ECLS-K BASE YEAR PUBLIC-USE DATA FILES AND ELECTRONIC CODEBOOK

Prepared by:

Westat
Rockville, Maryland
Karen Tourangeau
John Burke
Thanh Le
Siu Wan
Margaret Weant
Eugene Brown
Nancy Vaden-Kiernan
Elizabeth Rinker
Richard Dulaney
Kirsten Ellingsen
Brandon Barrett
Ismael Flores-Cervantes
Nicholas Zill

Educational Testing Service
Princeton, New Jersey
Judith Pollack
Donald Rock

University of Michigan
School of Education
Ann Arbor, Michigan
Sally Atkins-Burnett
Samuel Meisels

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

Education Statistics Services Institute
Washington, DC
Kristen Denton
Amy Rathbun
Jill Walston

TABLE OF CONTENTS

Chapter			Page
1	INTRODUCTION		1-1
	1.1	Background	1-2
	1.2	Conceptual Model	1-3
	1.3	Study Components	1-4
	1.4	Contents of Manual	1-5
	1.5	Differences Between the ECLS-K Restricted-Use Base Year Files and the ECLS-K Base Year Public-Use Files	1-6
2	DESCRIPTION OF DATA COLLECTION INSTRUMENTS		2-1
	2.1	The Direct Child Assessments	2-1
	2.2	Parent Interview	2-9
	2.3	Teacher Questionnaires	2-14
		2.3.1 Content of the ARS	2-14
		2.3.2 Teacher SRS	2-16
	2.4	Special Education Teacher Questionnaires	2-17
	2.5	Adaptive Behavior Scale	2-20
	2.6	School Administrator Questionnaire	2-21
	2.7	School Facilities Checklist	2-21
	2.8	School Records Abstract Form	2-21
	2.9	Salary and Benefits Questionnaire	2-23
3	ASSESSMENT AND RATING SCALE SCORES USED IN THE ECLS-K		3-1
	3.1	Direct Cognitive Assessment	3-1
		3.1.1 Number-Right Scores	3-1
		3.1.2 IRT Scale Scores	3-2
		3.1.3 Standardized Scores (T-Scores)	3-6
		3.1.4 Proficiency Scores	3-7
		3.1.5 Choosing the Appropriate Score for Analysis	3-11
		3.1.6 Reliabilities	3-13
	3.2	Indirect Cognitive Assessment (ARS)	3-15
		3.2.1 Comparison to Direct Cognitive Assessment	3-15
	3.3	IRT Scores Available for the ARS	3-15
	3.4	The Oral Language Development Scale (OLDS)	3-17

TABLE OF CONTENTS (continued)

Chapter				Page
	3.5	Psychomotor Assessment		3-18
	3.6	SRS		3-19
		3.6.1	Teacher SRS	3-20
		3.6.2	Parent SRS	3-21
4	SAMPLE DESIGN AND IMPLEMENTATION			4-1
	4.1	Selecting the Area Sample		4-1
	4.2	Selecting the School Sample		4-2
		4.2.1	Frame Construction	4-4
		4.2.2	School Measure of Size, Stratification, and Sample Selection	4-5
		4.2.3	Freshening the School Sample	4-6
		4.2.4	Sampling Children, Parents, and Teachers within Schools	4-8
	4.3	Calculation and Use of Sample Weights		4-10
		4.3.1	Types of Sample Weights	4-11
		4.3.2	Weighting Procedures	4-18
		4.3.3	Computation of School and Teacher Weights	4-20
		4.3.4	Computation of Child Weights	4-21
		4.3.5	Replicate Weights	4-24
		4.3.6	Characteristics of Sample Weights	4-24
	4.4	Variance Estimation		4-25
		4.4.1	Paired Jackknife Replication Method	4-26
		4.4.2	Taylor Series Method	4-28
		4.4.3	Specifications for Computing Standard Errors	4-28
	4.5	Design Effects		4-31
		4.5.1	Use of Design Effects	4-32
		4.5.2	Average Design Effects for the ECLS-K	4-33
5	DATA COLLECTION METHODS AND RESPONSE RATES			5-1
	5.1	Data Collection Methods		5-1
	5.2	Study Endorsements and School Recruitment		5-1
	5.3	Field Staff Training		5-4

TABLE OF CONTENTS (continued)

Chapter				Page
	5.4	Fall-Kindergarten Data Collection		5-7
		5.4.1	Advance Mailings	5-7
		5.4.2	Preassessment Visits	5-8
		5.4.3	Conducting the Direct Child Assessment	5-12
		5.4.4	Conducting the Parent Interview	5-14
		5.4.5	Teacher Data Collection	5-15
	5.5	Fall-Kindergarten Response Rates		5-15
	5.6	Spring-Kindergarten Data Collection		5-22
		5.6.1	Advance Contact with Respondents	5-22
		5.6.2	Preassessment Contact	5-22
		5.6.3	Conducting Direct Child Assessments	5-26
		5.6.4	Conducting the Parent Interview	5-27
		5.6.5	Teacher and School Data Collection	5-28
		5.6.6	Conducting Data Collection on Children Who Withdrew from Their Original School	5-29
		5.6.7	Collection of Head Start Data	5-30
	5.7	Data Collection Quality Control		5-31
		5.7.1	Child Assessment Observations	5-32
		5.7.2	Parent Validations	5-32
		5.7.3	School Validations	5-33
		5.7.4	Quality Control of the OLDS Scoring	5-34
		5.7.5	Assessor Effects	5-34
	5.8	Spring-Kindergarten Response Rates		5-35
	5.9	Base Year Response Rates		5-46
6	DATA PREPARATION			6-1
	6.1	Coding and Editing Specifications for CATI/CAPI		6-1
		6.1.1	Range Specifications	6-1
		6.1.2	Consistency Checks (Logical Edits)	6-2
		6.1.3	Coding	6-2
	6.2	Coding and Editing Specifications for Hard Copy Questionnaires		6-6
		6.2.1	Receipt Control	6-6
		6.2.2	Coding	6-6

TABLE OF CONTENTS (continued)

Chapter				Page
		6.2.3	Data Entry	6-8
		6.2.4	Data Editing	6-9
7	DATA FILE CONTENT AND COMPOSITE VARIABLES			7-1
	7.1		Identification Variables	7-1
	7.2		Missing Values	7-2
	7.3		Variable Naming Conventions	7-4
	7.4		Composite Variables	7-6
		7.4.1	Child Variables	7-6
		7.4.2	Family and Household Variables	7-8
		7.4.3	Teacher Variables	7-12
		7.4.4	School and Class Characteristics	7-13
		7.4.5	Parