6).
- In spring-third grade, most children had not yet demonstrated understanding of the author's craft or making connections between a problem in the narrative and similar life problems. Only 26 percent mastered the evaluation level (C5R2RPB8).

Comparisons of subgroups may be made by computing the mean probability for each group at a single point in time, or the mean gain for each group from one time to another. See section 3.1.6 for further discussion of measurement of gain.

3.1.5 Choosing the Appropriate Score for Analysis

Each of the types of scores described earlier measures children's achievement from a slightly different perspective. The choice of the most appropriate score for analysis purposes should be driven by the context in which it is to be used:

- A measure of overall achievement versus achievement in specific skills;
- An indicator of status at a single point in time versus growth over time; and
- A criterion-referenced versus norm-referenced interpretation.

3.1.5.1 Item Response Theory-Based Scores

The scores derived from the IRT model (IRT scale scores, T-scores, proficiency probabilities) were based on all of the child's responses to a subject area assessment. That is, the pattern of right and wrong answers, as well as the characteristics of the assessment items themselves, were used to estimate a point on an ability continuum, and this ability estimate, theta, then provided the basis for criterion-referenced and norm-referenced scores.

- **The IRT scale scores** are overall, **criterion-referenced** measures of status at a point in time. They are useful in identifying **cross-sectional differences** among subgroups in overall achievement level and provide a summary measure of achievement useful for correlational analysis with **status** variables, such as demographics, school type, or behavioral measures.

  The IRT scale scores were used as longitudinal measures of overall growth. Gains made at different points on the scale have qualitatively different interpretations. For example, children who made gains in recognizing letters and letter sounds are learning very different lessons from those who are making the jump from reading words to reading sentences, although the gains in number of scale score points may be the same. Comparison of gain in scale score points is most meaningful for groups that started with similar initial status.

- The **standardized scores (T-scores)** are also overall measures of status at a point in time, but they are **norm-referenced** rather than criterion-referenced. They do not answer the question, "What skills do children have?" but rather "**How do they compare with their peers**?" The transformation to a familiar metric with a mean of 50 and standard deviation of 10 facilitates comparisons in standard deviation units. T-score means may be used longitudinally to illustrate the **increase or decrease in gaps** in achievement among subgroups over time. T-scores are not recommended for measuring individual gains over time. The IRT scale scores or proficiency probability scores are used for that purpose.

- **Proficiency probability scores**, derived from the overall IRT model, are **criterion-referenced** measures of proficiency in **specific skills**. Because each proficiency score targets a particular set of skills, they are ideal for studying the **details of achievement**, rather than the single summary measure provided by the IRT scale scores and T-scores. They are useful as **longitudinal measures of change** because they show not only the extent of gains but also where on the achievement scale the gains are taking place. Thus, they can provide information on differences in skills being learned by different groups, as well as the relationships with processes, both in and out of school, that correlate with learning specific skills. For example, high socioeconomic status (SES) kindergarten children showed very little gain in the lowest reading proficiency level, letter recognition, because they were already proficient in this skill at kindergarten entry. At the same time, low SES children made big gains in basic skills, but most had not yet made major gains in reading words and sentences by the end of

kindergarten. Similarly, the best readers in third grade may be working on learning to make evaluative judgments based on reading material, which would show up as large gains in reading level 8. Less skilled readers may show their largest gains between first and third grade at levels 5 or 6, comprehension of words in context and literal inference. The proficiency level at which the largest change is taking place is likely to be different for children with different initial status, background, and school setting. Changes in proficiency probabilities over time may be used to identify the **process variables** that are effective in promoting achievement gains in specific skills.

3.1.5.2 Scores Based on Number Right for Subsets of Items (Non-IRT Based Scores)

The routing test number-right and item cluster scores do **not** depend on the assumptions of the IRT model. They were derived from item responses on specific subsets of assessment items, rather than estimates based on patterns of overall performance. Highest proficient level mastered also, in theory, was derived from item responses, although a relatively small number of IRT-based estimates were substituted for missing data.

- **Routing test number-right scores** for the third-grade reading, math, and science assessments are based on 15, 17, and 15 items respectively (20, 16, and 12 items for the K-1 reading, math and general knowledge assessments, respectively). They target specific sets of skills and cover a broad range of difficulty. These scores may be of interest to researchers because they are based on a specific set of assessment items, which was the same for all children who took the assessment.

- **Item cluster scores** in reading (e.g., Decoding Score Gr 3) and science (e.g., Life Science Gr 3) are based on a count of the number correct for a particular set of items. Users may wish to relate these scores to process variables to get a perspective that is somewhat different from that of the hierarchical levels of skills. However, with only three to five items in each of these item cluster scores, reliabilities tend to be relatively low.

- **Highest proficiency level mastered** is based on the same sets of items as the proficiency probability scores but consist of a set of dichotomous pass/fail scores, reported as a single highest mastery level. Pass/fail on each of the individual levels in the set is based on whether children were able to answer correctly at least three out of four actual items in each cluster. For about 20 percent of these scores, the item data was supplemented with IRT-based estimates to avoid complications associated with missing data that was not missing at random. The highest proficiency level mastered should be treated as an ordinal variable.

3.1.6 Measuring Gains

This section outlines approaches to measuring gains that rely on multiple criterion-referenced points to identify different patterns of student growth. It describes how analysts might use the proficiency probability scores to address policy questions dealing with subgroup differences in achievement growth over time.

Traditional approaches using a total scale score to measure change may yield uninformative if not misleading results. For example, analysis of the gain in total scale score points in reading between fall- and spring- kindergarten shows an average increase of about 10 points. Subgroup analysis shows nearly identical average gains of about the same magnitude for groups broken down by sex, race/ethnicity, SES, and school type, even though the *mean scores* for the subgroups are quite different. Similarly, each of these groups gained about 7 points, on average, on the mathematics scale during the same time, again starting from very different initial status. The *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99* (*ECLS-K) Psychometric Report for Kindergarten Through First Grade* (NCES 2002-05; U.S. Department of Education, National Center for Education Statistics, 2002) describes this analysis in detail.

It would be incorrect to conclude that because different subgroups of children are gaining quantitatively the same number of scale score points, they are learning the same things, or that these gains are qualitatively comparable in any sense. The problem is non-equivalence of scale units: children who gain 10 points at the low end of the scale, for example, by mastering letter recognition and letter sounds, are not learning the same things as more advanced children, who are achieving their 10 point gains by learning to read words and sentences.

The use of adaptive assessments increases the reliability of individual assessment scores by removing the sources of floor and ceiling effects. When assessment forms are matched to children's ability levels, all students have an equal chance to gain on the vertical scale. Depending on how adaptive the measure is, how the scale is constructed, and how even-handed the educational treatment, one may not observe large differences in individual children's amounts of gain in total scale score points. Individual and group differences in the *amount* of gain given a fairly standard treatment (e.g., a year of schooling) can be relatively trivial compared to individual and group differences in *where* the gains take place. It is more likely that one will see substantial subgroup differences in initial status than in gains, suggesting that the gains being made by individuals at different points on the score scale are qualitatively different.

Thus analysis of the total IRT scale score without explicitly taking into consideration where the gain takes place tells only part of the story.

The ECLS-K design utilized adaptive assessments to maximize the accuracy of measurement and minimize floor and ceiling effects, and then to develop an IRT-based vertical scale with multiple criterion-referenced points along that scale. These points, the 8 reading and 7 mathematics proficiency levels described in section 3.1.4, model critical stages in the development of skills. Criterion-referenced points serve two purposes at the individual level: (1) they provide information about changes in each child's mastery or proficiency at *each* level, and (2) they provide information about *where* on the scale the child's gain is taking place. This provides analysts with two options for analyzing achievement gains and relating them to background and process variables. First, gains in probability of proficiency at any level may be aggregated by subgroup, and/or correlated with other variables. Second, the location of maximum gain may be identified for each child by comparing the gains in probability for all of the levels, and focusing on the skills the child is acquiring during a particular time interval.

The probabilities of proficiency at any level may be averaged to estimate the proportion of children mastering the skills marked by that level. For example, the spring-first grade mean for mathematics level 5, "Multiply/Divide," was .23, analogous to 22 percent of the first-grade population demonstrating mastery of this set of items. The mean probability at the end of third grade, .75, is equivalent to a population mastery rate of 75 percent. While most children were making their largest gains at level 5, a small number of children were advancing their skills in solving word problems based on rate and measurement, level 7. The mastery rate for level 7 advanced from near zero at the end of first grade to about 14 percent at the end of third grade. These proportions, and the average gains in the proportions for this particular skill, would very likely be quite different for subgroups of children defined by various demographic and school-process categories. Similarly, gains at each level between time 1 and time 2 may be computed for individual children and treated as outcome variables in multivariate models that include background and process measures.

Another approach entails computing differences in probabilities of proficiency between time 1 and time 2 for all of the proficiency levels. The largest difference marks the mastery level where the largest gain for a given child is taking place: the "locus of maximum gain." The locus of maximum gain is likely to vary for different subgroups of children categorized according to variables of interest. Once having identified mutually exclusive groups of children according to the proximity of their gains to each

of the critical points on the developmental scale, one can treat the different types of gains as qualitatively different outcome measures to be explained by background and process variables.

Each different analytical approach provides a different perspective with respect to understanding student growth. While comparisons of scale score means may be used to capture information about children at a single point in time, analysis of gains in probability of proficiency is more likely to provide useful information about the contribution of background and process variables to gains in achievement over time. Examples of these approaches can be found in Rock and Pollack (2002).

Another important issue to be considered in analyzing achievement scores and gains is assessment timing: children's age at first assessment, assessment dates, and the time interval between successive assessments. Assessment dates ranged from September to November for fall data collections, and from March to June for spring rounds. At kindergarten entry, boys, on average, tend to be older than girls. Children assessed in November of their kindergarten year may be expected to have an advantage over children assessed in the first days or weeks of school. Substantial differences in intervals between assessments may also affect analysis of gain scores. Children assessed in September and June of kindergarten or first grade have more time to learn skills than children assessed in November and March. These differences in intervals may have a relatively small impact on analysis results for long time intervals, such as measuring gains from spring-first grade to spring-third grade, but may be more important within grade, especially fall- to spring-kindergarten. In designing an analysis plan, it is important to consider whether and how differences in ages, assessment dates and intervals may affect the results, to look at relationships between these factors and other variables of interest, and to compensate for differences if necessary.

3.1.7 Reliability

Reliability statistics appropriate for each type of score were computed for each subject area for each round of data collection. For the IRT-based scores, the reliability of the overall ability estimate, theta, is based on the variance of repeated estimates of theta. These reliabilities, ranging from .88 to .96, apply to all of the scores derived from the theta estimate, namely, the IRT scale scores, T-scores, and proficiency probabilities (see table 3-10). Alpha coefficients for the routing test number correct ranged from .75 to .86 for the third grade assessment forms (see table 3-11). The third grade reading alpha is somewhat lower than in earlier rounds, at least in part due to the third grade form having fewer items (15)

than the 20 items in the K-1 routing test. Conversely, the alpha coefficient for the math routing test was slightly higher in third grade. The increase in the number of mathematics routing items, from 16 in the K-1 form to 17 in third grade, probably accounts for a small part of this difference. Split-half reliabilities were computed for the item cluster scores in reading and science (see table 3-12). These reliabilities were higher for the reading clusters (.60 to .67) than for the science scores (.46 to .59). The difference in internal consistency statistics is due to the reading items being essentially replications of the same or similar tasks, while the science items had a greater diversity of content.

The score indicating the highest proficiency level mastered is based on a combination of raw item scores and IRT-based imputations. For a majority of students, highest level mastered could be determined on the basis of actual item responses alone, while for others imputations were required. Thus, the reported score represents a hybrid of different methodologies, as well as a collapsing of separate measurements into one.

Standard measures of reliability are not suitable for assessing the reliability of the highest proficiency level score, for several reasons. Split-half reliabilities at each individual proficiency level could be calculated for children with complete sets of item responses, but would not apply to levels or students whose item response data was supplemented with IRT-based estimates. Similarly, the reliability of the IRT theta would be relevant only to the small percentage of proficiency level scores for which IRT-based estimates alone provided the pass/fail determinations. Furthermore, the score denoting the highest level mastered reduces the series of pass/fail level scores to a single composite, so any reliability estimates obtained for individual levels would not necessarily represent the reliability of the single reported score.

As a result, the statistics reported in table 3-13 are not traditional reliability measures. Another approach was employed, based on the idea that reliability statistics are estimates of consistency of measurement under different circumstances, such as different sets of assessment items, or in this case, different methodologies. The two scoring methods, actual item responses and imputations, were compared, and statistics on the agreement between the methods presented in table 3-13 serve as reliability estimates for the highest level mastered scores. First, highest proficiency level mastered was determined on the basis of item responses alone for all of the children (about 80 percent in each round of data collection) who could be categorized without requiring imputations. Then, for the same group of children, highest level was obtained using only imputations, and ignoring actual item responses. The percent of agreement of the scores resulting from independent application of the two methods can be considered a

proxy for the reliability of the hybrid scoring procedure. The numbers in table 3-13 are the percentages of exact agreement between these two extremes, and the percentages of scores that were within one level of each other. When discrepancies were examined grade-by-grade and level-by-level, there did not appear to be a pattern of either method consistently overestimating or underestimating highest proficiency level when compared with the alternate method. The vast majority of the highest proficiency level scores in the data file were determined on the basis of item responses alone, or item responses supplemented by IRT estimates. The high level of exact-plus-adjacent agreement between the methods indicates that the IRT approach supports the use of the highest level score sufficiently well for use in aggregate statistics.

Tables 3-10 through 3-13 present the reliability statistics for all of the assessment scores.

Table 3-10. Reliability statistics of Item Response Theory (IRT)-based scores, by round of data collection and domain: School years 1998–99, 1999–2000, and 2001–02

Domain	Fall-kindergarten	Spring-kindergarten	Fall-first grade	Spring-first grade	Spring-third grade
Reading	.93	.95	.96	.96	.94
Mathematics	.92	.93	.94	.94	.95
Science	†	†	†	†	.88

† Not applicable. There was no science assessment prior to third grade.
NOTE: The IRT-based scores consist of the IRT scale scores, T-scores, and proficiency probabilities. See sections 3.1.2 and 3.1.4 for a discussion of these scores. Approximately 89 percent of the children interviewed were in third grade during the 2001–02 school year, 9 percent were in second grade, and less than 1 percent were in fourth grade.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first grade, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

Table 3-11. Reliability statistics of routing test number correct (alpha coefficient), by round of data collection and domain: School years 1998–99, 1999–2000, and 2001–02

Domain	Fall-kindergarten	Spring-kindergarten	Fall-first grade	Spring-first grade	Spring-third grade
Reading	.86	.88	.88	.86	.75
Mathematics	.78	.81	.83	.80	.86
Science	†	†	†	†	.75

† Not applicable. There was no science assessment prior to third grade.
NOTE: Approximately 89 percent of the children interviewed were in third grade during the 2001–02 school year, 9 percent were in second grade, and less than 1 percent were in fourth grade. See section 3.1 for a discussion of the routing tests and section 3.1.1 for a discussion of number-right scores.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first grade, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

Table 3-12. Split-half reliability statistics for item-cluster-based scores, by round of data collection and cluster: School years 1998–99, 1999–2000, and 2001–02

Type	Fall-kindergarten	Spring-kindergarten	Fall-first grade	Spring-first grade	Spring-third grade
Print Familiarity	.70	.68	.68	.60	†
Decoding Score	†	†	†	†	.67
Life Science	†	†	†	†	.59
Physical Science	†	†	†	†	.49
Earth Science	†	†	†	†	.46

† Not applicable. Cluster not collected.
NOTE: Approximately 89 percent of the children interviewed were in third grade during the 2001–02 school year, 9 percent were in second grade, and less than 1 percent were in fourth grade.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first grade, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

Table 3-13. Percent agreement of highest proficiency level-mastered score, by round of data collection: School years 1998–99, 1999–2000, and 2001–02.

Category	Fall-kindergarten	Spring-kindergarten	Fall-first grade	Spring-first grade	Spring-third grade
Reading					
Exact Agreement	68	57	57	59	54
Exact + Off by 1	97	95	95	95	94
Mathematics					
Exact Agreement	58	56	56	61	61
Exact + Off by 1	97	96	96	98	98

NOTE: Approximately 89 percent of the children interviewed were in third grade during the 2001–02 school year, 9 percent were in second grade, and less than 1 percent were in fourth grade.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first grade, and third grade data collections, school years 1998-99, 1999-2000, and 2001-02.

3.1.8 Validity

Evidence for the validity of the direct cognitive assessments was derived from several sources. A review of national and state performance standards, comparison with state and commercial assessments, the judgments of curriculum experts and teachers all provided input to test specifications. In addition, comparing the reading and mathematics field-test item pool scores with those obtained from an established instrument provided validity information.

The ECLS-K test specifications were derived from a variety of sources. For the third through fifth grade assessments, national and state performance standards in each of the domains were examined. The scope and sequence of materials from state assessments, as well as from major publishers, were also considered. The resulting ECLS-K fourth grade frameworks are similar to the NAEP fourth grade frameworks, with some differences due to ECLS-K formatting and administration constraints. The fourth grade frameworks were modified for third grade (and for the earlier K-1 forms). An expert panel of early elementary school educators, including curriculum specialists in the subject areas and teachers at the targeted grade levels from different regions of the country, examined the pool of items and the recommended allocations. The assessment specifications indicated target percentages for content strands within each of the subject areas. These percentages were matched as closely as possible in developing the field-test assessment item pool as well as in selecting items for the third-grade assessment forms. Some compromises in matching target percentages were necessary to satisfy constraints related to other issues, including linking to K-1 scales, avoiding floor and ceiling effects, and field-test item performance. This was especially true for the reading assessment, whose structure, i.e., several questions based on each reading passage, placed an additional constraint on the selection of items to match content strands. Experts in each of the subject areas then reviewed the proposed third-grade forms for appropriateness of content and relevance to the assessment framework.

An additional method of evaluating the construct validity of the reading and mathematics assessments was addressed by the inclusion of the Woodcock-McGrew-Werder Mini-Battery of Achievement (MBA; Woodcock, McGrew, and Werder, 1994) in the spring 2000 field test of third grade items. Selected field-test forms that included reading sections also included the MBA reading test, while the MBA mathematics test was administered along with field-test mathematics forms. Correlations were computed for the MBA scores with the theta estimates based on ECLS-K field-test responses. Test scores can be related to other measures only to the extent that they are consistent within themselves. Generally, a correlation between two variables cannot exceed the square root of the reliability of either variable. Reliabilities for the MBA were computed both with not-administered and omitted items treated as missing, and with these items treated as incorrect. The correlations of MBA with ECLS-K measures were quite close to the square roots of the reliabilities, indicating that the two assessments were measuring closely related skills. The correlations are presented in table 3-14.

Table 3-14. Validity coefficients for reading and mathematics field test item pools: School year 2001–02

Category	Reading	Mathematics
Reliability of MBA (computed both ways)	.84 and .86	.81 and .82
Square root of reliability	.92 and .93	.90 and .91
Correlation of MBA x ECLS grade 3 field assessment item pool	.83	.84

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

3.2 Indirect Cognitive Assessment

The Academic Rating Scale (ARS) was developed for the ECLS-K to measure teachers' evaluations of students' academic achievement in four domains: language and literacy (reading and writing), science, social studies, and mathematical thinking. Teachers rated the child's skills, knowledge, and behaviors on a scale from "Not Yet" to "Proficient" (see table 3-15). If a skill, knowledge, or behavior had not been introduced into the classroom yet, the teacher coded that item as N/A (not applicable). In third grade, the classroom teacher most knowledgeable of the child's academic achievement in the four domains may not be the primary or homeroom teacher. The primary teacher was asked to forward the rating form to the teacher most knowledgeable of the particular domain to complete the ratings. The differences between the direct and indirect cognitive assessments, and the scores available, are described here. For a discussion of the content areas of the ARS, see chapter 2, section 2.3.1.

Table 3-15. Academic Rating Scale response scale: School year 2001–02

Value	Response	Description
1	Not yet:	Child has not yet demonstrated skill, knowledge, or behavior.
2	Beginning:	Child is just beginning to demonstrate skill, knowledge, or behavior but does so very inconsistently.
3	In progress:	Child demonstrates skill, knowledge, or behavior with some regularity but varies in level of competence.
4	Intermediate:	Child demonstrates skill, knowledge, or behavior with increasing regularity and average competence but is not completely proficient.
5	Proficient:	Child demonstrates skill, knowledge, or behavior competently and consistently.
N/A	Not applicable:	Skill, knowledge, or behavior has not been introduced in classroom setting.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

3.2.1 Comparison to Direct Cognitive Assessment

The ARS was designed both to overlap and to augment the information gathered through the direct cognitive assessment battery. Although the direct and indirect instruments measure children's skills and behaviors within the same broad curricular domains with some intended overlap, several of the constructs they were designed to measure differ in significant ways. Most importantly, the ARS includes items designed to measure both the process and products of children's learning in school, whereas the direct cognitive battery measures only the products of children's achievement. Because of time and space limitations, the direct cognitive battery is less able to measure the process of children's thinking, including the strategies they use to read, solve math problems, or investigate a scientific phenomenon.

Due to time constraints, the direct cognitive battery does not include a scale of children's knowledge in social studies. On the ARS teachers reported the children's knowledge and understanding of civics, geography, history, culture, and economics.

The criterion-referenced indirect measures on the ARS are targeted to the specific grade level of the student and draw upon the daily observations made by teachers of the students in their class.

3.2.2 Rasch Scores Available for the Academic Rating Scale

A Rasch analysis was used to create measures of the reported performance of students on a hierarchy of skills, knowledge, and behavior. The Rasch Rating Scale model uses the pattern of ratings on items to determine an estimate of the difficulty of each item and to place each student on an interval scale set with a minimum score of one and a maximum score of five. The Rasch analysis showed that the reliability of the estimates of child ability was very high for all domains of the ARS (see table 3-16).

Table 3-16. Person separation reliability statistics[1] for the third-grade Rasch-based score, by category: School year 2001–02

Category	Grade 3
ARS Language and Literacy	.95
ARS Mathematical Thinking	.94
ARS Science	.95
ARS Social Studies	.93

[1]Person separation reliability is a measure of internal consistency and is analogous to the KR-20 and Cronbach's alpha. Person separation is the ratio of the adjusted standard deviation to the root mean standard error: ((S. D. of Measure)^2 – (RMSE)^2) / RMSE. Person separation reliability is the square of this separation statistic divided by one plus the separation squared (Linacre and Wright, 2000). SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

As mentioned, the ARS scores are scaled to have a low value of one and a high value of five to correspond to the 5-point rating scale that teachers used in rating children on these items. The item difficulties and student scores are placed on a common scale. Students have a high probability of receiving a high rating on items whose difficulty is below their scale score, and a lower probability of receiving a high rating on items above their scale score. Therefore, the scores children receive on the ARS subscales should not be interpreted as mean scores, but as the child's relative probability of success with the items. Students who received maximum ratings on all the items or minimum ratings on all the items are assigned an estimated score.

The variable names, descriptions, value ranges, weighted means, and standard deviations for the third grade (T5) ARS scores are shown in table 3-17. The description for each variable in the tables begins with a "T," indicating that it is a teacher questionnaire child-level variable. The items and the metric for the third grade ARS are different from the ARS ratings in earlier rounds of data collection, so the scores are not directly comparable to those for kindergarten and first grade. The students' scores are calculated relative to the item difficulty. With different items used across the grades and separate calibrations performed, the size of the metric differs from one grade to another.

On the ARS, teachers indicated "not applicable" when the knowledge, skill, or behavior had not been introduced to the classroom. Because some children might have already had this skill (from home or other opportunities for learning), the "not applicable" ratings were treated as missing data and the child's score was estimated based on the items on which the child was rated. Although the Rasch

program estimates scores for all children based on the information provided, the file includes only the scores of children who had more than 60 percent of the items in a scale rated. In other words, if 40 percent or more of the items in a scale were not rated, then the score was set to missing. Fewer than 1 percent of literacy and mathematics scores, and fewer than 5 percent of science and social studies scores, failed to meet the completeness criterion.

Table 3-17. Third grade Academic Rating Scale: variable names, descriptions, ranges, weighted means, and standard deviations: School year 2001–02

Variable name	Description	Range of values	Weighted mean	Standard deviation
T5ARSLIT	T5 Literacy ARS Score	1 - 5	3.27	0.89
T5ARSMAT	T5 Math ARS Score	1 - 5	3.08	0.75
T5ARSSCI	T5 Science ARS Score	1 - 5	3.16	0.94
T5ARSSOC	T5 Social Studies ARS Score	1 - 5	3.01	0.86

NOTE: Table estimates based on C5CW0 weight. See chapter 7, section 7.3 for variable naming conventions.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Tables 3-18 to 3-21 provide the estimates of difficulty for each of the items. Higher values mean that teachers rated fewer students as proficient on those items. Students would have a greater than 50 percent probability of receiving ratings of "5" on items below their ability level. Tables are provided for third grade items.

Table 3-18. Spring-third grade Academic Rating Scale language and literacy item difficulties: School year 2001–02

Item difficulty	Item number and abbreviated content
2.69	Q3. Conveys ideas clearly when speaking
2.72	Q6. Reads third grade books (fiction) independently with comprehension
2.74	Q5. Reads fluently
2.87	Q4. Uses various strategies to gain information
3.03	Q7. Reads and comprehends expository text
3.19	Q10. Makes mechanical corrections when reviewing a rough draft
3.22	Q8. Composes multi-paragraph stories/reports,
3.28	Q9. Rereads and reflects on writing, making changes to clarify or elaborate

NOTE: Items are arranged in order of difficulty. Higher values mean that teachers rated fewer students as proficient on those items. Students would have a greater than 50 percent probability of receiving ratings of "5" on items below their ability level.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 3-19. Spring-third grade Academic Rating Scale mathematical thinking item difficulties: School year 2001–02

Item difficulty	Item number and abbreviated content
2.45	Q7. Shows understanding of place value with whole numbers
2.53	Q3. Creates and extends patterns
2.71	Q5. Recognizes properties of shapes and relationships among shapes
2.73	Q9. Surveys, collects, and organizes data into simple graphs
2.78	Q4. Uses a variety of strategies to solve math problems
2.81	Q6. Uses measuring tools accurately
2.83	Q8. Makes reasonable estimates of quantities and checks answers
3.16	Q10. Models, reads, writes, and compares fractions
3.69	Q11. Divides a 3 digit number by a 1 digit number

NOTE: Items are arranged in order of difficulty. Higher values mean that teachers rated fewer students as proficient on those items. Students would have a greater than 50 percent probability of receiving ratings of "5" on items below their ability level.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 3-20. Spring-third grade Academic Rating Scale science item difficulties: School year 2001–02

Item difficulty	Item number and abbreviated content
2.77	Q5. Classifies and compares living and non-living things in different ways
2.84	Q3. Makes logical predictions when conducting scientific investigations
2.87	Q8. Demonstrates understanding of life science concepts
2.93	Q9. Demonstrates understanding of earth and space science concepts
2.97	Q6. Forms explanations and conclusions based on observation and investigation
3.01	Q7. Demonstrates understanding of physical science concepts
3.07	Q4. Communicates scientific information

NOTE: Items are arranged in order of difficulty. Higher values mean that teachers rated fewer students as proficient on those items. Students would have a greater than 50 percent probability of receiving ratings of "5" on items below their ability level.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 3-21. Spring-third grade Academic Rating Scale social studies item difficulties: School year 2001–02

Item difficulty	Item number and abbreviated content
2.56	Q7. Knows how to use maps and globes to locate and derive information
2.67	Q3. Identifies similarities and differences in group habits and living patterns
2.84	Q6. Recognizes the reciprocal influence of environment on people
2.88	Q5. Demonstrates understanding of the ways in which the past influences the present
3.19	Q4. Shows understanding of the purpose and structure of government functions
3.33	Q8. Demonstrates understanding of the U. S. economic system

NOTE: Items are arranged in order of difficulty. Higher values mean that teachers rated fewer students as proficient on those items. Students would have a greater than 50 percent probability of receiving ratings of "5" on items below their ability level.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

The ARS scale was designed to provide information on children's abilities at a given point in time, not necessarily over time. In addition, although some item stems are similar to those used in the kindergarten and first grade teacher questionnaires, the actual items include performance criteria that increase in difficulty from one time to the next. Moreover, the ARS scores are placed on different metrics relative to the item difficulty in a given grade. Therefore, change scores should not be calculated between time points. However, covariance models may be used to compare teacher's ratings of performance in different grades. Before using these variables in such analyses, the distribution of the samples should be assessed to determine if the assumption of normal distribution is met.

Tables 3-22 to 3-25 provide standard errors (SE) for each of the Rasch scores for third grade. The "Score" column is the sum of the raw score ratings. "Measure" is the Rasch-based score. The column labeled "SE" is the corresponding standard error of measurement for those scores. These standard errors can be used in analytic models to correct for the heteroskedasticity of scores.

Table 3-22. Spring-third grade Academic Rating Scale language and literacy standard errors: School year 2001–02

Score	Measure	Standard error (SE)	Score	Measure	Standard Error (SE)	Score	Measure	Standard error (SE)
8	1.00E	.45	19	2.45	.15	30	3.56	.15
9	1.32	.26	20	2.54	.15	31	3.66	.15
10	1.53	.20	21	2.64	.15	32	3.75	.15
11	1.68	.18	22	2.74	.16	33	3.85	.15
12	1.80	.16	23	2.84	.16	34	3.95	.16
13	1.90	.16	24	2.94	.16	35	4.06	.16
14	2.00	.15	25	3.05	.16	36	4.17	.17
15	2.09	.15	26	3.15	.16	37	4.30	.18
16	2.18	.15	27	3.26	.16	38	4.46	.21
17	2.27	.15	28	3.36	.16	39	4.68	.27
18	2.36	.15	29	3.46	.16	40	5.00E	.45

NOTE: The "Score" column is the sum of the raw score ratings. "Measure" is the Rasch-based score. The column labeled "Standard error (SE)" is the corresponding standard error of measurement for those scores. E=estimated extreme score.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 3-23. Spring-third grade Academic Rating Scale mathematical thinking standard errors: School year 2001–02

Score	Measure	Standard error (SE)	Score	Measure	Standard error (SE)	Score	Measure	Standard error (SE)
9	1.00E	.49	22	2.46	.14	35	3.41	.14
10	1.34	.28	23	2.53	.14	36	3.49	.15
11	1.55	.21	24	2.60	.14	37	3.57	.15
12	1.69	.18	25	2.67	.14	38	3.66	.15
13	1.79	.16	26	2.74	.14	39	3.75	.16
14	1.88	.15	27	2.82	.14	40	3.85	.17
15	1.96	.14	28	2.89	.14	41	3.97	.18
16	2.04	.14	29	2.96	.14	42	4.11	.20
17	2.11	.14	30	3.03	.14	43	4.29	.24
18	2.18	.13	31	3.11	.14	44	4.58	.32
19	2.25	.13	32	3.18	.14	45	5.00E	.52
20	2.32	.13	33	3.26	.14			
21	2.39	.13	34	3.33	.14			

NOTE: The "Score" column is the sum of the raw score ratings. "Measure" is the Rasch-based score. The column labeled "Standard error (SE)" is the corresponding standard error of measurement for those scores. E=estimated extreme score.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 3-24. Spring-third grade Academic Rating Scale science standard errors: School year 2001–02

Score	Measure	Standard error (SE)	Score	Measure	Standard error (SE)	Score	Measure	Standard error (SE)
7	1.00E	.40	17	2.36	.15	27	3.63	.19
8	1.29	.23	18	2.47	.15	28	3.80	.20
9	1.48	.16	19	2.58	.16	29	3.97	.19
10	1.61	.15	20	2.71	.17	30	4.12	.17
11	1.73	.15	21	2.85	.18	31	4.25	.16
12	1.83	.15	22	2.99	.17	32	4.37	.17
13	1.94	.15	23	3.12	.16	33	4.51	.18
14	2.05	.15	24	3.24	.16	34	4.71	.24
15	2.15	.15	25	3.36	.16	35	4.99E	.40
16	2.26	.15	26	3.48	.17			

NOTE: The "Score" column is the sum of the raw score ratings. "Measure" is the Rasch-based score. The column labeled "Standard error (SE)" is the corresponding standard error of measurement for those scores. E=estimated extreme score.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 3-25. Spring-third grade Academic Rating Scale social studies standard errors: School year 2001–02

Score	Measure	Standard error (SE)	Score	Measure	Standard error (SE)	Score	Measure	Standard error (SE)
6	1.00E	.45	15	2.47	.17	24	3.71	.19
7	1.33	.27	16	2.59	.17	25	3.86	.19
8	1.56	.21	17	2.72	.17	26	4.02	.20
9	1.73	.19	18	2.84	.17	27	4.19	.21
10	1.87	.18	19	2.97	.18	28	4.39	.23
11	2.00	.17	20	3.11	.18	29	4.65	.28
12	2.12	.17	21	3.25	.19	30	5.00E	.46
13	2.23	.17	22	3.40	.19			
14	2.35	.17	23	3.55	.19			

NOTE: The "Score" column is the sum of the raw score ratings. "Measure" is the Rasch-based score. The column labeled "Standard error (SE)" is the corresponding standard error of measurement for those scores. E=estimated extreme score.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Classroom teachers were asked to forward the ARS to the child's teacher most knowledgeable about the child's performance in each of the subject areas to complete. The majority of teachers rated more than one student on the ARS. The number of students rated by each teacher ranged from one to more than 20. The teacher ratings do not represent a systematic national sample of teachers. Each set of teacher ratings is linked to a sampled child, and teachers were asked to rate as many ECLS-K sample children as they had in class.

3.3 Teacher Social Rating Scale

The teacher Social Rating Scale (SRS) asked third-grade teachers to report how often students exhibited certain social skills and behaviors. Teachers rated individual students as part of a self-administered questionnaire. (In earlier rounds of data collection, SRS questions had been asked of both teachers and parents.) Teachers used a frequency scale (see table 3-26) to report on how often the student demonstrated the behavior described. See chapter 2, sections 2.3 and 2.3.2 for additional information on the teacher SRS instrument.

Table 3-26. Social Rating Scale response scale: School year 2001–02

Value	Response	Description
1	Never	Student never exhibits this behavior.
2	Sometimes	Student exhibits this behavior occasionally or sometimes.
3	Often	Student exhibits this behavior regularly but not all the time.
4	Very often	Student exhibits this behavior most of the time.
N/O.	No opportunity	No opportunity to observe this behavior.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Five teacher SRS scales were developed based on responses to the scale. The scale scores on all SRS scales are the mean rating on the items included in the scale. Scores were computed only if the student was rated on at least two-thirds of the items in that scale. The five social skill teacher scales are as follows: approaches to learning, self-control, interpersonal skills, externalizing problem behaviors, and internalizing problem behaviors. Although 24 of the 26 third grade SRS items were the same as items in the K-1 instrument, teachers may place different interpretations on the meaning of the items at different time points. Therefore these scores would be most appropriately used as covariates rather than as change scores.

Two items were added to the third grade scales due to a high number of maximum scores on the field assessment of these items. One item was added to the externalizing problem behavior scale ("child talks during quiet study time"). The other additional item "child follows classroom rules" was added to the SRS in an attempt to increase variance in the self-control scale. Analysis of the item responses indicated that it contributed strongly to the approaches to learning scale, increasing the variance and reliability of that scale. Thus, this item is included in the approaches to learning scale.

In third grade, examination of the responses suggested a different perception of a student's self-control and interpersonal social abilities. The self-control scale includes items on control of attention as well as control of emotions and behavior in interactions. Third-grade students who were rated higher on self-control were also rated higher on interpersonal skills that involved peers. Thus, in addition to the self-control and interpersonal social abilities scale scores, a peer relations scale score was included. This additional scale combines responses on both the interpersonal and self-control scale items that relate to peers.

Variable names for the teacher scores, descriptions, ranges, weighted means, and standard deviations for these scales are shown in table 3-27. About 90 percent of the children whose teachers provided social ratings data were in third grade during the round 5 data collection, and about 9 percent were in first or second grade. Numbers in the table are for third graders, with scores for children who at round 5 were still in first or second grade shown in parentheses. The number of children who had advanced to fourth or fifth grade by round 5 was too small to be analyzed separately.

Table 3-27. Teacher Social Rating Scale scores: variable names, descriptions, ranges, weighted means, and standard deviations: School year 2001–02

Variable name	Description	Range of values	Weighted mean	Standard deviation
T5LEARN	T5 Approaches to Learning	1 - 4	3.1 (2.7)	0.7 (0.7)
T5CONTRO	T5 Self-Control	1 - 4	3.2 (3.0)	0.6 (0.7)
T5INTERP	T5 Interpersonal	1 - 4	3.1 (2.8)	0.6 (0.7)
T5EXTERN	T5 Externalizing Problem Behaviors	1 - 4	1.7 (1.9)	0.6 (0.7)
T5INTERN	T5 Internalizing Problem Behaviors	1 - 4	1.6 (1.8)	0.5 (0.6)
T5SCINT	T5 Combo of Self-Control & Interpersonal (Peer Relations)	1 - 4	3.2 (2.9)	0.6 (0.6)

NOTE: Table estimates based on C5CW0 weight. Numbers outside of parentheses represent children in third grade at the time of assessment. Numbers within parentheses represent first and/or second graders at the time of assessment. See chapter 7, section 7.3 for variable naming conventions.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

The split-half reliabilities for the teacher SRS scales are high (see table 3-28). Reliabilities are nearly identical for third graders in round 5 and for children who were not yet in third grade, so the table contains only reliabilities for the whole sample.

Table 3-28. Split-half reliability for the teacher Social Rating Scale scores: School year 2001–02

Variable name	Description	Split-half reliability
T5LEARN	T5 Approaches to Learning	.91
T5CONTRO	T5 Self-control	.79
T5INTERP	T5 Interpersonal	.89
T5EXTERN	T5 Externalizing Problem Behaviors	.89
T5INTERN	T5 Internalizing Problem Behaviors	.76
T5SCINT	T5 Combo of Self-Control & Interpersonal (Peer Relations)	.92

NOTE: See chapter 7, section 7.3 for variable naming conventions.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Care should be taken when entering these scales into the same analysis due to problems of multicollinearity. The intercorrelations among the five SRS factors (excluding the combined peer relations scale) are high. The factor intercorrelations with the internalizing problem behaviors are the lowest. The absolute values of correlations among the teacher SRS factors range from 0.32 to 0.81, with nearly identical patterns for third graders and for children who were still in first or second grade.

3.4 Self-Description Questionnaire

For the first time in the ECLS-K, third grade students rated their perceived competence and interest in reading, mathematics, and all school subjects. They also rated their perceived competence and popularity with peers and reported on problem behaviors with which they might struggle. The "Externalizing Problems" scale included questions about anger and distractability, while "Internalizing Problems" scale included items on sadness, loneliness, and anxiety. For further description of the Self-Description Questionnaire (SDQ) see chapter 2. Students rated whether each item was "not at all true," "a little bit true," "mostly true," or "very true." Five scales were produced from the SDQ items. The scale scores on all SDQ scales represent the mean rating of the items included in the scale. Students who responded to the SDQ answered virtually of the questions, so treatment of missing data was not an issue. As with most measures of social-emotional behaviors, the distributions on these scales are skewed (negatively skewed for the positive social behavior scales, and positively skewed for the problem behavior scales). The reliability is lower for scales with only six items (see table 3-29).

Table 3-29. Self-Description Questionnaire scale reliabilities (alpha coefficient): School year 2001–02

Variable name	Description	Number of items	Alpha coefficient
C5SDQRDC	C5 SDQ Prcvd Interest/Competence - Reading	8	.87
C5SDQMTC	C5 SDQ Prcvd Interest/Competence - Math	8	.90
C5SDQSBC	C5 SDQ Prcvd Interest/Competence - All Sbj	6	.79
C5SDQPRC	C5 SDQ Prcvd Interest/Competence - Peer Rl	6	.79
C5SDQEXT	C5 SDQ Externalizing Problems	6	.77
C5SDQINT	C5 SDQ Internalizing Problems	8	.81

NOTE: See chapter 7, section 7.3 for variable naming conventions.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 3-30. Self-Description Questionnaire scale range, mean, and standard deviation (weighted): School year 2001–02

Variable name	Description	Range of Values	Weighted mean	Standard Deviation
C5SDQRDC	C5 SDQ Prcvd Interest/Competence - Reading	1 - 4	3.26	0.66
C5SDQMTC	C5 SDQ Prcvd Interest/Competence - Math	1 - 4	3.16	0.79
C5SDQSBC	C5 SDQ Prcvd Interest/Competence - All Sbj	1 - 4	2.92	0.66
C5SDQPRC	C5 SDQ Prcvd Interest/Competence - Peer Rl	1 - 4	3.03	0.65
C5SDQEXT	C5 SDQ Externalizing Problems	1 - 4	2.02	0.71
C5SDQINT	C5 SDQ Internalizing Problems	1 - 4	2.22	0.74

NOTE: Table estimates based on C5CW0 weight. See chapter 7, section 7.3 for variable naming conventions.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

REFERENCES

Linacre, J., and Wright, B. (2000). *WINSTEPS.* Chicago, IL: MESA Press.

Rock, D.A., and Pollack, J.M. (2002). *A Model-based Approach to Measuring Cognitive Growth in Pre-reading and Reading Skills During the Kindergarten* (ETS Research Report RR-02-18). Princeton, NJ: Educational Testing Service.

U.S. Department of Education, National Center for Education Statistics (2002). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99* (*ECLS-K) Psychometric Report for Kindergarten Through First Grade* (NCES 2002-005) by Donald A. Rock and Judith M. Pollack, Educational Testing Service, Elvira Germino Hausken, project officer. Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics (forthcoming). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99* (*ECLS-K) Third Grade Methodology Report*. Washington, DC: U.S. Department of Education.

Woodcock, R.W., McGrew, K.S., and Werder, J.K. (1994). *Woodcock-McGrew-Werder Mini-Battery of Achievement*, Itasca, IL: Riverside Publishing.

4. SAMPLE DESIGN AND IMPLEMENTATION

The Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) employed a multistage probability sample design to select a nationally representative sample of children attending kindergarten in 1998–99. In the base year the primary sampling units (PSUs) were geographic areas consisting of counties or groups of counties. The second-stage units were schools within sampled PSUs. The third and final stage units were students within schools.

The first grade data collection targeted base year respondents, where a case was considered responding if there was a completed child assessment or parent interview in fall- or spring-kindergarten. While all base-year respondents were eligible for the spring-first grade data collection, the effort for fall-first grade was limited to a 30 percent subsample. The spring student sample was freshened to include current first graders who had not been enrolled in kindergarten in 1998–99 and, therefore, had no chance of being included in the ECLS-K base year kindergarten sample. For both fall- and spring-first grade, only a subsample of students who had transferred from their kindergarten school was followed.

The third grade data collection targeted base year respondents and children sampled in first grade through the freshening operation. As in the first grade data collection, only a subsample of students who had transferred from their kindergarten school was followed. In third grade, however, the subsampling rate applied to transferred children was slightly higher: children whose home language is non-English (also known as children belonging to the language minority group) and who moved for the first time in third grade were followed at 100 percent. In other words, children belonging to the language minority group who did not move in first grade but moved in third grade were all followed into their new third grade schools. The higher subsampling rate allows for the preservation of this group in the sample for analytic reasons. Children not in the language minority group continued to be subsampled for followup if they moved in third grade.

4.1 Base Year Sample

In the base year, children were selected for the ECLS-K using a multistage probability design. The PSUs were counties or groups of counties selected with probability proportional to size (PPS). The basic PSU measure of size was the number of 5-year-olds, but this was modified to facilitate

the oversampling of Asian and Pacific Islanders (APIs) required to meet precision goals. In all, there were 100 PSUs selected for the ECLS-K. The 24 PSUs with the largest measure of size were designated self-representing (SR) and were included in the sample with certainty. The remaining non-SR PSUs were partitioned into 38 strata of roughly equal size. An initial cross-classification of census region with metropolitan statistical area (MSA) status created eight superstrata. These were further subdivided by percent minority, PSU measure of size (a composite count of five-year-old children), and 1988 per capita income. From each non-SR stratum, two PSUs were selected PPS without replacement using Durbin's Method (Durbin, 1967).

Table 4-1 summarizes the characteristics of the ECLS-K PSU sample.

Table 4-1. Distribution of the ECLS-K primary sampling unit (PSU) sample by self-representing (SR) status, metropolitan statistical area (MSA) status, and census region: School year 1998–99

			Census region			
SR status	MSA status	Total	Northeast	Midwest	South	West
Total		100	18	25	34	23
SR	MSA	24	6	5	6	7
Non-SR	MSA	52	10	12	18	12
Non-SR	Non-MSA	24	2	8	10	4

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten data collection, school year 1998–99.

In the second stage, public and private schools offering kindergarten programs were selected. For each PSU, a frame of public and private schools offering kindergarten programs was constructed using existing school universe files: the 1995–96 Common Core of Data (CCD; U.S. Department of Education, 1995–96) and the 1995–96 Private School Universe Survey (PSS; U.S. Department of Education, 1998). The 1995–96 Office of Indian Education Programs Education Directory was consulted in order to complete the list of Bureau of Indian Affairs (BIA) schools in the CCD file. For Department of Defense (DOD) domestic schools, a 1996 list of schools was obtained directly from the DOD. A procedure was implemented to identify kindergarten programs that would be operational at the time of ECLS-K's base year data collection, but that were not included in the frame just described. These were newly opened schools that were not listed in the CCD and the PSS, and schools that were in the CCD and the PSS but did not appear to offer kindergarten programs according to those sources. The selection of schools was systematic, with probability proportional to a weighted measure of size based on the number of kindergartners enrolled. As with the PSU sample, the measure of size was constructed

taking into account the desired oversampling of APIs. Public and private schools constituted distinct sampling strata. Within each stratum, schools were sorted to ensure good sample representation across other characteristics. In total, 1,280 schools were sampled from the original frame, and 133 from the freshened frame. Of these, 953 were public schools and 460 were private schools.

The characteristics of the ECLS-K school sample are presented in table 4-2. Schools that were discovered to be ineligible during recruitment have been omitted from the tabulation.

Table 4-2. Number of schools in the ECLS-K base year school sample, by selected school characteristics: School year 1998–99

Characteristic	Total	Sector	
		Public	Private
Total	1,277	914	363
Region			
Northeast	243	161	82
Midwest	298	210	88
South	418	306	112
West	318	237	81
Type of locale			
Large city	245	168	77
Midsize city	248	172	76
Urban fringe of large city	382	265	117
Urban fringe of midsize city	99	78	21
Large town	33	24	9
Small town	112	76	36
Rural	158	131	27
Religious affiliation			
Catholic	120	†	120
Other religious	149	†	149
Nonreligious, private	94	†	94
School type			
Regular	1,162	893	269
Ungraded	4	1	3
No grade beyond K	49	19	30
Unknown	62	1	61

† Not applicable.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten data collection, school year 1998–99.

The third stage sampling units were children of kindergarten age, selected within each sampled school. The goal of the student sample design was to obtain an approximately self-weighting

sample of students and at the same time to achieve a minimum required sample size for APIs who were the only subgroup that needed to be oversampled to meet the study's precision goals. For each sampled school, the field staff obtained a complete list of kindergartners enrolled. Two independent sampling strata were formed within each school, one containing API students and the second, all other students. Within each stratum, students were selected using equal probability systematic sampling, using a higher rate for the API stratum.[1] In general, the target number of children sampled at any one school was 24. Once the sampled children were identified, parent contact information was obtained from the school. The information was used to locate a parent or guardian and gain parental consent for the child assessment and for the parent interview. Table 4-3 presents characteristics of children sampled for the base year.

During the fall-kindergarten data collection, a census of kindergarten teachers was taken at each school. Each sampled child was linked to his or her kindergarten teacher. In spring-kindergarten, teacher-child linkages were reviewed and updated. If new kindergarten teachers had joined the school, they were added to the census of kindergarten teachers. Special education teachers who taught one or more sampled children were included in the spring-kindergarten data collection. If a sampled child received special education services from such a teacher, the teacher was linked to that child.

4.2 Fall-First Grade Subsample

A subsample of ECLS-K PSUs was selected for fall-first grade data collection. All 24 of the SR PSUs were retained. Of the 76 non-self-representing (NSR) PSUs, 38 were retained by sampling one PSU per stratum with equal probability.

Base year schools in the 62 fall-first grade sampled PSUs were stratified by frame source (original public, original private, freshened, etc.) and arranged in their original selection order. A 30 percent equal probability subsample of schools was drawn in the 24 SR PSUs and a 60 percent subsample of schools was drawn in the 38 NSR PSUs. In total 311 schools that had cooperated in either fall- or spring-kindergarten were selected. The characteristics of the base year cooperating schools selected for fall-first grade data collection are presented in table 4-4.

[1] See the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99* (*ECLS-K) Base Year Public-Use Data Files and Electronic Code Book: User's Manual* (NCES 2001–029; U.S. Department of Education, National Center for Education Statistics, 2000) for a detailed description of the base year sample.

Table 4-3. Number (unweighted) of children in the ECLS-K base year student sample, by selected characteristics: School year 1998–99

Characteristic	Total	Sector	
		Public	Private
Total	22,666	17,777	4,889
Region			
Northeast	4,262	3,045	1,217
Midwest	5,628	4,292	1,336
South	7,461	6,179	1,282
West	5,315	4,261	1,054
Type of locale			
Large city	4,550	3,365	1,185
Midsize city	4,728	3,569	1,159
Urban fringe of large city	6,470	4,945	1,525
Urban fringe of midsize city	1,644	1,434	210
Large town	714	577	137
Small town	1,905	1,485	420
Rural	2,655	2,402	253
Religious affiliation			
Catholic	2,510	†	2,510
Other religious	1,445	†	1,445
Nonreligious, private	934	†	934
School type			
Regular	21,436	17,390	4,046
Ungraded	56	24	32
No grade beyond K	663	338	325
Unknown	511	25	486
Composite child race			
White	11,723	8,533	3,190
Black	3,204	2,800	404
Hispanic, with race	1,749	1,455	294
Hispanic, without race	1,983	1,741	242
Asian	1,355	1,102	253
Pacific Islander	220	199	21
Native American	377	334	43
Multirace	511	416	95
Unknown	1,544	1,197	347
Highest parent level of education			
Less than high school	2,027	1,968	59
High school graduate	5,251	4,703	548
Vocational/technical	1,139	964	175
Some college	5,351	4,182	1,169
College graduate	4,004	2,568	1,436
Masters	1,429	850	579
Ph.D./professional	890	456	434
Unknown	2,575	2,086	489

† Not applicable.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten data collection, school year 1998–99.

Table 4-4. Number of base year cooperating schools selected for fall-first grade, by selected school characteristics: School year 1999–2000

Characteristic	Total	Sector	
		Public	Private
Total	311	228	83
Region			
Northeast	57	39	18
Midwest	83	59	24
South	99	77	22
West	72	53	19
Type of locale			
Large city	62	42	20
Midsize city	59	45	14
Urban fringe of large city	86	61	25
Urban fringe of midsize city	18	14	4
Large town	15	12	3
Small town	28	19	9
Rural	43	35	8
Religious affiliation			
Catholic	29	†	29
Other religious	33	†	33
Nonreligious, private	21	†	21
School type			
Regular	292	222	70
Ungraded	1	1	0
No grade beyond K	18	5	13

† Not applicable.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 first grade data collection, school year 1999–2000.

Fall-first grade data collection consisted of the direct child assessment and the parent interview. Data collection was attempted for every eligible child found still attending the school in which he or she had been sampled during kindergarten. “Eligible” is defined as a base year respondent (i.e., a child who had either a fall- or spring-kindergarten child assessment or parent interview). Base year nonrepondents would be adjusted for during weighting. Because of the additional burden of school recruiting, the cost of collecting data for a child who transferred from the school in which he or she was originally sampled greatly exceeds that for a child who stayed enrolled. To contain these costs, a random 50 percent of children were flagged to be followed for fall-first grade data collection in the event that they had transferred.

Except for children who were repeating kindergarten, all base year children sampled in schools with a high grade of kindergarten are de facto movers. Since many of these movers may move *en masse* to the same first grade school, steps were taken to follow these children at a higher rate. Using the information collected during spring-kindergarten, a list of destination schools was compiled for each such school. The destination school having the most movers was designated as primary, unless no such school had more than three movers. Children who moved *en masse* into a primary destination school in fall-first grade were treated as "nonmovers" and were not subsampled.

Prior to subsampling with equal probability, children were stratified into groups of nonmovers, movers with information identifying their new schools, and movers without such identifying information. A flag was created for each child indicating whether the child had been sampled to be followed.

Table 4-5 shows the characteristics of the children subsampled for fall-first grade. Region, locale, religious affiliation, and school type describe the school the child attended in kindergarten.

4.3 Spring-First Grade Sample

The ECLS-K spring-first grade data collection targeted all base year respondents. In addition the spring student sample was freshened to include current first graders who had not been enrolled in kindergarten in 1998–99 and, therefore, had no chance of being included in the ECLS-K base year kindergarten sample. While all students still enrolled in their base year schools were recontacted, only a 50 percent subsample of base year sampled students who had transferred from their kindergarten school was followed for data collection.

4.3.1 Subsampling Movers

In spring-first grade all children in a random 50 percent subsample of base year schools were flagged to be followed for data collection if they transferred from their base year school. (This is in contrast to fall-first grade where a random 50 percent of children in each of the 30 percent of schools

Table 4-5. Number (unweighted) of children subsampled for fall-first grade, by selected characteristics: School year 1999-2000

Characteristic	Total	Sector	
		Public	Private
Total	5,650	4,446	1,204
Region			
Northeast	1,000	759	241
Midwest	1,416	1,068	348
South	1,873	1,557	316
West	1,361	1,062	299
Type of locale			
Large city	1,154	816	338
Midsize city	1,109	874	235
Urban fringe of large city	1,558	1,205	353
Urban fringe of midsize city	320	276	44
Large town	306	246	60
Small town	518	390	128
Rural	685	639	46
Religious affiliation			
Catholic	535	†	535
Other religious	254	†	254
Nonreligious, private	415	†	415
School type			
Regular	5,374	4,338	1,036
Ungraded	24	24	0
No grade beyond K	138	84	54
Unknown	114	0	114
Composite child race			
White	3,131	2,288	843
Black	849	718	131
Hispanic, with race	419	345	74
Hispanic, without race	522	475	47
Asian	305	243	62
Pacific Islander	99	97	2
Native American	137	132	5
Multirace	163	127	36
Unknown	25	21	4

See notes at end of table.

Table 4-5. Number (unweighted) of children subsampled for fall-first grade, by selected characteristics: School year 1999–2000—Continued

Characteristic	Total	Sector	
		Public	Private
Highest parent level of education			
Less than high school	530	9	521
High school graduate	1,252	128	1,124
Vocational/technical	335	50	285
Some college	1,419	300	1,119
College graduate	1,038	358	680
Masters	398	157	241
Ph.D./professional	255	130	125
Unknown	423	72	351

† Not applicable.
NOTE: School characteristics (i.e., region, locale, religious affiliation, and school type) describe the school the child attended in kindergarten.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 first grade data collection, school year 1999–2000.

subsampled were flagged.) In order to maximize the amount of longitudinal data, care was taken during spring-first grade sampling to ensure that any child who had been flagged to be followed in fall-first grade would continue to be so.

In selecting the spring-first grade 50 percent subsample of schools where movers would be flagged for followup, the three primary strata were SR PSUs, NSR PSUs that had been selected for fall-first grade, and NSR PSUs that had not been selected for fall-first grade. Within these major strata, schools were grouped by frame source (original public, original private, new from Catholic dioceses, new from local governments, etc.). Finally within each frame source, schools were stratified by response status, and arranged in original selection order. Schools that had been part of the 30 percent fall-first grade sample were automatically retained. Then equal probability sampling methods were employed to augment the sample to the desired 50 percent. The net result of these procedures was that every base year selected school had a 50 percent chance of having its ECLS-K transfer students followed during spring-first grade, and any transfer student who had been followed in fall-first grade would still be followed in spring-first grade.

Table 4-6 shows the characteristics of the children in the spring-first grade sample, excluding freshened students. Region, locale, religious affiliation, and school type describe the school at which the child attended kindergarten.

Table 4-6. Number (unweighted) of children in spring-first grade sample excluding freshened students, by selected characteristics: School year 1999–2000

Characteristic	Total	Sector	
		Public	Private
Total	18,084	14,248	3,836
Region			
Northeast	3,339	2,434	905
Midwest	4,578	3,474	1,104
South	6,050	5,029	1,021
West	4,117	3,311	806
Type of locale			
Large city	3,459	2,575	884
Midsize city	3,761	2,797	964
Urban fringe of large city	5,140	3,991	1,149
Urban fringe of midsize city	1,288	1,126	162
Large town	576	466	110
Small town	1,578	1,215	363
Rural	2,282	2,078	204
Religious affiliation			
Catholic	2,091	†	2,091
Other religious	1,139	†	1,139
Nonreligious, private	606	†	606
School type			
Regular	17,277	13,971	3,306
Ungraded	40	24	16
No grade beyond K	420	235	185
Unknown	347	18	329
Composite child race			
White	10,208	7,472	2,736
Black	2,597	2,289	308
Hispanic, with race	1,460	1,220	240
Hispanic, without race	1,648	1,456	192
Asian	1,149	939	210
Pacific Islander	202	186	16
Native American	332	294	38
Multirace	434	347	87
Unknown	54	45	9

See notes at end of table.

Table 4-6. Number (unweighted) of children in spring-first grade sample excluding freshened students, by selected characteristics: School year 1999–2000—Continued

Characteristic	Total	Sector	
		Public	Private
Highest parent level of education			
Less than high school	1,529	1,491	38
High school graduate	3,779	3,356	423
Vocational/technical	1,078	926	152
Some college	4,211	3,313	898
College graduate	3,348	2,194	1,154
Masters	1,191	719	472
Ph.D./professional	749	395	354
Unknown	2,199	1,854	345

† Not applicable.
NOTE: School characteristics (i.e., region, locale, religious affiliation, and school type) describe the school the child attended in kindergarten.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 first grade data collection, school year 1999–2000.

4.3.2 Student Freshening

The spring-first grade student freshening used a half-open interval sampling procedure (Kish, 1965). The procedure was implemented in the same 50 percent subsample of ECLS-K base year schools where transfer students were flagged for followup. Each of these schools was asked to prepare an alphabetic roster of students enrolled in first grade and the names of ECLS-K kindergarten-sampled students were identified on this list. Beginning with the name of the first kindergarten-sampled child, school records were checked to see whether the student directly below in the sorted list attended kindergarten in the United States in fall 1998. If not, (1) that child was considered to be part of the freshened sample and was linked to the base year sampled student (i.e., was assigned that student's probability of selection) and (2) the record search procedure was repeated for the next listed child, and so forth. When the record search revealed that a child had been enrolled in kindergarten the previous year, that child was not considered part of the freshened sample and the procedure was begun all over again with the second base year sampled student name, and so on. Note: the student roster was "circularized" (i.e., the first name on the roster was considered to follow the last name on the roster in the implementation of the procedure). Student freshening brought 165 first graders into the ECLS-K sample, which increased the weighted survey estimate of the number of first graders in the United States by about 2.6 percent.

The student freshening procedure was not entirely free of bias. A first grader would have no chance of being in the ECLS-K first grade sample if he or she was enrolled in a school where neither the child nor any of his or her classmates had attended kindergarten in the United States in fall 1998. This would be a rare circumstance and is not thought to be an important source of bias. A more significant source of potential bias is nonresponse. One source of nonresponse inherent to the freshening plan was that the procedure only involved students who had not transferred from the school in which they had been sampled during the base year. A more detailed discussion of freshened student nonresponse can be found in section 5.7.2 of the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) User's Manual for the ECLS-K First Grade Public-Use Data Files and Electronic Code Book* (NCES 2002-135; U.S. Department of Education, National Center for Education Statistics, 2002).

4.4 Spring-Third Grade Sample

The sample of children for spring-third grade consists of all children who were base year respondents and children who were brought into the sample in spring-first grade through the sample freshening procedure described in section 4.3.2. Sample freshening was not implemented in third grade; hence no new students entered the sample.

While all students still enrolled in their base year schools were recontacted, slightly more than 50 percent of the base year sampled students who had transferred from their kindergarten school were followed for data collection. This subsample of students was the same 50 percent subsample of base year movers flagged for following in spring-first grade, with the addition of movers whose home language was not English (language minority students). The two special sampling procedures implemented in spring-third grade are described below.

4.4.1 Subsampling Movers

In spring-first grade all children in a random 50 percent subsample of base year schools were flagged to be followed for data collection if they transferred from their base year school. In order to maximize the amount of longitudinal data, care was taken during spring-first grade sampling to ensure that any child who had been flagged to be followed in fall-first grade would continue to be followed. The spring-first grade sampling procedure for movers is described in section 4.3.1. In spring-third grade,

children who were followed in spring-first grade were retained in the sample (i.e., the mover followup still targeted the same 50 percent subsample of children in the base year schools). In addition, language minority children who moved between first and third grade were followed with certainty as described below.

4.4.2 Language Minority Children

In addition to the subsample of movers to be followed described above, children whose home language was not English and who moved between spring-first grade and spring-third grade were all retained rather than being subsampled at the 50 percent rate. Operationally, this means that children whose home language was not English who were not flagged for followup in the previous round had their flags switched. This only affects children who had not moved out of the original sample schools before third grade. If they had moved before third grade, than their flags were not switched and they continued not to be followed. This modification to the mover followup procedure provides a larger sample of children whose home language is not English for analytic purposes. The mover followup activities that originally targeted a 50 percent subsample of children in base year schools resulted in a 54 percent subsample with the addition of language minority children.

Table 4-7 shows the characteristics of children in the spring-third grade sample, excluding freshened students. Region, locale, religious affiliations, and school type describe the school at which the child attended kindergarten.

4.5 Sample Attrition

In a longitudinal study, sample attrition due to nonresponse and change in eligibility status is expected. The sample of respondents decreases with each round of data collection. In the case of the ECLS-K, a combination of field and sampling procedures was applied that caused the sample to increase after the fall-kindergarten data collection, but then decrease in spring-first grade and again in spring-third grade.

Table 4-7. Number (unweighted) of children in spring-third grade sample excluding freshened students, by selected characteristics: School year 2001–02

Characteristic	Total	Sector	
		Public	Private
Total	16,670	13,166	3,504
Region			
Northeast	3,102	2,274	828
Midwest	4,208	3,187	1,021
South	5,522	4,607	915
West	3,838	3,098	740
Type of locale			
Large city	3,150	2,344	806
Midsize city	3,385	2,536	849
Urban fringe of large city	4,747	3,705	1,042
Urban fringe of midsize city	1,194	1,033	161
Large town	536	428	108
Small town	1,491	1,149	342
Rural	2,167	1,971	196
Religious affiliation			
Catholic	1,924	†	1,924
Other religious	1,036	†	1,036
Nonreligious, private	544	†	544
School type			
Regular	15,930	12,901	3,029
Ungraded	34	23	11
No grade beyond K	391	222	169
Unknown	315	20	295
Composite child race			
White	9,348	6,853	2,495
Black	2,238	1,977	261
Hispanic, with race	1,450	1,222	228
Hispanic, without race	1,547	1,367	180
Asian	1,115	911	204
Pacific Islander	196	180	16
Native American	305	273	32
Multirace	432	351	81
Unknown	39	32	7

See notes at end of table.

Table 4-7. Number (unweighted) of children in spring-third grade sample excluding freshened students, by selected characteristics: School year 2001–02—Continued

Characteristic	Total	Sector	
		Public	Private
Highest parent level of education			
Less than high school	1,586	1,543	43
High school graduate	3,536	3,196	340
Vocational/technical	935	801	134
Some college	4,500	3,621	879
College graduate	3,517	2,352	1,165
Masters	1,324	825	499
Ph.D./professional	813	429	384
Unknown	459	399	60
Home language			
Not English	4,409	3,676	733
English	12,261	9,490	2,771

† Not applicable.
NOTE: School characteristics (i.e., region, locale, religious affiliation, and school type) describe the school the child attended in kindergarten.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

The first procedure was refusal conversion in spring-kindergarten, resulting in a number of schools that agreed to participate in the study after having refused to do so in the previous round. From these schools, 1,426 children were sampled and added to the initial sample. The second procedure was sample freshening in spring-first grade as described in section 4.3.2. This brought in 165 eligible children to add to the sample of 21,192 base year respondents who remained eligible after the base year. A base year responding child was defined as one with at least one direct cognitive test score in fall- or spring-kindergarten or whose parent responded to the family structure section of the parent instrument in fall- or spring-kindergarten. The third and last procedure, applied in first and third grades, required that a subsample of children who moved out of their original sample schools not be followed into their new schools, as described in sections 4.3.1 and 4.4.1, resulting in a decrease in the sample.

Table 4-8 shows the sample size for each round of data collection of the ECLS-K, and the response status of the children in each round. Fall-first grade is not included in this table, as it pertains only to a subsample of the ECLS-K children. Tables 4-9 and 4-10 show the same children separately by the original sample school type (public/private).

Table 4-8. Number (unweighted) of children in the ECLS-K sample, by response status and data collection round: School years 1998–99, 1999–2000, and 2001–02

Data collection round	Unweighted sample size	Response status				
		Ineligibles	Unknown eligibility	Non-followed movers	Nonrespondents	Respondents
Fall-kindergarten	21,387	31	†	†	1,672	19,684
Spring-kindergarten	22,813[1]	147	†	†	2,088	20,578
Spring-first grade	21,357[2]	56	202	2,850	925	17,324
Spring-third grade	21,357	122	289	4,117	1,524	15,305

† Not applicable.
[1] 1,426 children were sampled from refusal-converted schools.
[2] 21,192 children remained eligible after the base year. In addition, 165 children were sampled via the sample freshening procedure.
NOTE: Response status is defined in terms of completed child assessment OR completed family structure data of the parent interview. Children who died or moved out of the country are classified as ineligible. Children who moved and were subsampled for followup but could not be located were treated as belonging to the unknown eligibility category. A portion of children who moved was subsampled out and not followed into their new schools.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

Table 4-9. Number (unweighted) of public school children in the ECLS-K sample, by response status and data collection round: School years 1998–99, 1999–2000, and 2001–02

Data collection round	Unweighted sample size	Response status				
		Ineligibles	Unknown eligibility	Non-followed movers	Nonrespondents	Respondents
Fall-kindergarten	17,003	23	†	†	1,324	15,656
Spring-kindergarten	17,894[1]	117	†	†	1,676	16,101
Spring-first grade	16,784[2]	45	181	2,164	733	13,661
Spring-third grade	16,784	99	250	3,129	1,236	12,070

† Not applicable.
[1] 891 public school children were sampled from refusal-converted schools.
[2] 16,638 public school children remained eligible after the base year. In addition, 146 public school children were sampled via the sample freshening procedure.
NOTE: Response status is defined in terms of completed child assessment OR completed family structure data of the parent interview. Children who died or moved out of the country were classified as ineligible. Children who moved and were subsampled for followup but could not be located were treated as belonging to the unknown eligibility category. A portion of children who moved was subsampled out and not followed into their new schools.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first, and third grade data collections, school years 1998–1999, 1999-2000, and 2001–02.

Table 4-10. Number (unweighted) of private school children in the ECLS-K sample, by response status and data collection round: School years 1998–99, 1999–2000, and 2001–02

Data collection round	Unweighted sample size	Response status				
		Ineligibles	Unknown eligibility	Non-followed movers	Nonrespondents	Respondents
Fall-kindergarten	4,384	8	†	†	348	4,028
Spring-kindergarten	4,919[1]	30	†	†	412	4,477
Spring-first grade	4,573[2]	11	21	686	192	3,663
Spring-third grade	4,573	23	39	988	288	3,235

† Not applicable.
[1] 535 private school children were sampled from refusal-converted schools.
[2] 4,554 private school children remained eligible after the base year. In addition, 19 private school children were sampled via the sample freshening procedure.
NOTE: Response status is defined in terms of completed child assessment OR completed family structure data of the parent interview. Children who died or moved out of the country were classified as ineligible. Children who moved and were subsampled for followup but could not be located were treated as belonging to the unknown eligibility category. A portion of children who moved was subsampled out and not followed into their new schools.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

The number of children who participated in the base year and first grade and third grade data collections is 13,698 (10,900 in original public schools and 2,798 in original private schools). This represents 64 percent of the base year respondents or 60 percent of children sampled for the base year.

4.6 Calculation and Use of Sample Weights

As in previous years, the ECLS-K data were weighted to compensate for differential probabilities of selection at each sampling stage and to adjust for the effects of nonresponse. As in the first grade year, only child-level weights were computed for third grade. The use of these weights is essential to produce estimates that are representative of the cohort of children who were in kindergarten in 1998–99 or in first grade in 1999–2000. Since the third grade sample was not freshened with third graders who did not have a chance to be sampled in kindergarten or first grade (as was done in first grade), estimates from the ECLS-K third grade data are representative of the population cohort rather than all third graders in 2001–02. The estimated number of children from the ECLS-K is approximately 96 percent of all third graders. While the vast majority of children in third grade in the 2001–02 school year are members of the cohort, third graders who repeated second or third grade and recent immigrants are not covered. Data were collected from teachers and schools to provide important contextual information about the environment for the sampled children. The teachers and schools are not representative of third

grade teachers and schools in 2001–02. For this reason, the only weights produced from the study are for making statements about children, including statements about the teachers and schools of those children.

Several sets of weights were computed for third grade. As in previous years, there are several survey instruments administered to sampled children and their parents, teachers and schools: cognitive and physical assessments for children; parent instruments; several types of teacher instruments; and school instruments. The stages of base year sampling in conjunction with differential nonresponse at each stage and the diversity of survey instruments require that multiple sampling weights be computed for use in analyzing the ECLS-K data. Several combinations of kindergarten through third grade longitudinal weights were also computed. Details on these longitudinal weights are available in chapter 9. This section describes the different types of third grade cross-sectional weights, how they were calculated, how they should be used, and their statistical characteristics.

4.6.1 Types of Sample Weights

Three sets of cross-sectional weights were computed for children in the third grade sample. These weights are defined as follows:

- C5CW0 is nonzero if the child has completed assessment data or the child was excluded from direct assessment due to a disability.
- C5PW0 is nonzero if the child has completed parent interview.
- C5CPTW0 is nonzero if the child has completed assessment data and parent interview data and teacher data from questionnaire part B.

Careful consideration should be given to the choice of a weight for a specific analysis since it depends on the type of data analyzed. Each set of weights is appropriate for a different set of data or combination of sets of data. Exhibit 4-1 summarizes how the different types of cross-sectional weights should be used. Cross-sectional weights are used to provide estimates for the third grade data collection. Details under “to be used for analysis of . . .” provide guidance based on whether the data to be used with the weights were collected through the child assessments, parent interviews, or teacher questionnaires.

Exhibit 4-1. ECLS-K third grade cross-sectional weights: School year 2001-02

Weight	to be used for analysis of ...
C5CW0	third grade direct child assessment data, alone or in conjunction with any combination of (a) a limited set of child characteristics (e.g., age, sex, race/ethnicity), (b) any third grade teacher questionnaire A, B or C data, and (c) data from the school administrator questionnaire or school fact sheet.
C5PW0	third grade parent interview data alone or in combination with (a) third grade child assessment data, (b) third grade teacher questionnaire A, B, or C data, and (c) data from the school administrator questionnaire or school fact sheet. *Exception*: If data from the parent AND child assessment AND teacher questionnaire A or B (not C) are used then C5CPTW0 should be used.
C5CPTW0	third grade direct child assessment data combined with third grade parent interview data AND third grade teacher data alone or in conjunction with data from the school administrator or school fact sheet or facilities checklist.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) third grade data collection, school year 2001–02.

Weight C5CW0 is used to estimate child-level characteristics or assessment scores for third grade. Examples of such estimates are the percent of third grade children who are male, the percent of children who are API, the percent of children who are 9 years old when they begin third grade, and the mean reading score of children in third grade. These weights exist not only for children who were administered a child assessment but also for children who could not be assessed due to a disability.[2] These children were not administered the ECLS-K direct cognitive battery, but their background characteristics such as age, sex, race/ethnicity, and characteristics of their parents, teachers, classrooms, and schools are available from the parent interviews, the teacher questionnaires, the school administrator questionnaire, and the school fact sheet. The academic and social rating scores (see chapter 3) from teachers are also available for children with disabilities, regardless of whether they completed the direct child assessment.

When analyzing child assessment data in conjunction with teacher data collected in third grade, weight C5CW0 should be used. An example of the use of C5CW0 is in the analysis of the relationship between children's approaches to learning as rated by their teachers, the teacher's type of teaching certification, and the children's cognitive skills and knowledge. Some data may be missing because some teachers did not complete the questionnaire, but these are the most appropriate weights for

[2] In kindergarten and first grade, children who were not proficient in English due to a non-English or non-Spanish home language (LM/not Spanish) also had weights even though they were not administered a child assessment. In third grade, this is no longer applicable, since there were no children not assessed due to language ability.

this type of analysis. However, different weights should be used for analysis of child data in conjunction with both parent and teacher data (C5CPTW0).

C5PW0 is used for child-level estimates associated with data collected through the parent interview. Examples are the percent of children whose mothers are currently employed, the percent of children who are in a particular type of child care, and the percent of children who are read to every day. These weights should not be used for estimates solely using direct child assessment data but should be used when analyzing parent and child assessment data together, for example, when exploring the relationship between home literacy behaviors and children's reading skills.

C5CPTW0 is used when child direct assessment *and* teacher *and* parent data are combined in an analysis; for example, in the analysis of the relationship between parent education, teacher education, and children's reading knowledge and skills. These weights should be not be used for estimates using only direct child assessment data or only parent interview data.

Careful consideration should be given to which set of weights is appropriate for the desired analysis. Using the wrong weights will result in more biased or inefficient estimates. For example, if C5CPTW0 were used in an analysis of child and teacher/classroom data only, then the resulting estimates will be inefficient compared to estimates using C5CW0. The lower parent response causes C5CPTW0 to result in lower sample size with positive weights. There may be combinations of data from a different source for which no weights were developed, but most analyses are possible from the weights provided.

The distribution of schools by number of sampled students with nonzero third grade weights and the mean number of sampled students with nonzero weights per school are useful in analysis using hierarchical linear modeling. These are given in table 4-11. In third grade, 70 percent of schools in the sample have five or fewer ECLS-K students with nonzero third grade weights; 94 percent of these schools with small numbers of children are schools where students transferred to (not in tables). For this reason, schools are classified in table 4-11 on the basis of the number of students who had never transferred schools.

Table 4-11. Distribution of originally sampled schools by number of children with nonzero weights and by type of third grade sample weights: School year 2001–02

Sample	Number of cases					Mean cases per school
	1 – 5	6 – 10	11 – 15	16 – 20	21 – 27	
Spring-third grade						
C5CW0	93	224	344	219	18	12
C5PW0	129	301	309	151	9	11
C5CPTW0	180	278	249	131	7	10

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) third grade data collections, school year 2001–02.

4.6.2 Weighting Procedures

The third grade sample included all base year respondents as defined earlier and a supplemental sample of first graders brought in through a sample freshening procedure implemented in spring-first grade. Only a subsample of children who moved from the schools they were attending when they were sampled originally was followed into their new schools. However, children who belong to the language minority group and who had not moved out of the original sample schools at anytime during the first grade year were all followed into their new third grade schools if they moved from the original sample school during their third grade year. The weighting procedures were divided into three main stages.

The first stage of weighting was to compute an initial child weight that reflects the following:

- Adjustment of the school base weight for base year school-level nonresponse;
- Adjustment of the child weights for base year child-level nonresponse; and
- Adjustment of the base year child weight for subsampling of schools for freshening in first grade (for children sampled in first grade only).

The procedures used in this first stage are the same as for the first grade year. They are described again for completeness.

The second stage of weighting was to adjust the initial child weight computed in the first stage for the following:

- Subsampling of movers; and
- Child-level nonresponse.

The third and last stage was to rake the weights adjusted in the second stage to sample-based control totals.

The computation of the initial child weights is described in section 4.6.3. The subsequent weight adjustments are described in section 4.6.4. Section 4.6.5 describes the different types of weights computed for spring-third grade.

In general, in each adjustment to the weight, the adjustment factor is multiplied by the weight in the prior step to get the adjusted weight. This fact is not repeated in the discussions of the weight adjustments in the following sections, only the computation of the adjustment factor is discussed.

4.6.3 Computation of Spring-First Grade Initial Child Weights

As mentioned earlier, the first stage of weighting was to compute an initial child weight that reflects: (1) the adjustment of the school base weight for base year school-level nonresponse (school-level weights), (2) the adjustment of the child weights for base year child-level nonresponse (child-level weights), and (3) the adjustment of the base year child weight for subsampling of schools for freshening in first grade (child-level weights, for children sampled in first grade only). These weights were already computed for spring-first grade. For completeness, they are described below, in section 4.6.3.1 for the school-level weights, and in section 4.6.3.2 for the child-level weights.

4.6.3.1 Base Year Nonresponse-Adjusted School Weights

This weight is the same as that computed for the first grade data collection. It was computed as the school base weight adjusted for base year school-level nonresponse. The base weight for each school was the inverse of the probability of selecting the PSU (county or group of counties) multiplied by

the inverse of the probability of selecting the school within the PSU. For schools selected in the base year through the frame freshening procedure, an additional factor equal to the inverse of the selection probability of the district or diocese was included in the base weight.

A base year responding school was an original sample school with at least one child with a positive C1CW0, C2CW0, C1PW0, or C2PW0 weight. C1CW0 is positive for LM/not Spanish children, children with disabilities and children with at least one direct cognitive test score in fall-kindergarten. C1PW0 is positive for children whose parents completed the family structure questions of the parent interview in fall-kindergarten. C2CW0 and C2PW0 weights are positive under similar circumstances but for spring-kindergarten. Schools that did not meet this condition are nonrespondents and their weights distributed across responding units (at the school level) in this stage. The base year school weight was adjusted within nonresponse weighting classes created in the base year using the Chi-squared Automatic Interaction Detector (CHAID) and variables with known values for both respondents and nonrespondents. School characteristics used for constructing nonresponse cells were the school type (public, Catholic private, non-Catholic private, or nonsectarian private), the school locale (large city, midsize city, suburb of large city, suburb of midsize city, large town, small town, or rural area), the region where the school is located (Northeast, Midwest, South, or West), and the size classification of the school in terms of school enrollment. Once the weighted nonresponse cells were determined, the nonresponse adjustment factors are the reciprocals of the response rates within the selected nonresponse cells.

4.6.3.2 Base Year Child Weights

As mentioned earlier, two groups of children were fielded in spring-third grade: base year respondents, and eligible children who were sampled in first grade as part of the sampling freshening procedure. The base year child weights for the two groups were the same as those computed for the first grade year. A description of them follows.

4.6.3.2.1 Base Year Child Weights for Base Year Respondents

As previously described, a base year respondent was defined as one with at least one direct cognitive test score in fall- or spring-kindergarten or whose parent responded to the family structure section of the parent instrument in fall- or spring-kindergarten. In terms of weights, a base year

respondent is a sampled child with a positive fall- or spring-kindergarten weight (i.e., C1CW0, C2CW0, C1PW0 or C2PW0 weights). The base year child weight is the product of the base year nonresponse-adjusted school weight and the inverse of the within school selection probability of the child, adjusted for child-level nonresponse. The nonresponse weighting classes included school characteristics from the school nonresponse adjustments such as school type, locale, region, school enrollment class, and child characteristics such as age group, sex, and race/ethnicity. These weighting classes are similar to those used for the child weights in the base year. For a description of the computation of child weights in the base year, see chapter 4, section 4.3.4 of the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99* (*ECLS-K) Base Year Public-Use Data Files and Electronic Code Book: User's Manual* (NCES 2001–029: U.S. Department of Education, National Center for Education Statistics, 2000).

4.6.3.2.2 Base Year Child Weights for Eligible Children Sampled in First Grade

Since each child sampled in first grade was directly linked to a child sampled in kindergarten, the first step was to compute a weight for the children who were sampled in kindergarten that reflected the school freshening subsampling and the school freshening nonresponse (some schools refused to provide information needed for freshening). This weight was then linked back to the child sampled in first grade and further adjusted for nonresponse due to not obtaining the data from the sample of freshened children.

First the school base year weight adjusted for school nonresponse (as described in section 4.6.3.1) was adjusted for the subsampling of schools for freshening. Student freshening was done in the same 50 percent subsample of schools that were flagged for following movers in spring-first grade. The school freshening subsampling adjustment factor was computed as:

- 0 if the school was not in the set of schools subsampled for freshening[3] and
- The sum of base year nonresponse-adjusted school weights for all schools over the sum of base year nonresponse-adjusted school weights for schools subsampled for freshening, if the school was in the set of schools subsampled for freshening.

This adjustment was done within cells defined by school type and census region.

[3] These weights, used only to link children sampled in first grade to children sampled in kindergarten, sum up to zero in schools not subsampled for freshening, meaning that there are no children sampled in those schools through freshening.

The freshening procedure could not be applied in all designated schools because some schools did not provide the information needed for freshening. These schools are considered freshening nonrespondents. The school weight adjusted for freshening subsampling was then adjusted for this type of nonresponse. The school freshening nonresponse adjustment factor was calculated as the sum of weights of the freshening-adjusted school weights for all schools designated for freshening over the sum of weights of the freshening-adjusted school weights for schools who responded to freshening. In both the numerator and denominator of this factor, the school measure of size was incorporated; the school measure of size is relevant because the weights will be used for child-level estimates, not school-level estimates. The nonresponse cells for this adjustment were created using school type and urbanicity.

Next, the school-adjusted weight was multiplied by the inverse of the within school selection probability of the child in the base year to obtain a base year child weight. The base year child weight was then adjusted for base year child nonresponse because children who did not respond in the base year could not be linked to children in first grade in spring 2000. The adjustment factor was computed as the sum of the base year child weights of all base year children over the sum of the base year child weights of base year respondents within each nonresponse cell. The nonresponse cells were created using school characteristics such as school type, locale, region, school enrollment class, and child characteristics such as age group, sex, and race/ethnicity.

Only children who did not move from their original schools were designated as links to children in the freshening procedure. The children who moved and were followed into their new schools were not identified to participate in the freshening process in their new schools. As a result, all the children who moved were considered nonrespondents to the freshening process. Additionally, nonmovers and movers who were not in first grade were not eligible for freshening (e.g., if a child was in kindergarten in spring 2000, he or she would be linked only to other kindergarten children and thus was not eligible for the freshening of first graders). Adjustment was necessary to account for these two groups of children and was done in two steps.

In the first step, adjustment was done for movers whose grade was unknown. A portion of the movers was assumed to be in first grade. In the second step, the weights were adjusted for children who were in first grade but who were not identified to participate in the freshening process because they moved into a new school. For this two-step adjustment, each child was classified as a (a) mover in first grade, (b) mover in another grade, (c) mover with unknown grade, (d) nonmover in first grade, and (e) nonmover in another grade.

The first step adjustment for movers whose grade was unknown was computed as

- 0, if the child was a mover with unknown-grade (group c);
- 1, if the child was a nonmover, in first grade or another grade (group d or e); and
- The sum of the nonresponse-adjusted base year child weights (computed in the step before) of all movers (group a, b, or c) over the sum of the nonresponse-adjusted base year child weights of movers with known grade (group a or b), if the child was a mover with known grade (group a or b).

The second step adjustment for movers who could not be used as links for freshening was computed as

- 0, if the child was a first grade mover (group a);
- 1 if the child was in a grade other than first grade (group b or e); and
- The sum of the weights adjusted in step 1 of all first graders (group a or d) over the sum of the weights adjusted in step 1 of nonmovers in first grade (group d), if the child was a nonmover in first grade (group d).

This two-step adjustment was done within cells defined by school type and census region.

The weights thus created for children sampled in kindergarten were then linked to the children that they brought into the sample in first grade through sample freshening. In other words, the weight of the child sampled in first grade was defined at this point to be the weight computed for the child sampled in kindergarten that was responsible for bringing the first grader into the sample.

For the next step in the computation of the spring-first grade child weights, the two groups of children—base year respondents and children sampled in first grade through sample freshening—were put together, and a common variable and label were used to designate the initial child weight. This is the base year child weight as computed above for each group of children.

4.6.4 Computation of Spring-Third Grade Child Weights

The initial child weights described in section 4.6.3 were adjusted for movers between the base year and third grade and nonresponse in third grade, and raked to sampled-based control totals to obtain the final spring-third grade child weights.

4.6.4.1 Adjustment for Movers

First, the initial child weights were adjusted to reflect the subsampling of movers. In the ECLS-K, a child could move more than once and at different times. For example, a child could move out of his original sample school because the school did not have grades higher than kindergarten. Then he could move again between first and third grade. Once a child was identified as a mover, he stayed a mover (unless he moved back to the original sample school). The spring-first grade follow flags were maintained for all children in the spring-third grade sample except for children whose home language was not English. For these language minority children, their spring-first grade flags were switched to 1 if they were not already equal to 1 and if they had not already been subsampled out because they moved in spring-first grade. Thus, children who moved out of their original sample school were followed in the random 50 percent of schools where the follow flag was set to 1, and language minority children were followed at 100 percent if they had not moved previously. The adjustment factor for subsampling movers was computed as follows:

- 1, if the child was not a mover;
- 0, if the child was a mover and the value of the follow flag was 0; and
- The sum of initial child weights of children who were movers over the sum of initial child weights of children who were movers and whose follow flags have value 1, if the child was a mover whose follow flag has value 1.

For the third category, the adjustment factor was computed within cells created using the following characteristics: whether children were sampled in kindergarten or first grade, whether they were movers in spring-first grade, whether they were language minority children, the school type of their original sample school, and the region where their original sample school was located. Seven children with large weights had their weights trimmed by half. However, the weights were not redistributed because the total sum of weights was re-established in the raking procedure that came later.

4.6.4.2 Adjustment for Nonresponse

After the adjustment for subsampling movers, the child weights were adjusted for nonresponse. As in spring-first grade, the nonresponse adjustment was done in two steps. In the first step, the adjustment was for children whose eligibility was not determined (unknown eligibility). A portion of children of unknown eligibility was assumed to be ineligible. In the second step, the adjustment was for eligible nonrespondents. To carry out these adjustments, each child was classified as (a) an eligible respondent, (b) an eligible nonrespondent, (c) ineligible (out of the country or deceased) or (d) of unknown eligibility (mover who could not be located). The first adjustment factor (for children of unknown eligibility) was computed as

- 0, if the child was of unknown eligibility (group d) and
- The sum of the mover adjusted weights of all children (any group) over the sum of the mover adjusted weights of children who were eligible respondents, eligible nonrespondents or ineligible (group a, b or c), if the child was not of unknown eligibility.

The second adjustment factor (for eligible nonrespondents) was computed as

- 0, if the child was an eligible nonrespondent (group b) and
- The sum of the weights adjusted in the first step of eligible children (group a or b) over the sum of the weights adjusted in the first step of eligible responding children (group a), if the child was an eligible respondent.

In both steps of the adjustment, separate nonresponse classes were created for movers and nonmovers using various combinations of response status of child assessments and parent interviews in the base year as well as whether children belong to the language minority group, the type of household collected from the parent interviews (C5PW0 only), and the school type including whether the child was homeschooled (C5CPTW0 only).

4.6.4.3 Raking to Sample-Based Control Totals

To reduce the variability due to the subsampling of schools and movers, the child weights were then raked to sample-based control totals computed using the initial child weights computed in section 4.6.3. The child records included in the file used for computing the control totals are records of

base year respondents and records of eligible children sampled in first grade, including records of children who became ineligible in spring-third grade. The sum of weights thus calculated is the estimated number of third graders in spring 2002. In the previous steps, the weights of the nonresponding children were distributed to the responding children while the weights of the ineligible children were not affected by this weighting step. The weights of the ineligible children are set to zero at the end of this process because these children are not included in the analysis of the spring-third grade data. The reason for including the ineligible children in the raking step is that these children were included in the sampled-based control totals.

Before raking the C5CPTW0 weights, 11 responding movers had their nonresponse-adjusted weights trimmed and the excess weight redistributed among the remaining responding movers so that the sum of weights before trimming was equal to the sum of weights after trimming.

The raking factor was computed separately within raking cells as the sample-based control total for the raking cell over the sum of the nonresponse-adjusted weights for children in the same cell. Raking cells (also known as raking dimensions) were created using school and child characteristics collected in the base year or first grade year: school type, region, urbanicity, sex, age, race/ethnicity, SES, language minority status, whether sampled in kindergarten or first grade, and if sampled in kindergarten, mover status.

4.6.5 Types of Weights and Their Use

The different types of cross-sectional weights are described in section 4.6.1 and their use was summarized in exhibit 4-1. They were all created as described in sections 4.6.4.2 and 4.6.4.3, but the definition of which children were eligible respondents varied. The adjustment for movers was done once, then the resulting weights were adjusted for nonresponse separately for C5CW0, C5PW0 and C5CPTW0.

4.6.5.1 Weights To Be Used With Direct Child Assessment Data (C5CW0)

In spring-third grade, responding children for this type of weight were eligible children who had spring-third grade scorable direct child cognitive assessment data, or children with disabilities who according to specifications in their IEP could not participate in the assessments. A child was eligible if he

or she was a base year respondent or freshened in first grade. Children who transferred to schools and were not flagged to be followed, who moved out of the country or were deceased were ineligible. In spring-third grade, responding children were classified using rules similar to those used in spring-first grade.

Table 4-12 shows the number of children who were not assessed due to the following special situations: children with disabilities, children who moved out of their original sample schools and were not flagged to be followed, children who moved and were flagged to be followed but could not be located or moved into a school in a nonsampled county, and children who moved outside of the country or who were deceased. Of these, only children with disabilities had weights.

Table 4-12. Number of children who were not assessed in spring-third grade, by special situations: School year 2001–02

Special situation	Number of children	
	Unweighted	Weighted
Spring-first grade		
Children with disabilities[1]	75	22,334
Moved from original sample schools		
Subsampled, not to be followed	4,117	784,957
Nonlocatable or moved to nonsampled PSU	1,481	282,443
To be followed but were ineligible in spring	72	13,599

[1] These children's individualized education programs (IEPs) specifically prohibited assessments.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

4.6.5.2 Weights To Be Used With Parent Data (C5PW0)

The weight C5PW0 is to be used with parent interview data. In spring-third grade, a respondent was defined as a child for whom the family structure section (FSQ) in that child's parent interview for the corresponding round was completed. Note that this weight is at the child level even though the data were collected from the parents; they sum to third grade children, not to the parents of third grade children.

4.6.5.3 Weights To Be Used With a Combination of Child Direct Assessment Data and Parent Interview Data and Teacher Data (C5CPTW0)

The weight C5CPTW0 is to be used for analysis involving child, parent, and teacher data. A respondent for this type of weight was defined as a child who had scorable cognitive assessment data for spring-third grade (or children with disabilities), whose parent completed the FSQ section of the parent interview for spring-third grade, and whose teacher completed part B of the teacher questionnaire.

4.6.6 Replicate Weights

For each weight included in the data file, a set of replicate weights was calculated. Replicate weights are used in the jackknife replication method to estimate the standard errors of survey estimates. All adjustments to the full sample weights were repeated for the replicate weights.

For spring-third grade, there are 90 replicate weights. Each set of replicate weights has the same prefix in the variable name as the full sample weight. For example, the replicate weights for C5CW0 are C5CW1 through C5CW90. The method used to compute the replicate weights and how they are used to compute the sampling errors of the estimates are described in section 4.7.

4.6.7 Characteristics of Sample Weights

The statistical characteristics of the sample weights are presented in table 4-13. For each type of weight, the number of cases with nonzero weights is presented together with the mean weight, the standard deviation, the coefficient of variation (i.e., the standard deviation as a percentage of the mean weight), the minimum weight, the maximum weight, the skewness, the kurtosis, and the sum of weights.

The difference in the estimate of the population of students (sum of weights) between rounds of data collection and types of weight is due a combination of factors, among them: (1) the number of first graders who became ineligible in third grade (due to death, leaving the country, or being a nonsampled mover), and (2) the adjustment of the weights for the children of unknown eligibility.

Table 4-13. Characteristics of the first grade child-level weights: School year 1999–2000

Sample	Number of cases	Mean	Standard deviation	CV (× 100)	Minimum	Maximum	Skewness	Kurtosis	Sum
C5CW0	14,470	272.18	242.53	89.10	1.54	3,376.78	3.21	18.45	3,938,513
C5PW0	13,489	291.92	241.71	82.80	1.63	3,654.05	3.23	18.83	3,937,758
C5CPTW0	10,395	378.75	435.34	114.94	2.58	5,209.19	3.38	15.25	3,937,125

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

4.7 Variance Estimation

The precision of the sample estimates derived from a survey can be evaluated by estimating the variances of these estimates. For a complex sample design such as the one employed in the ECLS-K, replication and Taylor Series methods have been developed. These methods take into account the clustered, multistaged characteristics of sampling and the use of differential sampling rates to oversample targeted subpopulations. For the ECLS-K, in which the first-stage self-representing sampling units, (i.e., PSUs) were selected with certainty and the first-stage non-self-representing sampling units were selected with two units per stratum, the paired jackknife replication method (JK2) is recommended. This section describes the JK2 and the Taylor Series estimation methods.

4.7.1 Paired Jackknife Replication Method

In this method, a survey estimate of interest is calculated from the full sample. Subsamples of the full sample are then selected to calculate subsample estimates of the same parameter. The subsamples are called *replicates*, and the subsample estimates are called *replicate estimates*. The variability of the replicate estimates about the full sample estimate is used to estimate the variance of the full sample estimate. The variance estimator is computed as the sum of the squared deviations of the replicate estimates from the full sample estimate:

$$v\left(\hat{\theta}\right) = \sum_{g=1}^{G}\left(\hat{\theta}_{(g)} - \hat{\theta}\right)^2,$$

where

θ is the survey estimate of interest,
$\hat{\theta}$ is the estimate of θ based on the full sample,
G is the number of replicates formed, and
$\hat{\theta}_{(g)}$ is the g^{th} replicate estimate of θ based on the observations included in the g^{th} replicate.

The variance estimates of selected survey items presented in section 4.8 were produced using WesVar and JK2.

Replicate weights were created to be used in the calculation of variance estimates. Each replicate weight was calculated using the same adjustment steps as the full sample weight but using only the subsample of cases that constitute each replicate. For the original ECLS-K design in the base year, replicate weights were created taking into account the Durbin method of PSU selection. The Durbin method selects two first-stage units per stratum without replacement, with probability proportional to size and a known joint probability of inclusion.

In the ECLS-K PSU sample design, there were 24 SR strata and 38 NSR strata. Among the 38 NSR strata, 11 strata were identified as Durbin strata and were treated as SR strata for variance estimation. The purpose of the Durbin strata is to allow variances to be estimated as if the first-stage units were selected with replacement. This brings the number of SR PSUs to 46 (24 original SR PSUs and 22 Durbin PSUs from the 11 Durbin strata). The remaining 54 NSR PSUs are in 27 NSR strata; thus 27 replicates were formed, each corresponding to one NSR stratum. For the SR strata, 63 replicates were formed. The 90 replicates will yield about 76 degrees of freedom for calculating confidence intervals for many survey estimates.

As stated earlier, the sample of PSUs was divided into 90 replicates or variance strata. The 27 NSR strata formed 27 variance strata of two PSUs each; each PSU formed a variance unit within a variance stratum. All schools within an NSR PSU were assigned to the same variance unit and variance stratum. Sampled schools in the 46 SR PSUs were grouped into 63 variance strata. In the SR PSUs, schools were directly sampled and constituted PSUs. Public schools were sampled from within PSU while private schools were pooled into one sampling stratum and selected systematically (except in the SR PSUs identified through the Durbin method where private schools were treated as if they were sampled from within PSU). Schools were sorted by sampling stratum, school type (from the original sample or newly selected as part of freshening), type of frame (for new schools only), and their original order of selection (within stratum). From this sorted list, they were grouped into pairs within each sampling

stratum; the last pair in the stratum may be a triplet if the number of schools in the stratum is odd. This operation resulted in a number of ordered preliminary variance strata of two or three units each. The first ordered 63 strata were then numbered sequentially from 1 to 63; the next ordered 63 strata were also numbered sequentially from 1 to 63, and so on until the list was exhausted, thus forming the desired 63 variance strata.

In strata with two units, a unit being a PSU in the case of NSR PSUs and a school in the case of SR PSUs, the base weight of the first unit was doubled to form the replicate weight, while the base weight of the second unit was multiplied by zero. In strata with three units, two variance strata were created: in the first variance stratum, the base weight of two of the three units was multiplied by 1.5 to form the replicate weight and the base weight of the last unit was multiplied by zero; in the second variance stratum, the base weight of a different group of two units was multiplied by 1.5, and the base weight of the third unit was multiplied by zero. Multiplying the base weight in a unit by zero is equivalent to dropping one unit as required by the jackknife method. All adjustments to the full sample weights were repeated for the replicate weights. For each full sample weight, there are 90 replicate weights with the same weight prefix.

A child sampled in first grade through the freshening process was assigned to the same replicate as the originally sampled child to whom the child was linked. When the child sampled in first grade was assigned a full sample weight (see section 4.6.3.2), he or she was assigned the replicate weights in the same manner.

To reflect the variability of the control totals in the sample-based raking, a set of replicate control totals was created. Each replicate was then raked to the corresponding replicate-based control totals. This resulted in each replicate retaining the variability associated with the original sample estimates of the control totals.

The replicate weights can be used with software such as WesVar, SUDAAN and AM.

4.7.2 Taylor Series Method

The Taylor Series method produces a linear approximation of the survey estimate of interest; then the variance of the linear approximation can be estimated by standard variance formulas. The stratum

and first-stage unit (i.e., PSU) identifiers needed to use the Taylor Series method were assigned taking care to ensure that there were at least two responding units in each stratum. A stratum that did not have at least two responding units was combined with an adjacent stratum. For the ECLS-K, the method of stratifying first-stage units was the same for each type of cross-sectional weight. For each type of weight, the sample size was examined, then strata were combined when the sample size was not adequate. The sequential numbering of strata and first-stage units was done separately for each weight. Consequently, there is a different set of stratum and first-stage unit identifiers for each set of weights.

Stratum and first-stage unit identifiers are provided as part of the ECLS-K data file and can be used with software such as SUDAAN, STATA, SAS or AM. They are described in table 4-14.

Table 4-14. ECLS-K Taylor Series stratum and first-stage unit identifiers

Variable name	Description
C5TCWSTR	Sampling stratum—spring-third grade C-weights
C5TCWPSU	First-stage sampling unit within stratum—spring-third grade C-weights
C5TPWSTR	Sampling stratum—spring-third grade P-weights
C5TPWPSU	First-stage sampling unit within stratum—spring-third grade P-weights
C5CPTSTR	Sampling stratum—spring-third grade CPT-weights
C5CPTPSU	First-stage sampling unit within stratum—spring-third grade CPT-weights

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

4.7.3 Specifications for Computing Standard Errors

Specifications for computing standard errors are given in table 4-15. For each type of analysis described in the table, users can choose the replication method or the Taylor Series method for computing standard errors.

Table 4-15. Specifications for computing standard errors, spring-third grade: School year 2001–02

		Computing standard errors					Approximating sampling errors
		Replication method (WesVar, SUDAAN or AM)			Taylor Series method (SUDAAN, STATA, SAS or AM)		
Type of analysis	Full sample weight	ID	Replicate weights	Jackknife method	Sample design	Nesting variables	DEFT (Average root design effect)
Spring-third grade cross-sectional	C5CW0	CHILDID	C5CW1 – C5CW90	JK2	WR[1]	C5TCWSTR C5TCWPSU	1.841
	C5PW0	PARENTID	C5PW1 – C5PW90	JK2	WR[1]	C5TPWSTR C5TPWPSU	
	C5CPTW0	CHILDID	C5CPTW1 – C5CPTW90	JK2	WR[1]	C5CPTSTR C5CPTPSU	

[1] WR = with replacement, specified only if using SUDAAN. WR is the only option available if using SAS, STATA or AM.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

For the replication method, the full sample weight, the replicate weights, and the method of replication are required parameters. All analyses of the ECLS-K data should be done using JK2. As an example, to compute spring-third grade child-level estimates (e.g., mean reading scores) and their standard errors, users need to specify CHILDID in the ID box of the WesVar data file screen, C5CW0 as the full sample weight, C5CW1 to C5CW90 as the replicate weights, and JK2 as the method of replication.

For the Taylor Series method using SUDAAN, STATA, SAS or AM, the full sample weight, the sample design, the nesting stratum and PSU variables are required. For the same example above, the full sample weight (C5CW0), the stratum variable (C5TCWSTR), and the PSU variable (C5TCWPSU) must be specified. The "with replacement" sample design option, WR, must also be specified if using SUDAAN.

The last column in table 4-15 gives the average root design effect that can be used to approximate the standard errors for each type of analysis. For a discussion of the use of design effects, see section 4.8.1.

4.8 Design Effects

An important analytic device is to compare the statistical efficiency of survey estimates from a complex sample survey such as the ECLS-K, with what would have been obtained in a hypothetical and usually impractical simple random sample (SRS) of the same size. In a stratified clustered design like the ECLS-K, stratification generally leads to a gain in efficiency over simple random sampling, but clustering has the opposite effect because of the positive intracluster correlation of the units in the cluster. The basic measure of the relative efficiency of the sample is the *design effect*, defined as the ratio, for a given statistic, of the variance estimate under the actual sample design to the variance estimate that would be obtained with an SRS of the same sample size:

$$DEFF = \frac{Var_{DESIGN}}{Var_{SRS}}.$$

The root design effect, *DEFT*, is defined as:

$$DEFT = \sqrt{DEFF} = \frac{SE_{DESIGN}}{SE_{SRS}},$$

where *SE* is the standard error of the estimate.

4.8.1 Use of Design Effects

Methods of computing standard errors for the ECLS-K are replication and Taylor Series linearization. If a statistical analysis software package such as SPSS (Statistical Program for the Social Sciences) is used, the standard errors should be corrected using *DEFT*, since these programs calculate standard errors, assuming the data were collected with a simple random sample. The standard error of an estimate under the actual sample design can be approximated as follows:

$$SE_{DESIGN} = \sqrt{DEFF \times Var_{SRS}} = DEFT \times SE_{SRS}.$$

Packages such as SAS or SPSS can be used to obtain Var_{SRS} and SE_{SRS}. Alternatively, Var_{SRS} and SE_{SRS} can be computed using the formulas below for means and proportions.

Means:
$$Var_{SRS} = \frac{1}{n}\frac{\sum_{1}^{n} w_i \left(x_i - \overline{x_w}\right)^2}{\sum_{1}^{n} w_i} = SE_{SRS}^2,$$

where w_i are the sampling weights, n is the number of respondents in the sample, and the sample mean $\bar{x}_w$ is calculated as follows:

$$\bar{x}_w = \frac{\sum_{1}^{n} w_i x_i}{\sum_{1}^{n} w_i}.$$

Proportions: $$Var_{srs} = \frac{p(1-p)}{n} = SE_{SRS}^2,$$

where p is the weighted estimate of proportion for the characteristic of interest and n is the number of cases in the sample.

In both cases of means and proportions, the standard error assuming SRS should be multiplied by *DEFT* to get the approximate standard error of the estimate under the actual design.

4.8.2 Average Design Effects for the ECLS-K

In the ECLS-K, a large number of data items were collected from students, parents, teachers, and schools. Each item has its own design effect that can be estimated from the survey data. One way to produce design effects for analysts' use is to calculate them for a number of variables and average them. The averaging can be done overall and for selected subgroups. The tables that follow show estimates, standard errors, and design effects for selected means and proportions based on the ECLS-K third grade child, parent, teacher, and school data. For each survey item, the tables present the number of cases for which data are nonmissing, the estimate, the standard error taking into account the actual sample design (Design SE), the standard error assuming SRS (SRS SE), the root design effect (DEFT), and the design effect (DEFF). Standard errors (Design SE) were produced in WesVar using JK2 based on the actual ECLS-K complex design.

For each survey estimate, the variable name as it appears in the ECLS-K first grade Electronic Code Book is also provided in the table. For more information on the variables used in this section, refer to chapter 3, which describes the assessment and rating scale scores used in the ECLS-K, and chapter 7, which has a detailed discussion of the other variables.

Standard errors and design effects for the child-level items are presented in tables 4-16 and 4-17. The survey items were selected so that there was a mix of items from the direct child assessment, the parent interview, and the teacher child-level questionnaire They include the different scores from the direct child assessment, the social rating scores as provided by teachers, characteristics of the parents, and

characteristics of the students as reported by the parents and teachers. For a small number of estimates, the data were subset to cases where the estimate is applicable; for example, the proportion of children who have access to the internet is only for children in households with a computer.

Table 4-16 presents design effects for the third grade sample, with a median design effect of 3.3. Table 4-17 presents the median design effects for subgroups based on school type, child's sex and race/ethnicity, geographic region, level of urbanicity, and the socioeconomic scale (SES quintiles) of the parents. The median design effect varies from 1.3 (Pacific Islanders) to 3.6 (children in small towns and rural areas). The variation in the design effects is largely a function of the sample size as well as the homogeneity of the children within schools.

In spring-third grade, as in spring-first grade, design effects are not computed for items from the teacher and school administrator's questionnaires since there are no teacher or school weights computed for the first grade year nor for the third grade year. Although standard errors and design effects may also be calculated for the teacher and school administrator's questionnaires at the child level, they are quite large compared to those typically found for the ECLS-K data. Design effects for teacher and school items are large because the intraclass correlation is 100 percent for children in the same school and very high for children in the same class; children attending the same school have the same school data, and children in the same class have the same teacher data.

Table 4-16. ECLS-K standard errors and design effects by selected child and parent variables, for the full sample—child and parent data: School year 2001–02

Survey item	Variable name	Number of cases	Estimate	Design SE[1]	SRS SE[2]	DEFT[3]	DEFF[4]
Child scores (mean)							
Reading scale score	C5R2RSCL	14,280	106.05	0.437	0.173	2.521	6.354
Math scale score	C5R2MSCL	14,374	83.25	0.433	0.152	2.843	8.083
Science scale score	C5SSCALE	14,351	33.46	0.222	0.083	2.661	7.079
Self-described : Externalizing problems	C5SDQEXT	14,379	2.02	0.012	0.006	2.063	4.254
Self-described : Internalizing problems	C5SDQINT	14,379	2.23	0.013	0.006	2.187	4.782
Self-described : Competence in math	C5SDQMTC	14,379	3.16	0.011	0.006	1.718	2.951
Self-described : Competence in peer relation	C5SDQPRC	14,378	3.03	0.007	0.005	1.323	1.750
Self-described : Competence in reading	C5SDQRDC	14,379	3.26	0.009	0.005	1.690	2.855
Self-described : Competence in all subject	C5SDQSBC	14,379	2.92	0.008	0.005	1.497	2.240
Approaches to learning-Teacher	T5LEARN	11,701	2.99	0.010	0.007	1.526	2.330
Self-control-Teacher	T5CONTRO	11,592	3.17	0.010	0.006	1.651	2.725
Interpersonal-Teacher	T5INTERP	11,558	3.05	0.009	0.006	1.407	1.979
Externalizing problems-Teacher	T5EXTERN	11,676	1.73	0.010	0.006	1.760	3.097
Internalizing problems-Teacher	T5INTERN	11,577	1.67	0.008	0.005	1.590	2.529
Child and parent characteristics from parent interview (percent)							
Lived in single parent family	P5HFAMIL	13,489	27.80	0.627	0.386	1.626	2.644
Lived in two-parent family	P5HFAMIL	13,489	69.70	0.711	0.396	1.797	3.229
Mom worked 35 hours+/week	P5HMEMP	9,790	67.08	0.782	0.475	1.646	2.709
Primary care is center-based	P5PRIMNW	4,765	36.41	1.315	0.697	1.887	3.561
Primary care is home-based	P5PRIMNW	4,765	63.59	1.315	0.697	1.887	3.561
Parents had high school or less	W3PARED	13,489	33.42	0.882	0.406	2.172	4.716
Household income category below median	W3INCCAT	13,489	42.88	0.995	0.426	2.335	5.450
Parent attended PTA	P5ATTENP	13,470	43.03	0.887	0.426	2.080	4.327
Practiced reading, writing, numbers every day	P5RDWRNM	13,364	48.90	0.609	0.433	1.407	1.981
Visited library	P5LIBRAR	13,362	54.01	0.807	0.431	1.871	3.500
Used computer 1-2 times per week	P5HOMECM P5COMPWK	10,671	41.57	0.625	0.477	1.311	1.718
Had internet access	P5HOMECM P5INTACC	10,417	85.61	0.466	0.344	1.356	1.840
Used computer 1-2 times per week for homework	P5HOMECM P5CMPEDU	10,394	52.58	0.593	0.490	1.210	1.465
Had family rule for TV	P5TVHOME P5TVRULE	13,285	91.58	0.323	0.241	1.342	1.802
Did homework 3-4 times per week	P5OFTDHW	13,344	40.55	0.675	0.425	1.588	2.522
Have someone help with reading homework	P5HELPR	13,161	98.50	0.164	0.106	1.546	2.389
Completely true that child and self have close time	P5WARMCL	13,021	68.48	0.572	0.407	1.404	1.971
Took away privilege when child angry	P5HITPRV	13,016	62.17	0.996	0.425	2.342	5.486
Self-reported in very good health	P5HEALTH	12,995	86.81	0.560	0.297	1.885	3.555
Household received food stamp in last 12 months	P5FSTAMP	13,256	15.15	0.729	0.311	2.341	5.479
Child characteristics from teacher questionnaire C (percent)							
Enrolled in third grade	T5GRADE	11,721	88.33	0.619	0.297	2.086	4.351
Average in language skills	T5RTLANG	11,581	70.78	0.773	0.423	1.828	3.343
Average in science/social studies	T5RTSCI	11,500	79.68	0.891	0.375	2.375	5.639
Average in math skills	T5RTMTH	11,544	77.21	0.802	0.390	2.054	4.219

See notes at end of table.

Table 4-16. ECLS-K standard errors and design effects by selected child and parent variables, for the full sample—child and parent data: School year 2001–02—Continued

Survey item	Variable name	Number of cases	Estimate	Design SE[1]	SRS SE[2]	DEFT[3]	DEFF[4]
Child characteristics (mean)							
Age of child in months	R5AGE	14,393	111.22	0.093	0.040	2.346	5.506
Child's BMI	CBMI	13,866	18.69	0.044	0.033	1.332	1.775
Child's household size	P5HTOTAL	13,489	4.56	0.024	0.012	1.960	3.843
Number of children <18 in child's HH	P5LESS18	13,489	2.52	0.020	0.010	2.009	4.036
Number of siblings in HH	P5NUMSIB	13,489	1.55	0.018	0.010	1.850	3.421
Number of hours watched TV after dinner	P5TVAFDH	13,263	0.89	0.012	0.007	1.700	2.891
Median						1.813	3.343
Mean						1.841	3.559
Standard deviation						0.397	1.567
Coefficient of variation						0.216	0.440
Minimum						1.210	1.465
Maximum						2.843	8.083

[1]Design SE is the standard error under the ECLS-K sample design. For an explanation of this statistic, see section 4.8.
[2] SRS SE is the standard error assuming simple random sample. For an explanation of this statistic, see section 4.8.
[3] DEFT is the root design effect. For an explanation of this statistic, see section 4.8.
[4] DEFF is the design effect. For an explanation of this statistic, see section 4.8.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 4-17. ECLS-K median design effects for subgroups—child, parent, and teacher questionnaire part C data: School year 2001–02

Subgroups	Spring-third grade	
	DEFT[1]	DEFF[2]
All students	1.813	3.343
School type		
Public	1.669	2.841
Private	1.788	3.203
Catholic private	1.796	3.262
Other private	1.585	2.576
Sex		
Male	1.606	2.641
Female	1.512	2.326
Race/ethnicity		
White	1.717	2.945
Black	1.514	2.305
Hispanic	1.392	1.962
Asian	1.481	2.178
Pacific Islander	1.134	1.277
American Indian	1.387	1.905
Other	1.369	1.870
Region		
Northeast	1.675	2.812
Midwest	1.789	3.208
South	1.765	3.290
West	1.663	2.765
Urbanicity		
Central city	1.608	2.601
Urban fringe and large town	1.636	2.658
Small town and rural area	1.903	3.558
SES quintiles		
First	1.462	2.128
Second	1.430	2.049
Third	1.429	2.037
Fourth	1.376	1.868
Fifth	1.453	2.112

[1]DEFT is the root design effect. For an explanation of this statistic, see section 4.8.
[2]DEFF is the design effect. For an explanation of this statistic, see section 4.8.
NOTE: Each median is based on 44 items.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

REFERENCES

Durbin, J. (1967). Design of Multi-stage Surveys for the Estimation of Sampling Errors. *Journal of the Royal Statistical Society*, Series C (Applied Statistics), Vol. 16, 152-164.

Kish, L. (1965). *Survey Sampling*. New York: John L. Wiley & Sons.

U.S. Department of Education, National Center for Educational Statistics (1995–96). *Common Core of Data, Public School Universe Survey*. Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics (2000). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) Base Year Public-Use Data Files and Electronic Code Book: User's Manual* (NCES 2001–029) Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics (2002). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) User's Manual for the ECLS-K First Grade Public-Use Data Files and Electronic Code Book* (NCES 2002-135). Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics (1998). *Private School Universe Survey, 1995–96*. Washington, DC: U.S. Department of Education.

U.S. Department of the Interior, Bureau of Indian Affairs (1995-96). *Office of Indian Education Programs Education Directory*. Washington, DC: U.S. Department of the Interior.

5. DATA COLLECTION METHODS AND RESPONSE RATES

The following sections discuss the data collection procedures and response rates in the third grade data collection phase of the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K). Section 5.1 gives an overview of the data collection methods. Detailed information is provided on field staff training (section 5.2), preassessment school contacts (section 5.3), spring-third grade data collection (section 5.4), and quality control procedures (section 5.5). Spring-third grade completion rates are presented and discussed in section 5.6.

5.1 Overview of Data Collection Methods

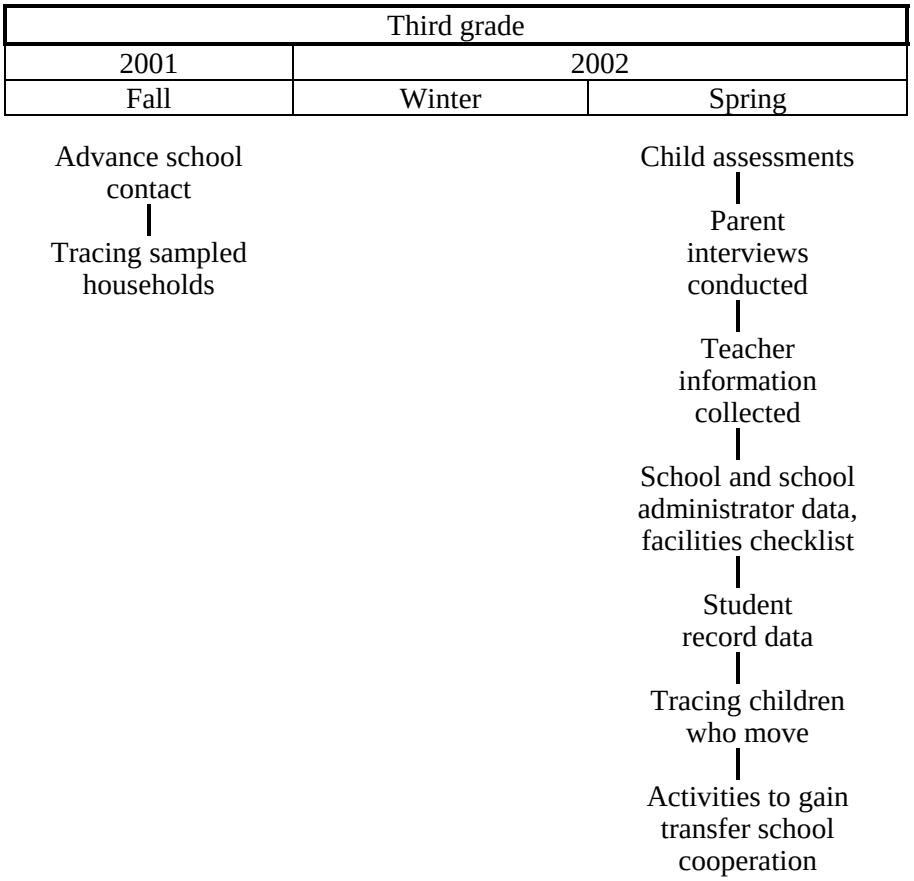
The ECLS-K third grade data collection was conducted in the fall and spring of the 2001–02 school year. Fall data collection included contacting sampled schools to set appointments to conduct the child assessments in the spring of the school year, verifying the parent consent procedures, linking children to teachers, identifying children who had withdrawn from the school, and obtaining locating information about their new schools. Spring data collection included the direct child assessments, parent interviews, teacher and school questionnaires, student record abstract, and facilities checklist. The activities, begun in fall data collection, to locate children and gain cooperation of the schools into which they transferred continued in spring data collection. The content and timeline of the third grade data collections are shown in exhibit 5-1.

Computer-assisted personal interviewing (CAPI) was the mode of data collection for the child assessments, and telephone and in-person computer-assisted interviewing (CAI) was the mode of data collection for the parent interview; self-administered questionnaires were used to gather information from teachers, school administrators, and student records. The facilities checklist was completed by field staff.

5.2 Field Staff Training

Several in-person training sessions were conducted to prepare staff for the third grade data collection. In the fall of 2001, supervisors were trained to contact original schools and recruit transfer

Exhibit 5-1. Timeline of third grade data collection: School year 2001–02



SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

schools. In the spring of 2002, three trainings were held: one for staff who only conducted parent interviews, one for field supervisors, and one for assessors. Field supervisors managed all the data collection activities within their assignment, supervising the assessors and interviewers and conducting child assessments and parent interviews. Assessors conducted the child assessments and parent interviews. Interviewers only conducted parent interviews.

5.2.1 Advance Contact and Recruitment Training

In September 2001, field supervisors were trained for 3 days to contact original sampled schools and transfer schools to set up the data collection in the spring. A total of 50 field supervisors and

2 field managers completed training. Topics included an overview of study activities to date, verifying parent consent procedures, identifying and locating children who moved from the schools they attended in the first grade, identifying the teachers of ECLS-K children and linking them to those children, and exercises on scheduling schools efficiently within a work area.

As in the first grade training, advance contact and recruitment training was conducted using the automated Field Management System (FMS). The FMS was used throughout the data collection period to enter information about the sampled children, parents, teachers, and schools and to monitor production on all data collection activities. The field supervisors entered information into the FMS during training presentations, thus acquiring hands-on experience with the FMS and all field procedures prior to beginning data collection. The field supervisors completed role plays and exercises that involved entering information into the FMS.

5.2.2 Spring-Third Grade Training

Field supervisors, interviewers, and assessors were trained for the spring-third grade data collection in three sessions in February and March 2002. Prior to the March in-person training session, supervisors and assessors completed 8 hours of home study training on the study design, field procedures, and computer keyboard skills.

5.2.2.1 Parent Interviewer-Only Training

Supervisors and staff assigned to complete only parent interviews during the spring data collection attended a 2-day training in February 2002. Trainers presented the content of the parent interview and discussed protocols for interviewing. The interviewers practiced using the CAI system on laptops during interactive lectures and role plays. Supervisors had an additional day of training to practice using the FMS to organize and track production and to discuss management techniques for overseeing their teams of interviewers. Eight (8) supervisors and 66 interviewers completed training.

5.2.2.2 Field Supervisor Training

Field supervisor training preceded the assessor training and lasted for 3 days. The topics covered in the field supervisor training session included reviewing materials from the fall school recruitment, role plays to practice contacting school coordinators, identifying and locating children who moved from their first grade schools, identifying the regular and special education teachers of ECLS-K children and linking them to those children, distributing and following up on teacher questionnaires and school administrator questionnaires, completing the facilities checklist, and conducting quality control observations. Field supervisors were also trained to use the FMS, and the field supervisors entered information into the FMS during training presentations. Seventy-seven (77) field supervisors completed training.

5.2.2.3 Assessor Training

The assessor training sessions included an overview of study activities to date, interactive lectures based on the direct child assessments and the parent interview, practice parent interviews in pairs using role-play scripts, practice direct child assessments using role-play scripts, direct child assessment precertification exercises on each form of the direct child assessments, techniques for parent refusal avoidance, and strategies for building rapport with children. A major goal of the assessor training was to train field staff in the proper procedures to conduct the direct child assessments. This included following standardized procedures for administration of all assessment items as well as giving children neutral praise with the sampled children. The sessions provided trainees with hands-on experience with all the direct child assessment materials and procedures and the CAI programs prior to data collection. Interactive lectures and role plays were also used to train field staff in administering the parent interviews. Trainees practiced entering information into the CAI system on laptop computers during training presentations on conducting the direct child assessments and parent interview. Assessor training lasted for 5 days; field supervisors were also trained to perform all assessor activities. Two hundred sixty-six (266) assessors and 77 field supervisors completed training.

5.2.2.4 Certification of the Child Assessors

In order to ensure that the supervisors and assessors who completed training administered the direct child assessments in a standardized manner, all field staff completed certification exercises. Certification was composed of written exercises on each level form of each of the assessment domains (e.g., the red form of reading, which corresponds to a low difficulty level) and an observation of each trainee administering the assessment to children specifically recruited for the training sessions. Each level form of an assessment domain was reviewed in detail during an interactive lecture. Time was then given to each trainee to review and practice administering it individually. After the individual practice, written exercises were distributed.

The written exercises were used to ensure that each trainee understood the coding rules for selected open-ended questions with particularly complex scoring rubrics. Each exercise included certain assessment items from the level form that was just discussed, with an assortment of possible responses. The trainees were instructed to score each response as either correct or incorrect. The exercises were then scored by the co-trainer during the next training session. Trainees who did not achieve a passing score were asked to attend a training session in the evening to review the items. These trainees then re-took the same exercises that they had previously failed to pass.

Most trainees passed the written exercises on the first attempt. All of the trainees who had to re-take the exercises after the remedial evening session achieved a passing score. Just over a quarter of the trainees (26 percent or 84 trainees) did not pass at least one element of the certification exercises on the first attempt. Nineteen (19) percent did not pass the reading exercises, with only a small percentage not passing the math (2 percent) and science (5 percent) exercises. Most likely, this was due to the complexity of the reading scoring rubrics and the unfamiliarity of the exercises themselves (reading exercises were distributed first, with math and science exercises on later days). Once additional training was given, all of the trainees passed the exercises on the second attempt. Refer to the *Early Childhood Longitudinal Study, Kindergarten Class of 1989–1999, Third Grade Methodology Report* (U.S. Department of Education, National Center for Education Statistics, forthcoming [c]) for additional detail.

In the final stage of the certification process, the trainees were observed conducting a direct child assessment with children brought on site to the training session. Training staff who were already certified on the assessment observed trainees as they administered parts (e.g., routing test and the yellow reading level—see chapter 3) of the assessment to third grade children. They rated the trainees on skills

such as rapport with the child, avoidance of coaching or use of inappropriate probing, following proper administration procedures, and pacing. While the trainee administered the assessment, an observer certified on the assessment simultaneously coded the child's answers to preselected open-ended questions. After the assessment was completed, the observer brought up a screen in the CAPI program that displayed the assessor's coding of the open-ended questions. The answers recorded by the assessor were compared with those recorded by the observer. Discrepancies in any of the recorded answers were included in the assessor's overall score on a certification form.

Table 5-1 presents the results of the training certification. Trainees who scored 85 percent or above were certified as qualified to administer the child assessments. Trainees who scored between 70 and 84 percent were required to complete remedial training and an additional certification in the field before beginning assessments.

Table 5-1. Number and percent of trainees, by scores on training certification form: School year 2001–02

Trainees	Number	Percent
Total	343	100
Score on certification form		
85 percent or above	337	98.2
70–84 percent	6	1.8
Below 70 percent	0	0

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

The majority of the trainees (98.2 percent) scored at or above 85 percent on the certification form, with only 1.8 percent scoring between 70 and 84 percent. None of the trainees failed to meet the 70 percent threshold on the assessment certification form. All trainees who needed remedial training were certified qualified to administer the child assessments after they conducted a second assessment on a third grade child who was not part of the ECLS-K sample.

5.3 Fall Preassessment School Contact

Beginning in September 2001, all participating ECLS-K schools, i.e., schools that participated in fall or spring of kindergarten or first grade, were contacted by telephone to prepare for the

spring data collection. When children were identified as transferring to another school, the child's new school (and district, if necessary) was recruited.

5.3.1 Advance Mailings

In September 2001, an advance package was mailed via Federal Express to all participating ECLS-K schools asking them to prepare for the preassessment contact telephone call. The schools were asked to identify a school staff coordinator to serve as a liaison with the study (in returning schools, this person was usually the coordinator from previous rounds of data collection). The advance package contained study findings from first grade and an overview of third grade data collection activities. The school coordinators were asked to complete an information form about the ECLS-K sampled children prior to the telephone call.

5.3.2 Preassessment Contact

The preassessment contact was made by telephone between September and November 2001. The preassessment school contact was successful in meeting two important goals: (1) contacting original sampled schools to set up the spring assessment and (2) identifying children who withdrew from their spring-first grade school. Schools were determined to be ineligible for third grade data collection if no sampled children were currently enrolled. Original sampled schools became ineligible if second grade was the highest grade in the school or if the school had closed, that is, was no longer operational. More transfer schools were determined to be ineligible as children transferred out of them into other schools. During the preassessment contact, the field supervisor contacted the school coordinator to schedule the dates of the assessment visit for original sampled schools, identified ECLS-K sampled children who were no longer enrolled at the school, collected locating information for those children, identified each enrolled child's regular and special education teacher, reviewed parental consent status, obtained information on special accommodations[1] during assessment for the enrolled sampled children, and answered any questions the school coordinator may have had.

[1] Accommodations included in the data collection protocol were special setting accommodations, scheduling/timing accommodations, presence of a health care aide, or use of an assistive device.

5.3.2.1 Identifying ECLS-K Sampled Children Who Withdrew From the School

Field supervisors asked the school coordinators to identify ECLS-K children who transferred out of the school. If the school records indicated where the children had transferred, then the field supervisors asked the school coordinator to provide the names, addresses, and telephone numbers of these transfer schools. Of those children who transferred, only a subset were followed to their new school (see section 4.4.1 in chapter 4 for more detail on how mover children were subsampled). If the new school belonged to a district that was new to the study, the district was contacted and recruited before any contact was made with the school. If the district was already cooperating, the new school was contacted and recruited directly.

5.3.2.2 Reviewing Information About ECLS-K Sampled Children

Field supervisors collected information from the school coordinators about the ECLS-K sample children still enrolled in the school, including the child's current grade, the name and classroom of the child's regular teacher, and whether or not the child had an Individualized Education Plan (IEP). If the child had an IEP, then the name and classroom of the child's special education teacher were collected, along with whether the child required any accommodations to participate in the direct cognitive assessment. The accommodations to the third grade direct cognitive assessment were the same as those for the kindergarten and first grade direct cognitive assessments. Field supervisors contacted the teachers of the ECLS-K children as necessary for any of this information.

5.3.2.3 Reviewing Parent Consent

Although parental consent was obtained in the base year (and, in some schools, in the first grade year), field supervisors reviewed the parental consent with the school coordinator to determine if the base year or first grade consent was acceptable for third grade. If the schools required consent to be re-obtained or changed the type of consent that was required (e.g., from implicit to explicit), parent letters and consent forms were mailed either to the school for distribution to parents or directly to parents from Westat, based on the schools' preference. Parents were requested to return signed consent forms to the school coordinator.

5.3.2.4 Contacting Families of Home-Schooled Children

As part of the advance school contact, children who were home schooled in previous rounds were identified. The status of home-schooled children who were identified in round 1 through 4 was verified with their parents and updated as necessary. In addition, some home-schooled children were identified by the schools during the preassessment contact. Their status was also verified with their parents during data collection. Parents of these children were contacted in September through November 2001 to determine if the child was still home schooled or had enrolled in a school. If the child had enrolled in a school, the new school was contacted and recruited into the study. Parents of children who were still schooled at home were notified about the next round of data collection in the spring.

5.3.3 Preparing for Spring-Third Grade Data Collection

In order to ensure that as many of the sampled children as possible were contacted in the spring, locating efforts were undertaken in the winter of 2001. Staff in Westat's Telephone Research Center (TRC) traced children who could not be located during the preassessment school contact phase. TRC staff also used the Internet, telephone directories, and other means to locate these children and their households. When children and/or households were found, the new school and contacting information was entered into the computer database, for fielding in the spring. Table 5-2 presents the results of this effort. See section 5.4.4 for more details about children who transferred schools in third grade.

A mailing to the post office requesting change of address information for sampled households was also conducted in the winter of 2001.

5.4 Spring-Third Grade Data Collection

All children who were assessed during the base year or for whom a parent interview was completed in the base year were eligible to be assessed in the spring-third grade data collection. Eligibility for the study was not dependent on the child's current grade, that is, children were eligible whether they were promoted to third grade or were retained in second grade.

Table 5-2. Results of the Telephone Research Center's locating efforts, spring–third grade: School year 2001–02

Result	Number	Percent
Total cases worked	781	100.0
Located and entered into database	307	39.3
Unlocatable	426	54.5
Out of scope	27	3.5
Final refusal	16	2.0
Partially located	3	0.4
Unable to locate due to language barrier	2	0.3

NOTE: "Unlocatable" means that the children and their households could not be found using the available tracing and locating strategies; "out of scope" means that a child was ineligible to participate; "final refusal" means that the child's family indicated that they did not want to participate; "partially located" means that the tracing and locating effort yielded some information about the child, but not enough to definitively locate the child; "unable to locate due to language barriers" means that the household language was not English and no staff were available who were bilingual in that language.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

As in previous rounds of data collection, the field staff were organized into work areas, each with a data collection team consisting of one field supervisor and three or more assessors. The data collection teams were responsible for all data collection activities in their work areas; they conducted the direct child assessments and the parent interviews, distributed and collected all school and teacher questionnaires, and completed school facilities checklists. The majority of field staff members in third grade were continuing from previous rounds of data collection; a few new staff were hired in areas where no experienced ECLS-K staff lived.

5.4.1 Preassessment School Contact

Based on the information collected in the fall of 2001, packets of hard-copy teacher and school administrator questionnaires were assembled and mailed to schools in February 2002, along with letters confirming the scheduled visits to the school. Teachers and school administrators were asked to either complete the questionnaires for pickup on assessment day, or to return the questionnaires in a Federal Express mailer that was provided in the packet. Letters were also mailed to parents reminding them of the spring-third grade data collection activities.

Field supervisors conducted most preassessment activities by telephone starting in March 2002. The preassessment activities for these schools were similar to those conducted in previous rounds of data collection and included confirming the assessment date and the receipt of the hard-copy questionnaires and arranging for space to conduct the assessments.

5.4.2 Timeline of the Direct Child Assessments

The direct child assessments were conducted from March through June 2002, the same time of year as in prior spring data collections. Conducting the child assessments began in March with 91 percent of the assessments completed between April and May and a small percentage (9 percent) completed in June. In year-round schools, assessment teams made multiple visits to the school, visiting when each track was in session to assess the sampled children.

5.4.2.1 Conducting the Direct Child Assessments

The direct child assessments were usually conducted in a school classroom or library. Before conducting the assessments, field supervisors and assessors set up the room for the assessments. They followed procedures for meeting children that were agreed upon during the preassessment contact with the school. Each child was signed out of his or her classroom prior to the assessments and signed back into the classroom upon the conclusion of the assessments. When scheduling schools in the fall, an attempt was made to conduct the direct child assessments at about the same point in time from the beginning of school year and at the end of the year to increase the chances that exposure to instruction was about the same for all children. The third grade direct child assessments averaged 94 minutes.

Table 5-3 displays the number of completed child assessments for each round of data collection, including spring-third grade. All of the assessments in spring-third grade were completed in English. Most (74.6 percent) of these assessments were completed in original schools, although the number of assessments in transfer schools has grown at each data collection point. In spring-third grade, a quarter of the sample was assessed in transfer schools.

Table 5-3. Completed child assessments by round of data collection and selected characteristics: School years 1998–99, 1999–2000, and 2001–02

Characteristic	Fall-kindergarten		Spring-kindergarten		Fall-first grade		Spring-first grade		Spring-third grade	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Child assessments completed	19,147	100.0	19,987	100.0	5,297	100.0	16,622	100.0	14,502	100.0
Assessed										
In English, no accommodation[1]	17,019	88.9	18,342	91.8	4,848	91.5	15,460	93.0	13,565	93.5
In Spanish	1,008	5.3	724	3.6	176	3.3	286	1.7	†(a)	†(a)
In other language	410	2.1	229	1.1	33	0.6	37	0.2	†(a)	†(a)
With accommodation[1]	515	2.7	579	2.9	195	3.7	761	4.6	814	5.6
Excluded	88	0.5	70	0.4	28	0.5	47	0.3	74	0.5
Partial complete	107	0.6	43	0.2	17	0.3	31	0.2	49	0.3
Child assessments completed	19,147	100.0	19,987	100.0	5,297	100.0	16,622	100.0	14,502	100.0
Original sampled school	19,147	100.0	19,463	97.4	4,867	91.9	14,830	89.2	10,820	74.6
Transfer school	†(b)	†(b)	524	2.6	430	8.1	1,792	10.8	3,682	25.4

†(a) Not applicable. The assessment was conducted only in English in third grade.
†(b) Not applicable. There were no transfer schools in fall-kindergarten.
[1]The term *accommodation* in this table is the field operational definition of accommodation, which includes the wearing of glasses and hearing aids. These types of aids were systematically tracked to ensure that every child had the same chance at a successful assessment. With this information, assessors could prompt a child, for example, to get her glasses before being assessed.
NOTE: This table reflects final production numbers prior to statistical adjustment. This table does not include children who were subsampled out in fall- and spring-first grade and spring-third grade (see section 5.4.4.) These numbers should not be used to estimate student mobility.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first, and third grade data collections, school years 1998–99, 1999–2000, and 2001–02.

5.4.2.2 Accommodations and Exclusions

Approximately 1 percent of participating children in third grade required accommodations or were excluded from the direct child assessments. Children were excluded from the direct assessments because of a disability, e.g., blindness or deafness, that was not accommodated by the ECLS-K direct assessments or their Individualized Education Plan prevented their participation in assessments or required an accommodation not offered in the assessments. Accommodations offered in the assessments were as follows: alternative setting, scheduling, or timing; health care aide present; or the use of an assistive device. Table 5-4 presents the number of children excluded from and requiring an accommodation to the direct child assessment procedures in the spring of third grade.

Table 5-4. Number of children excluded from and requiring an accommodation in the spring-third grade assessments: School year 2001–02

Category	Number of children
Exclusions	
Excluded for disability	74
Accommodations[1]	
Alternative setting accommodation	33
Scheduling/timing accommodation	65
Health care aide present	6
Assistive device	4

[1]The term *accommodation* in this table includes only those accommodations offered during the assessment, such as an alternative setting.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

5.4.3 Conducting the Parent Interview

Parent interview procedures mirrored those of the base year and first grade. The parent interview was administered, primarily by telephone interview using CAI, between March and July 2002.

Sixteen percent of the parent interviews were completed in March, 54 percent were completed in April and May, and 30 percent were completed in June or later. The parent interview averaged 62 minutes. As in previous rounds of data collection, the parent interview was conducted in person if the respondent did not have a telephone. Table 5-5 contains the number of parent interviews per round, including spring-third grade. In third grade, only 2.4 percent of all completed parent interviews were conducted in person; 7.2 percent of all completed parent interviews were conducted in a language other than English with 95.8 percent of completed non-English interviews conducted in Spanish. A special effort to build parent response rates was conducted between July 5 and 31, 2002 and yielded an additional 7.3 percentage points in the response rate. Almost 8 percent (7.7 percent) of the parent interviews were not completed because of locating problems.

5.4.4 Conducting Data Collection on Children Who Withdrew From Their Previous Round School

While contacting schools, field supervisors asked school coordinators to identify children who had withdrawn from the school since the spring of first grade. School staff were asked whether they knew the name and address of the school to which the child transferred, as well as any new information about the child's household address. For the children who had moved from their spring-first grade school and were not part of the sample to be followed, information was collected only from the school personnel and not parents. For children who had withdrawn from their spring-first grade school and were identified to be followed (i.e., were part of the sample of movers), supervisors also consulted parents and other contacts for information on the children's new school. This information was entered into the FMS and processed at Westat for data collection.

Table 5-6 presents the status of the children who were identified as movers in third grade; 9,889 children were identified as having transferred from the school in which they were enrolled during the spring of first grade. Of the 9,889 children who moved in spring-third grade, 5,668 were in scope, i.e., children selected to be followed, and followed (57.3 percent of total movers). The remaining 4,221 mover children were out of scope and were not followed; no child assessments or parent interview was conducted for these children.

Table 5-5. Number and percent of completed parent interviews by data collection mode, language, and wave of data collection: School years 1998–99, 1999–2000, and 2001–02

Data collection modes and language	Fall-kindergarten		Spring-kindergarten		Fall-first grade		Spring-first grade		Spring-third grade	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Parent interviews completed	17,997	100.0	18,907	100.0	5,073	100.0	15,576	100.0	13,504	100.0
In person	618	3.4	619	3.3	211	4.2	456	2.9	319	2.4
By phone	17,379	96.6	18,288	96.7	4,862	95.8	15,120	97.1	13,185	97.6
Parent interviews completed	17,997	100.0	18,907	100.0	5,073	100.0	15,576	100.0	13,504	100.0
In English	17,379	96.6	17,482	92.5	4,717	93.0	14,319	91.9	12,416	91.9
In Spanish	618	3.4	1,321	7.0	351	6.9	1,071	6.9	932	6.9
In other language	0	0	81	0.4	0	0	75	0.5	41	0.3
Partial complete	0	0	23	0.1	5	0.1	111	0.7	115	0.9

NOTE: This table reflects final production numbers prior to statistical adjustment.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 kindergarten, first, and third grade data collections, school years 1998–2002.

Table 5-6. Number and percent of spring-third grade children who moved from their spring-first grade school, by scope and completion category: School year 2001–02

Scope and completion category	Spring-third grade	
	Number of children	Percent
Total movers	9,889	100.0
Out-of-scope[1]	4,221	42.7
Did not follow[2]	4,102	97.2
Moved to outside of U.S.[2]	117	2.7
Deceased[2]	2	<1
In-scope and followed[1]	5,668	57.3
Completed assessment[3]	3,682	65.0
Unlocatable[3]	607	10.7
Nonsampled primary sample unit[3]	871	15.4
Assessment refused[3]	323	5.7
Not assessed/absent[3]	185	3.3

[1] Percent based on total movers.
[2] Percent based on out-of-scope children.
[3] Percent based on in-scope children.
NOTE: Detail may not sum to totals due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Different data collection strategies were followed for children who moved, depending on where they moved to and the status of their new school. Data collection was attempted for children who moved and were flagged as "follow" in spring-third grade in the following ways:

- Data collected for children moving into cooperating base year sampled schools included the child assessments in the school, school administrator questionnaire, regular and/or special education teacher questionnaires, facilities checklist, and student record abstract forms.
- Data collected for children moving into nonsampled schools in base year cooperating districts included the child assessments in the school, school administrator questionnaires, regular and/or special education teacher questionnaires, and student record abstract forms, if school permission was obtained. If school permission was not obtained, the assessments were conducted in the home and no school or teacher data were collected. Parent interviews were attempted for all children.
- For children moving into transfer schools that refused, schools in sampled districts that refused, or originally sampled schools that were ineligible when sampled because they did not have kindergarten classes, the direct child assessments were conducted in the home. No school or teacher data were collected. Parent interviews were attempted for all children.

- For children moving into schools in nonsampled districts or dioceses:
  - If the school was within the primary sampling unit (PSU), data collected included the child assessments in the school, school administrator questionnaire, regular and/or special education teacher questionnaires, facilities checklist, and student record abstract forms, if school permission was obtained. If school permission was not obtained, the assessments were conducted in the home and no school or teacher data were collected. Parent interviews were attempted for all children.
  - If the school was outside the PSU, no child, school, or teacher data were collected. The parent interview was still attempted.
- For children who were not enrolled in school in the spring (including children who were home schooled), data collected included the child assessments in the home if the child was in the sampled PSU. If the child was outside the sampled PSU, no child assessment or school or teacher data were collected. Parent interviews were attempted for all children.

Of the children who were identified as movers in third grade and who were selected to be followed, 15.4 percent moved into a school outside the PSU, and 10.7 percent could not be located. Assessments were completed for 65 percent of the movers who were followed in the spring-third grade data collection, and parent interviews were completed for 68 percent of these children.

5.4.5 Teacher and School Data Collection

Data were collected from school administrators, regular classroom teachers, and special education teachers from March through June 2002.

The school and teacher questionnaires were mailed to the school coordinators in February 2002. Using the child-teacher linkage information collected in the fall, a packet of questionnaires was assembled for each regular and special education teacher. The regular teacher packet included a cover letter, a sheet explaining the study and its goals, and teacher questionnaire part A, teacher questionnaire part B, and teacher questionnaire part C for each student who had been linked to the teacher in the fall. The special education teacher packet contained a cover letter and summary sheet, special education teacher questionnaire part A, and special education teacher questionnaire part B for each sampled student linked to the teacher. Packets were bundled together by school and mailed to the school coordinator for distribution. If the school and/or teacher and school administrator were not identified in the fall advance

contact, then the supervisor gathered the relevant information during the preassessment call in the spring and mailed the packets at that time.

Teachers were asked to complete individual ratings for the sampled children in their classrooms, and they were paid $7 for each child rating (teacher questionnaire part C) they completed. In addition, school staff were asked to complete a student record abstract after the school year closed and were paid $7 for every student record abstract completed. Field supervisors also completed a facilities checklist for each sampled school.

During the field period, field supervisors followed up with school administrators and teachers by telephone and visits to the schools to ensure that questionnaires were not missing critical information and that completed questionnaires were mailed to Westat. To improve response rates, in early September 2002 a package was mailed to all schools with outstanding school fact sheets or student record abstracts with a request to complete and return questionnaires. School staff were prompted by telephone for the return of the questionnaires and abstracts through October 2002. The hard-copy followup increased child-level responses rates for the school fact sheet by 10 percent and the student record abstract by 12 percent.

5.5 Data Collection Quality Control

Continuous quality assurance procedures were employed during all data collection activities, but with a particular focus on the assessments. The procedures were incorporated throughout all stages of the study (e.g., during instrument development, in the staff training program, through assessment certification, and as part of the ongoing staff observations and evaluation activities).

Data collection quality control efforts began with the additional development and testing of redesigned sections of the CAI/CAPI applications and the FMS. As sections of these applications were re-programmed, extensive testing of the entire system was conducted to verify that the systems were working properly from all perspectives. This testing included review by project design staff, statistical staff, and the programmers themselves. Quality control processes continued with the development of field procedures that maximized cooperation and thereby reduced the potential for nonresponse bias.

Quality control activities continued during training and data collection. During assessor training, field staff practiced conducting the parent interview in pairs and practiced the direct child assessments with third grade-aged children brought to the training site for this purpose. The supervisors and assessors were certified on the child assessments using the Training Certification Form. When the fieldwork began, field supervisors observed each assessor conducting child assessments and made telephone calls to parents to validate the interview. Field managers made telephone calls to the schools to collect information on the school activities for validation purposes.

5.5.1 Child Assessments Observations

Field supervisors conducted on-site observations of the child assessments and completed the child observation form. Our quality control plan called for conducting two observations for each of the 266 assessors who completed training. The first observation was to be within 2 weeks of the start of the assessments, and the second observation within 3 weeks of the first observation. These procedures were followed for the majority of assessors (over 80 percent), but some assessors were observed only once due to the school year ending or to the travel distance involved.

A standardized observation form was used to evaluate the assessor's performance in conducting the child assessments. The assessor was rated in three areas:

1. Rapport building and working with the child—use of neutral praise and the assessor's response to various child behaviors.

2. Cognitive assessment activities—reading questions verbatim, the use of acceptable probes, the use of appropriate hand motions, and the absence of coaching.

3. Specific assessment activities—correctly coding answers to open-ended questions in the assessments and following administration procedures.

The field supervisors recorded their observations on the form and then reviewed the form with the assessor. The most frequent problems observed were not reading the items verbatim and inappropriate gesturing. Feedback was provided to the assessors on the strengths and weaknesses of their performance and, when necessary, remedial training was provided in areas of weakness. Table 5-7 presents the result of the observations.

Table 5-7. Results of the child assessments observations, spring-third grade: School year 2001-02

Observations[1]	Number	Percent
Total	468	100
Score on certification form		
85 percent or above	468	100
70–84 percent	0	0
Below 70 percent	0	0

[1] A total of 258 assessors were initially observed. (Eight trainees chose not to continue with data collection.) Of these, 200 received a second observation visit and 10 received a third observation visit. No assessor scored below 85 percent on any of the observation visits.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

5.5.2 Validation of Parent Interviews

Approximately 10 percent of the completed parent interviews were called back by a field supervisor (i.e., validated). The first parent interview completed by an assessor was always validated. Over the course of the field period, a running count of an assessor's completed parent interviews was maintained, and each tenth completed parent interview was selected for validation, thus ensuring that 10 percent of each assessor's cases were selected for validation. The parent validation was approximately 5 minutes long and was conducted by telephone.

Field supervisors used a standardized parent validation script to make validation calls to parents. The script covered the following topics:

- Verification of the child's name, date of birth, and sex; and
- Eight to ten questions from the parent interview that were re-asked of the parent.

5.5.3 Validations of School Visits

To ensure that assessments proceeded smoothly, a validation call was completed with the school principal in at least two of each supervisor's assigned schools in the spring-third grade data collection.

Field managers conducted the school validations by telephone. The first school that each team completed was called to ascertain how well the preassessment and assessment activities went. If the feedback from the school was positive, the fifth school that each team completed was called. If any problems were indicated in the first validation call, immediate action was taken with the field supervisor. The validation feedback was discussed with the supervisor and remedial action was taken, including in-person observation of the supervisor's next school, if necessary. In spring-third grade, a total of 155 school visits were validated with no negative reports of the assessment team or study made by school staff.

Field managers used a standardized script to call the school principals. The script covered the following topics:

- An overall rating of how the assessments went;
- Feedback about the study from the children and teachers;
- Suggestions for improving procedures and making it easier for a school to participate and;
- General comments and suggestions.

5.5.4 Assessor Interrater Reliability

As part of the child assessments observation described in section 5.5.1, field supervisors completed an assessment certification form for each observation they conducted. An important element of this form was the "validation items." With the exception of the reading routing test, all of the assessments included at least one item that both the observer and the assessor scored. These items had open-ended responses that called for interpretation on the part of the assessor to determine whether a child's response was correct. By comparing the extent to which assessors and observers agreed on scoring these validation items, a measure of interrater reliability was obtained. Interrater reliability provided a measure of the accuracy of the assessor's scoring compared with the standard, the observer's.

Table 5-8 contains the results of these comparisons. As can be seen, overall interrater reliability is very high throughout all the levels. It is highest for math (97 percent or better on all levels) and lowest for reading, with the reading yellow level (the medium reading level) showing

Table 5-8. Interrater reliability on child assessment validation item, by subject area and level: School year 2001–02

Subject and level	Number of observations	Number of validation items	Percent agreement: assessors and observers[1]
Reading	462	13	90
Routing	†	0	†
Low (Red)	118	3	93
Middle (Yellow)	262	5	87
High (Blue)	82	5	94
Mathematics	464	7	99
Routing	464	1	100
Low (Red)	135	1	99
Middle (Yellow)	186	2	99
High (Blue)	143	3	97
Science	453	12	93
Routing	453	3	95
Low (Red)	140	1	100
Middle (Yellow)	228	4	89
High (Blue)	85	4	93

† Not applicable.
[1] Percent agreement was calculated as follows: number of validation items observed in which observer agreed with the assessor/number of validation items observed.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

the lowest percent agreement (87 percent). The reading yellow level path received a relatively large number of observations (262) and also contained a relatively large number of validation items (5) compared with some of the other paths. Thus, there was greater opportunity for disagreement on this path compared with the others. The science yellow level (the medium science level) also had a relatively higher opportunity for disagreement (228 observations and 4 validation codes) and it, too, exhibited a somewhat lower interrater reliability (89 percent) compared with some of the other paths. The reliability, however, even on these more difficult paths, was high and demonstrated that the assessors accurately coded open-ended items. More details on the instruments can be found in the *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) Psychometric Report for the Third Grade* (U.S Department of Education, National Center for Education Statistics, forthcoming [b]).

5.6 Spring-Third Grade Completion Rates

In the sections that follow, spring-third grade completion rates are presented for three groups of students: (1) students sampled in kindergarten, (2) students sampled in first grade through the freshening procedure, and (3) both groups combined. Completion rates were computed with the same procedures used for spring-first grade to allow for comparisons of completion rates between the two years of data collection following the base year. For spring-first grade and spring-third grade, the sample of children is the same: base year respondents (i.e., children who had either a fall- or spring-kindergarten child assessments or parent interview) and children sampled in spring-first grade as part of sample freshening as described in section 4.3.2.

5.6.1 Students Sampled in Kindergarten

Tables 5-9 to 5-11 present weighted and unweighted child-level completion rates for spring-third grade data collection, broken out by school characteristics. These rates pertain to children who were sampled as part of the kindergarten cohort in the base year. (Rates for students sampled in first grade through the student sample freshening procedure can be found in table 5-19.) For the ECLS-K, a completion rate is a response rate conditioned on the results of an earlier stage of data collection. For the group of children sampled in kindergarten, all completion rates are conditioned on the case having been a base year respondent.

Relative to spring-first grade, the overall completion rates for the child assessments (80.8 percent) and the parent interview (77.8 percent) both decreased 7 percentage points in spring-third grade. The decrease is almost completely due to the increase in the number of children who moved outside of the sampled PSUs or moved within the sampled PSUs but could not be located (the numbers slightly more than doubled in both cases). These children are included in the category labeled “Unknown” for each of the different school characteristics (tables 5-9 and 5-11). The category includes children who were unlocatable as their whereabouts were unknown, and those children who had moved into a nonsampled county. If no information concerning the child’s school was available, they were included in the “Unknown” category. The completion rates for the child assessments are quite high and uniform across school characteristics (ranging from 97.1 percent in schools with 750 or more enrolled to 100 percent in

Table 5-9. Number of completed child-level cases and child-level completion rates among 1998–99 kindergartners in spring-third grade, by survey type and school characteristics: School year 2001–02

School characteristic[1]	Child assessments			Parent interview		
		Completion rate			Completion rate	
	Completes[2]	Weighted	Unweighted	Completes[3]	Weighted	Unweighted
All school types	14,349	80.8	86.1	13,392	77.8	80.3
School type						
Public	11,671	98.1	98.5	10,268	86.3	86.6
Private	2,629	98.7	99.2	2,452	92.6	92.5
Catholic	1,662	99.0	99.3	1,546	92.9	92.4
Other private	967	98.1	99.0	906	92.3	92.7
Unknown school type	49	2.6	2.3	672	36.2	31.0
Type of locale						
Large city	2,431	97.8	98.3	2,038	80.9	82.4
Mid-size city	2,474	98.9	99.3	2,216	88.6	88.9
Urban fringe of large city	4,169	97.4	98.0	3,723	86.9	87.5
Urban fringe of mid-size city	942	97.7	98.0	832	86.2	86.6
Large town	375	100.0	100.0	355	93.4	94.7
Small town	1,032	99.4	99.2	967	90.0	93.0
Rural—outside MSA	1,498	98.8	99.4	1,350	89.7	89.6
Rural—inside MSA	1,201	98.2	98.7	1,075	88.0	88.3
Unknown	227	8.5	9.7	836	39.5	35.6
School size (total enrollment)						
1 to 299	3,078	98.2	98.8	2,809	89.8	90.2
300 to 499	4,562	98.5	98.9	4,118	88.5	89.2
500 to 749	4,043	98.5	98.7	3,569	87.3	87.2
750 or more	2,505	97.1	97.9	2,116	81.4	82.7
Unknown	161	6.2	7.0	780	38.3	34.1

See notes at end of table.

Table 5-9. Number of completed child-level cases and child-level completion rates among 1998–99 kindergartners in spring-third grade, by survey type and school characteristics: School year 2001–02—Continued

School characteristic[1]	Child assessment			Parent interview		
		Completion rate			Completion rate	
	Completes[2]	Weighted	Unweighted	Completes[3]	Weighted	Unweighted
Percent non-White enrolled						
0 – 10	4,580	99.3	99.4	4,269	92.3	92.7
11 – 49	4,594	97.5	98.2	4,179	88.3	89.4
50 – 89	2,564	98.0	98.4	2,197	83.8	84.3
90 – 100	2,412	97.8	98.2	1,932	78.5	78.6
Unknown	199	7.3	8.6	815	38.9	35.1
Region						
Northeast	2,667	98.6	98.8	2,411	88.8	89.3
Midwest	3,677	98.1	98.5	3,405	90.8	91.2
South	4,674	97.6	98.3	4,038	83.5	85.0
West	3,291	98.6	98.8	2,880	87.2	86.5
Unknown	40	2.2	1.9	658	35.8	30.5

[1] Based on ECLS-K survey data and not data from the sampling frame.
[2] Reading, math, or science assessment was scorable or child was disabled.
[3] Family structure portion of parent interview was completed.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 5-10. Number of completed child-level cases and child-level completion rates among 1998–99 kindergartners in spring-third grade, by survey type and school characteristics: School year 2001–02

School characteristic[1]	School administrator questionnaire			School fact sheet		
		Completion rate			Completion rate	
	Completes[2]	Weighted	Unweighted	Completes[2]	Weighted	Unweighted
All school types	12,361	66.1	73.3	14,064	76.5	83.4
School type						
Public	9,559	78.4	80.7	10,976	91.1	92.6
Private	2,352	85.8	88.7	2,545	94.6	96.0
Catholic	1,515	88.5	90.5	1,623	95.9	97.0
Other private	837	81.4	85.7	922	92.5	94.4
Unknown school type	450	12.0	19.0	543	14.5	22.9
Type of locale						
Large city	1,832	69.0	74.1	2,246	88.6	90.8
Mid-size city	2,138	82.2	85.8	2,349	93.1	94.3
Urban fringe of large city	3,271	73.9	76.9	3,894	89.7	91.6
Urban fringe of mid-size city	792	78.0	82.4	903	91.0	94.0
Large town	374	99.4	99.7	375	100.0	100.0
Small town	933	88.0	89.7	1,015	97.2	97.6
Rural – outside MSA	1,401	93.0	93.0	1,480	97.7	98.2
Rural – inside MSA	1,053	81.7	86.5	1,132	89.9	93.0
Unknown	567	15.3	22.2	670	17.4	26.3
School size (total enrollment)						
1 to 299	2,883	90.0	92.6	3,024	95.9	97.1
300 to 499	3,934	81.9	85.2	4,328	92.4	93.8
500 to 749	3,234	76.5	79.0	3,807	90.4	93.0
750 or more	1,859	72.8	72.6	2,320	90.2	90.6
Unknown	451	11.6	18.1	585	15.5	23.5

See notes at end of table.

Table 5-10. Number of completed child-level cases and child-level completion rates among 1998–99 kindergartners in spring-third grade, by survey type and school characteristics: School year 2001–02—Continued

School characteristic[1]	School administrator questionnaire			School fact sheet		
		Completion rate			Completion rate	
	Completes[2]	Weighted	Unweighted	Completes[2]	Weighted	Unweighted
Percent non-White enrolled						
0 – 10	4,017	85.2	87.2	4,412	94.6	95.8
11 – 49	3,973	82.1	85.0	4,379	91.1	93.6
50 – 89	1,970	73.1	75.6	2,389	90.6	91.6
90 – 100	1,921	74.0	78.2	2,265	90.9	92.2
Unknown	480	12.2	19.0	619	16.2	24.5
Region						
Northeast	2,035	72.8	75.4	2,460	89.5	91.1
Midwest	3,349	86.9	89.7	3,575	94.4	95.7
South	3,864	79.0	81.3	4,454	91.4	93.7
West	2,668	76.0	80.1	3,039	89.9	91.3
Unknown	445	11.8	18.9	536	14.4	22.8

[1] Based on ECLS-K survey data and not data from the sampling frame.

[2] A completed questionnaire was defined as one that was not completely left blank.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 5-11. Number of completed child-level cases and child-level completion rates among 1998–99 kindergartners in spring-third grade, by survey type and school characteristics: School year 2001–02

School characteristic[1]	Teacher questionnaire part A			Teacher questionnaire part B			Teacher questionnaire part C		
		Completion rate			Completion rate			Completion rate	
	Completes[2]	Weighted	Unweighted	Completes[2]	Weighted	Unweighted	Completes[2]	Weighted	Unweighted
All school types	11,770	62.4	69.8	11,741	62.3	69.6	11,802	62.7	70.0
School type									
Public	9,274	75.7	78.3	9,242	75.5	78.0	9,301	76.0	78.5
Private	2,416	87.3	91.1	2,418	87.2	91.2	2,425	87.4	91.5
Catholic	1,557	90.2	93.0	1,564	90.5	93.4	1,575	91.4	94.1
Other private	859	82.7	87.9	854	82.0	87.4	850	81.3	87.0
Unknown school type	80	2.2	3.4	81	2.3	3.4	76	2.2	3.2
Type of locale									
Large city	1,649	62.0	66.7	1,647	62.0	66.6	1,654	62.4	66.9
Mid-size city	2,087	80.0	83.7	2,079	79.5	83.4	2,096	80.2	84.1
Urban fringe of large city	3,322	74.5	78.1	3,295	73.9	77.5	3,329	74.6	78.3
Urban fringe of mid-size city	801	78.1	83.4	804	78.9	83.7	808	79.5	84.1
Large town	363	96.0	96.8	370	98.2	98.7	368	97.7	98.1
Small town	985	92.9	94.7	982	92.8	94.4	977	92.1	93.9
Rural—outside MSA	1,356	85.3	90.0	1,363	85.8	90.4	1,369	86.6	90.8
Rural—inside MSA	1,027	80.1	84.4	1,020	79.4	83.8	1,023	79.5	84.1
Unknown	180	5.4	7.1	181	5.5	7.1	178	5.4	7.0
School size (total enrollment)									
1 to 299	2,832	87.9	90.9	2,817	87.3	90.5	2,829	87.8	90.8
300 to 499	3,911	81.2	84.7	3,920	81.4	84.9	3,924	81.5	85.0
500 to 749	3,181	75.3	77.7	3,164	75.0	77.2	3,199	75.8	78.1
750 or more	1,730	65.1	67.6	1,723	65.0	67.3	1,735	65.6	67.8
Unknown	116	3.2	4.7	117	3.3	4.7	115	3.3	4.6

See notes at end of table.

Table 5-11. Number of completed child-level cases and child-level completion rates among 1998–99 kindergartners in spring-third grade, by survey type and school characteristics: School year 2001–02—Continued

School characteristic[1]	Teacher questionnaire part A			Teacher questionnaire part B			Teacher questionnaire part C		
		Completion rate			Completion rate			Completion rate	
	Completes[2]	Weighted	Unweighted	Completes[2]	Weighted	Unweighted	Completes[2]	Weighted	Unweighted
Percent non-White enrolled									
0 – 10	4,216	88.2	91.5	4,206	87.9	91.3	4,211	88.0	91.4
11 – 49	3,839	77.9	82.1	3,835	77.9	82.0	3,872	78.6	82.8
50 – 89	1,963	73.4	75.3	1,951	73.2	74.8	1,960	73.9	75.2
90 – 100	1,605	62.2	65.3	1,601	62.0	65.2	1,611	62.3	65.6
Unknown	147	4.0	5.8	148	4.0	5.9	148	4.1	5.9
Region									
Northeast	2,168	77.1	80.3	2,155	76.7	79.8	2,170	77.8	80.4
Midwest	3,272	85.0	87.6	3,275	85.1	87.7	3,292	85.3	88.2
South	3,822	76.2	80.4	3,815	75.9	80.3	3,842	76.6	80.8
West	2,433	69.2	73.1	2,420	68.9	72.7	2,427	69.1	72.9
Unknown	75	2.0	3.2	76	2.1	3.2	71	2.0	3.0

[1] Based on ECLS-K survey data and not data from the sampling frame.

[2] A completed questionnaire was defined as one that was not completely left blank.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

schools in large towns), except for the "unknown" category. Similarly, the completion rates for the parent interviews were uniform across school characteristics (ranging from 80.9 percent for children in large cities to 93.4 percent for children in large towns), except for the "unknown" category. The "unknown" category aside, both the child assessments and the parent interview completion rates increased between spring-first grade and spring-third grade for all school characteristics. The completion rates by mover status are discussed later, but the rates of completing all the instruments are much lower for children who moved than for those who did not move.

Table 5-10 shows that the overall weighted completion rate is 66.1 percent for the school administrator questionnaire and 76.5 percent for the school fact sheet. The completion rate for the school fact sheet is about 10 percent higher than that for the school administrator questionnaire due to the continued data collection in fall 2002 that affected the school fact sheet and the student record abstract. The completion rate for the school administrator questionnaire is about 10 percentage points lower than the corresponding rate in spring-first grade. Note that there was no school fact sheet in spring-first grade. Once again, the increase in the movers is largely responsible for the lower rates, as discussed below. The completion rates for the school administrator questionnaire range from 69.0 percent for children in large cities to 99.4 percent for those in large towns (ignoring the "unknown" category). Rates for the school fact sheet follow the same pattern. Excluding the movers, the completion rate for the school administrator questionnaire is much higher, with an overall rate of 86.7 percent, only slightly lower than the spring-first grade rate of 88.7 percent. In the case of the school fact sheet, the rate for nonmovers is 95.8 percent. It is worth noting that the completion rates for the school administrator questionnaire are lower for schools with higher percentages of minorities, a phenomenon also observed in previous rounds for the school administrator questionnaire. However, this disparity decreased considerably in the first-grade year and in spring-third grade compared to the base year, reflecting the success of increased data collection efforts targeted toward these schools.

All three of the teacher questionnaires were completed at an overall rate of 62 to 63 percent. The completion rates have substantial variation when broken out by school characteristics, even when the "unknown" category is ignored. The completion rates are 90 percent or more for Catholic schools and for schools in large and small towns. Schools in large cities, schools with 750 students or more, and schools with 90 percent or more minority enrollment have completion rates in the 60s. The "unknown" categories have, by far, the lowest completion rates.

As noted above, the rate at which the survey instruments were completed varies markedly by mover status and within movers, by whether or not the child was located and followed. As shown in table 5-12 the completion rate for the child assessments was 94.5 percent for children still enrolled in their base year school. For movers it dropped by close to 9 points to 85.6 percent for those who were located and followed, and for those not located or followed due to a move to a non-ECLS-K PSU, it was zero. The parent interview completion rates varied from 84.5 percent for nonmovers to 74.8 percent for movers who were located and followed for the purposes of the child assessments, to 51.2 percent for movers who could either not be located or were not followed for the purposes of the child assessments. Even though children who had moved to a non-ECLS-K PSU were not administered the child assessments, a parent interview was conducted by telephone wherever possible, leading to the 51.2 percent response rate for this category.

Table 5-13 shows that the school administrator questionnaire completion rate is about 30 points lower for movers compared to nonmovers, even when the children who had moved were located and followed. For the school fact sheet, it is about 20 points lower for movers than for nonmovers. There are several reasons for this difference: located movers were not always assessed in schools; new schools in which movers enrolled had a lower level of commitment to the ECLS-K and often refused to complete the school administrator questionnaire; and some of these schools were contacted too late in the school year for them to consider completing it. The completion rate for nonmovers was 86.7 percent for the school administrator questionnaire and 95.8 percent for the school fact sheet. For located and followed movers it was 56.0 and 74.5 percent for the school administrator questionnaire and for the school fact sheet, respectively.

For all three teacher questionnaires the completion rates were approximately 82 percent if the child had not moved; about 54 percent if the child moved, was located, and followed; and just about 0 percent if not located or followed (table 5-14). A handful of children who could not be located but had teacher data was due to the fact that, if they move during the term, and teachers in their old schools already filled out the questionnaires but teachers in the new schools did not, the teacher data from the old schools were attached to these children. The reasons for lower completion rates from teachers if the child moved are similar to the reasons that affected the school administrator questionnaire and school fact sheet completion rates for movers.

Table 5-12. Number of completed child-level cases and child-level completion rates among 1998–99 kindergartners in spring-third grade, by survey type and child's mover status: School year 2001–02

Mover status[1]	Child assessments			Parent interview		
		Completion rate			Completion rate	
	Completes[2]	Weighted	Unweighted	Completes[3]	Weighted	Unweighted
All students	14,349	80.8	86.1	13,392	77.8	80.3
Mover status						
Mover	2,791	60.9	62.6	3,063	68.0	68.8
Located, followed	2,791	85.6	85.1	2,451	74.8	74.7
Other[4]	0	0.0	0.0	612	51.2	52.1
Nonmover	11,558	94.5	94.6	10,329	84.5	84.6

[1] This is the mover status used in weighting, which does not consider children who moved into identified destination schools as movers.
[2] Reading, math or science assessment was scorable or child was disabled.
[3] Family structure portion of parent interview was completed.
[4] This category includes movers who could not be located, and movers who moved into nonsampled PSUs.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 5-13. Number of completed child-level cases and child-level completion rates among 1998–99 kindergartners in spring-third grade by survey type and child's mover status: School year 2001–02

Mover status[1]	School administrator questionnaire			School fact sheet		
		Completion rate			Completion rate	
	Completes[2]	Weighted	Unweighted	Completes[2]	Weighted	Unweighted
All students	12,361	66.1	73.3	14,064	76.5	83.4
Mover status						
Mover	1,739	37.2	37.4	2,327	49.6	50.0
Located, followed	1,739	56.0	54.5	2,327	74.5	72.9
Other[3]	0	0.0	0.0	0	0.0	0.0
Nonmover	10,622	86.7	87.0	11,737	95.8	96.1

[1] This is the mover status used in weighting, which does not consider children who moved into identified destination schools as movers.
[2] A completed questionnaire was defined as one that was not completely left blank.
[3] This category includes movers who could not be located, and movers who moved into nonsampled PSUs.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 5-14. Number of completed child-level cases and child-level completion rates among 1998–99 kindergartners in spring-third grade, by survey type and child's mover status: School year 2001–02

Mover status[1]	Teacher questionnaire part A			Teacher questionnaire part B			Teacher questionnaire part C		
		Completion rate			Completion rate			Completion rate	
	Completes[2]	Weighted	Unweighted	Completes[2]	Weighted	Unweighted	Completes[2]	Weighted	Unweighted
All students	11,770	62.4	69.8	11,741	62.3	69.6	11,802	62.7	70.0
Mover status									
Mover	1,656	35.7	35.6	1,645	35.6	35.3	1,650	35.7	35.5
Located, followed	1,653	53.7	51.8	1,641	53.4	51.4	1,646	53.7	51.6
Other[3]	3	0.1	0.2	4	0.2	0.3	4	0.2	0.3
Nonmover	10,114	81.5	82.8	10,096	81.4	82.7	10,152	81.9	83.1

[1] This is the mover status used in weighting, which does not consider children who moved into identified destination schools as movers.
[2] A completed questionnaire was defined as one that was not completely left blank.
[3] This category includes movers who could not be located, and movers into nonsampled PSUs.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Tables 5-15 to 5-17 present child-level completion rates for the spring-third grade data collection, this time broken out by child characteristics for children who were sampled as part of the kindergarten cohort in the base year. When the "unknown" categories are not included, the differences in completion rates by sex and by year of birth are very small, but for race and ethnicity they are more substantial. For the child assessments the completion rate was highest for Asians and Pacific Islanders (84.1 percent and 84.9 percent, respectively) and lowest for American Indians or Alaska Natives (75.5 percent). For the parent interview it was highest for Whites (82.9 percent), and lowest for Blacks (67.0 percent) and Asian students (68.6 percent). The ECLS-K sample of Pacific Islanders is very clustered and has unusually high completion rates for the instruments filled out by school personnel, 80.9 percent for the school administrator questionnaire, 84.6 percent for the school fact sheet, and about 74 percent for each of the teacher questionnaires. The lowest completion rate for the school administrator questionnaire is for Blacks (57.1 percent); for the school fact sheet it is for American Indians or Alaska Natives (66.0 percent). For the teacher questionnaires the lowest rates are in the 53 to 55 percent range and are associated with Blacks, Hispanics, and American Indians or Alaska Natives. Since 60 percent of the Black and Hispanic students fielded in spring-third grade are enrolled in high minority schools (50 percent or higher), this may be associated with lower levels of response for the school administrator questionnaire from high minority schools. Of the 32 percent of Black and Hispanic students with no school administrator questionnaire data, roughly half are enrolled in high minority schools.

In addition to the child assessments, parent interview, teacher questionnaires, school administrator questionnaires and school fact sheets whose completion rates have been summarized in the preceding tables, various other types of data were collected during spring-third grade as well. Table 5-18 presents counts of completes and weighted and unweighted completion rates at the overall student level for these other data collection efforts. The facilities checklist has a 77.5 percent completion rate, which is about 11 points higher than that for the school administrator questionnaire but only 1 point higher than the rate for the school fact sheet, the two other school-level survey instruments. The student record abstract, which was to have been completed for all students except for those who moved and could not be located, has a 67.0 percent completion rate. Data were also collected during spring-third grade from the special education teachers for children in special education programs (fewer than 1,200). The completion rates for these instruments are higher than for the regular teacher questionnaires, 73.0 percent for part A, which captures teacher information, and 72.8 percent for part B, which relates to individual students who receive special education services.

Table 5-15. Number of completed child-level cases and child-level completion rates among 1998–99 kindergartners in spring-third grade, by survey type and selected child characteristics: School year 2001–02

Child characteristic[1]	Child assessments			Parent interview		
		Completion rate			Completion rate	
	Completes[2]	Weighted	Unweighted	Completes[3]	Weighted	Unweighted
All students	14,349	80.8	86.1	13,392	77.8	80.3
Sex						
Male	7,285	80.3	85.7	6,858	78.3	80.7
Female	7,064	81.5	86.6	6,534	77.3	80.1
Unknown sex	0	0.0	0.0	0	0.0	0.0
Race/ethnicity						
White (not Hispanic)	8,119	81.2	86.9	8,000	82.9	85.6
Black (not Hispanic)	1,872	78.0	83.6	1,570	67.0	70.2
Hispanic	2,575	82.2	85.9	2,295	73.9	76.6
Asian	963	84.1	86.4	757	68.6	67.9
Pacific Islander	171	84.9	87.2	161	82.3	82.1
American Indian or Alaska Native	248	75.5	81.3	242	75.5	79.3
Other	382	81.7	88.4	360	78.8	83.3
Unknown race/ethnicity	19	42.9	48.7	7	7.8	17.9
Year of birth						
1992	4,179	80.5	85.7	3,922	77.3	80.5
1993	10,086	81.1	86.4	9,397	78.2	80.5
Other/unknown	84	59.6	71.2	73	53.3	61.9

[1] Based on ECLS-K survey data and not on data from the sampling frame.
[2] Reading, math, or science assessment was scorable or child was disabled.
[3] Family structure portion of parent interview was completed.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 5-16. Number of completed child-level cases and child-level completion rates among 1998–99 kindergartners in spring-third grade, by survey type and selected child characteristics: School year 2001–02

Child characteristic[1]	School administrator questionnaire			School fact sheet		
		Completion rate			Completion rate	
	Completes[2]	Weighted	Unweighted	Completes[2]	Weighted	Unweighted
All students	12,361	66.1	73.3	14,064	76.5	83.4
Sex						
Male	6,241	65.0	72.6	7,119	75.7	82.9
Female	6,108	67.2	73.9	6,933	77.3	83.9
Unknown sex	12	100.0	100.0	12	100.0	100.0
Race/ethnicity						
White (not Hispanic)	7,309	70.8	78.2	8,061	79.1	86.2
Black (not Hispanic)	1,489	57.1	64.6	1,810	70.5	78.5
Hispanic	2,062	61.1	66.5	2,484	75.2	80.1
Asian	786	64.9	69.6	920	78.1	81.4
Pacific Islander	172	80.9	86.9	178	84.6	89.9
American Indian or Alaska Native	212	61.9	67.1	221	66.0	69.9
Other	305	60.4	70.3	358	72.4	82.5
Unknown race/ethnicity	26	68.2	66.7	32	83.7	82.1
Year of birth						
1992	3,681	68.1	74.8	4,115	76.9	83.6
1993	8,602	65.4	72.7	9,856	76.4	83.3
Other/unknown	78	54.5	65.0	93	69.4	77.5

[1] Based on ECLS-K survey data and not on data from the sampling frame.

[2] A completed questionnaire was defined as one that was not completely left blank.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

Table 5-17. Number of completed child-level cases and child-level completion rates among 1998–99 kindergartners in spring-third grade, by survey type and selected child characteristics: School year 2001–02

Child characteristic[1]	Teacher questionnaire part A			Teacher questionnaire part B			Teacher questionnaire part C		
		Completion rate			Completion rate			Completion rate	
	Completes[2]	Weighted	Unweighted	Completes[2]	Weighted	Unweighted	Completes[2]	Weighted	Unweighted
All students	11,770	62.4	69.8	11,741	62.3	69.6	11,802	62.7	70.0
Sex									
Male	5,930	61.4	69.0	5,921	61.4	68.9	5,951	61.7	69.3
Female	5,840	63.7	70.7	5,820	63.4	70.4	5,851	63.9	70.8
Unknown sex	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
Race/ethnicity									
White (not Hispanic)	7,174	68.6	76.8	7,158	68.4	76.6	7,186	68.8	76.9
Black (not Hispanic)	1,405	53.7	61.0	1,402	53.6	60.8	1,409	54.0	61.1
Hispanic	1,819	54.7	58.7	1,820	54.7	58.7	1,837	55.2	59.3
Asian	726	59.8	64.2	714	58.7	63.2	720	59.1	63.7
Pacific Islander	158	74.2	79.8	157	73.8	79.3	157	73.8	79.3
American Indian or Alaska Native	189	53.3	59.8	189	53.3	59.8	191	53.1	60.4
Other	288	54.3	66.4	289	55.0	66.6	289	55.1	66.6
Unknown race/ethnicity	11	21.8	28.2	12	27.7	30.8	13	29.0	33.3
Year of birth									
1992	3,491	63.3	71.0	3,488	63.2	70.9	3,497	63.5	71.1
1993	8,215	62.3	69.4	8,188	62.2	69.2	8,241	62.6	69.7
Other/unknown	64	38.6	53.3	65	40.2	54.2	64	39.6	53.3

[1] Based on ECLS-K survey data and not on data from the sampling frame.

[2] A completed questionnaire was defined as one that was not completely left blank.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, School year 2001–02.

Table 5-18. Number of completed instruments and child-level completion rates among 1998–99 kindergartners for additional data collected in spring-third grade, by survey type: School year 2001–02

Survey instrument	Completes	Completion rate	
		Weighted	Unweighted
Facilities checklist[1]	14,280	77.5	84.7
Student record abstract[1]	12,359	67.0	73.3
Special education part A[1]	875	73.0	75.0
Special education part B[1]	870	72.8	74.6

[1] A completed instrument was defined as one that was not completely left blank.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

5.6.2 Students Sampled in First Grade

In spring-first grade, the student sample was freshened to include first graders who had no chance of selection in the base year because they did not attend kindergarten in the United States or were in first grade in the fall of 1998. (For a detailed description of the freshening procedure see section 4.3.2.) This same group of students was followed into spring-third grade. Nonresponse in the freshened student sample could occur at two stages: during the procedure for sampling schools for freshening and identifying children to be used as freshening links in spring-first grade (first component) and then during data collection from the freshened students in spring-third grade (second component). The first component alone can further be decomposed into two sources: attrition due to entire schools refusing to implement the freshening procedure (*the school term*), and attrition due to ECLS-K sampled children moving to other schools (*the child term*). To contain costs, students who transferred from schools targeted for freshening were not used as links to identify freshened students, even when they were otherwise followed for data collection. These movers were considered freshening nonrespondents in the *child term*.

Table 5-19 presents weighted completion rates for freshened students. The two components of the completion rates are presented separately in table 5-19. The actual completion rates are the products of the two components. Since no freshening was done in the third grade, the first component is identical to that for spring-first grade. It is separated into *a school term* and *a child term* as described earlier. For this component, the completion rate is defined as the freshening completion rates, as opposed

Table 5-19. Number of completed child-level cases and child-level completion rates among freshened 1999–2000 first graders in spring-third grade, by freshening component and survey type: School year 2001–02

Survey instrument	Completes	Completion rate[1]	
		Weighted	Unweighted
First component of completion rate	7,135	64.3	77.9
School term[2]	7,192	65.4	78.5
Child term[3]	7,135	98.3	99.2
Second component of completion rate			
Child assessments[4]	121	78.3	76.1
Parent interview[5]	97	63.7	61.0
Teacher questionnaire part A[6]	86	53.1	54.1
Teacher questionnaire part B[6]	85	52.2	53.5
Teacher questionnaire part C[6]	82	50.9	51.6
Special education part A[6]	12	67.6	70.6
Special education part B[6]	13	73.3	76.5
School administrator questionnaire[6]	102	65.8	64.2
School fact sheet[6]	124	78.9	78.0
Facilities checklist[6]	126	80.3	79.2
Student records abstract[6]	103	68.7	64.8
Overall completion rate			
Child assessments[4]	121	50.4	59.3
Parent interview[5]	97	41.0	47.5
Teacher questionnaire part A[6]	86	34.2	42.1
Teacher questionnaire part B[6]	85	33.6	41.6
Teacher questionnaire part C[6]	82	32.7	40.2
Special education part A[6]	12	43.5	55.0
Special education part B[6]	13	47.1	59.6
School administrator questionnaire[6]	102	42.3	50.0
School fact sheet[6]	124	50.8	60.7
Facilities checklist[6]	126	51.7	61.7
Student records abstract[6]	103	44.2	50.5

[1] The first component is the completion rate for sample freshening in first grade. The second component is the completion rate for the survey instruments. The product of the two components is the overall completion rate for the survey instruments.
[2] The school term is the completion rate for schools targeted for freshening.
[3] The child term is the completion rate for children in schools that agreed to the freshening procedure.
[4] Reading, math, or science assessment was scorable or child was disabled.
[5] Family structure portion of parent interview was completed.
[6] A completed questionnaire was defined as one that was not completely left blank.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

to the survey instrument completion rates found in the second component. The weighted freshening completion rate for children in schools targeted for freshening (*the school term*) is 65.4 percent. The reasons that schools did not participate in the freshening process included refusing or being unable to provide the requested information in order to complete the procedures. Within the schools that agreed to freshen, the freshening completion rate is 98.3 percent, the slight loss due to students who transferred to other schools (*the child term*). Multiplying these two terms together gives a first component completion rate of 64.3 percent. Note that the first component rate for spring-third grade is identical to the first component rate for spring-first grade.

The second component varies by survey instrument. The rates for the paper-and-pencil instruments range from 50.9 percent for the child-level teacher questionnaire to 80.3 percent for the facilities checklist and are uniformly lower than for the kindergarten sample. The child assessments at 78.3 percent are 3 points lower than for the kindergarten sample and the parent interview, at 63.7 percent, is 14 points lower. The final completion rate for each instrument is the product of the two components. Because of the low rates at the first stage, these range from a high of 51.7 percent for the facilities checklist to a low of 32.7 percent for the child-level teacher questionnaire.

5.6.3 Spring-Third Grade Completion Rates—All Children

Table 5-20 presents final spring-third grade completion rates for children sampled in kindergarten, children sampled in first grade, and all children combined. Because children sampled in first grade represent such a small fraction of the total population of children, their inclusion in the computation of the completion rate brings down the rates for all children by less than one percent relative to the children sampled in kindergarten rates, even though the completion rates for children sampled in first grade are lower than the kindergarten rates.

Table 5-20. Number of completed child-level cases and child-level completion rates in spring-third grade, by sampling timeframe and survey type: School year 2001–02

Survey instrument	Children sampled in kindergarten			Children sampled in first grade			All children		
		Completion rate			Completion rate			Completion rate	
	Completes	Weighted	Unweighted	Completes	Weighted	Unweighted	Completes	Weighted	Unweighted
Child assessments[1]	14,349	80.8	86.1	121	50.4	59.3	14,470	80.1	85.9
Parent interview[2]	13,392	77.8	80.3	97	41.0	47.5	13,489	76.9	80.1
Teacher questionnaire part A[3]	11,770	62.4	69.8	86	34.2	42.1	11,856	61.7	69.6
Teacher questionnaire part B[3]	11,741	62.3	69.6	85	33.6	41.6	11,826	61.6	69.4
Teacher questionnaire part C[3]	11,802	62.7	70.0	82	32.7	40.2	11,884	62.0	69.7
Special education part A[3]	875	73.0	75.0	12	43.5	55.0	887	72.3	74.8
Special education part B[3]	870	72.8	74.6	13	47.1	59.6	883	72.2	74.5
School administrator questionnaire[3]	12,361	66.1	73.3	102	42.3	50.0	12,463	65.5	73.1
School fact sheet[3]	14,064	76.5	83.4	124	50.8	60.7	14,188	75.9	83.2
Facilities checklist[3]	14,280	77.5	84.7	126	51.7	61.7	14,406	76.9	84.5
Student records abstract[3]	12,359	67.0	73.3	103	44.2	50.5	12,462	66.5	73.1

[1] Reading, math, or science assessment was scorable or child was disabled.
[2] Family structure portion of parent interview was completed.
[3] A completed questionnaire was defined as one that was not completely left blank.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

5.6.4 Spring-Third Grade Completion Rates Conditioned on Child Assessments

Table 5-21 shows the completion rates for the child assessments, the parent interviews, and the school and teacher instruments for children who have nonzero child weights (C5CW0>0). These are children whose spring-third grade reading, math, or science assessments were scorable, or children who could not be assessed because of disabilities. For these children, the completion rate for the child assessments should be 100 percent. The less than 100 percent rate shown when children sampled in kindergarten are combined with children sampled in first grade is due to the school freshening nonresponse for children sampled in first grade.

When the completion rates are conditioned on the presence of the child weight (i.e., completion rates for children with C5CW0>0), they are at least 9 points higher than the unconditioned completion rates for all instruments but the special education questionnaires. For these last two instruments, the difference between the number of completes for the conditioned and unconditioned rates is very small; hence the conditioned rates are not affected as much as for the other instruments. For all the other instruments, the conditioned completion rates are higher by 9 points for the parent interviews and as high as 17 points for the student record abstract. These numbers are the differences between the unconditioned rates (table 5-20) and conditioned rates (table 5-21), hence not shown in any table.

Since data were collected from schools, parents, teachers, and children, there were many opportunities for sources to contribute differentially to nonresponse, and this is reflected in the varying completion rates in the tables in this section. These completion rates differ not only by survey instruments, but within each survey instrument they are also different by school and child characters. A separate report examines the potential for bias in estimates produced from the ECLS-K third grade data. Since analysis of the third grade data is conditioned on the base year—only base year respondents were included in the collection of first grade and third grade data—the analysis of nonresponse bias is built on the base year nonresponse bias analysis (see *Analysis of Nonresponse Bias in the Base Year Early Childhood Longitudinal Survey, Kindergarten Class of 1998–99*, [U.S. Department of Education, National Center for Education Statistics, forthcoming (a)]).

Table 5-21. Number of completed child-level cases and child-level completion rates in spring-third grade that includes children with scorable reading, math, or science assessments or children not assessed due to disabilities, by sampling timeframe and survey type: School year 2001–02

Survey instrument	Children sampled in kindergarten			Children sampled in first grade			All children		
		Completion rate			Completion rate			Completion rate	
	Completes	Weighted	Unweighted	Completes	Weighted	Unweighted	Completes	Weighted	Unweighted
Child assessments[1]	14,349	100.0	100.0	121	64.3	77.9	14,470	99.2	99.8
Parent interview[2]	12,564	86.8	87.6	90	49.7	57.9	12,654	85.9	87.4
Teacher questionnaire part A[3]	11,644	78.0	81.4	86	43.7	55.4	11,730	77.2	81.2
Teacher questionnaire part B[3]	11,614	77.8	81.2	85	43.0	54.7	11,699	77.0	81.0
Teacher questionnaire part C[3]	11,684	78.4	81.7	82	41.9	52.8	11,766	77.5	81.5
Special education part A[3]	854	74.1	75.8	12	43.5	55.0	866	73.4	75.6
Special education part B[3]	847	73.7	75.2	13	47.1	59.6	860	73.1	75.1
School administrator questionnaire[3]	11,846	80.1	82.8	93	50.6	59.9	11,939	79.4	82.6
School fact sheet[3]	13,447	92.5	94.0	113	60.4	72.7	13,560	91.7	93.8
Facilities checklist[3]	13,670	94.0	95.5	116	62.0	74.7	13,786	93.3	95.3
Student records abstract[3]	12,282	84.2	85.8	103	56.5	66.3	12,385	83.6	85.6

[1] Reading, math, or science assessment was scorable or child was disabled.

[2] Family structure portion of parent interview was completed.

[3] A completed questionnaire was defined as one that was not completely left blank.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 third grade data collection, school year 2001–02.

REFERENCES

U.S. Department of Education, National Center for Education Statistics (forthcoming [a]). *Analysis of Nonresponse Bias in the Base Year Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K).* By Mike Brick, John Burke, and Thanh Lê. Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics (forthcoming [b]). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Psychometric Report for the Third Grade.* Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics (forthcoming [c]). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Third Grade Methodology Report.* Washington, DC: U.S. Department of Education.

This page intentionally left blank.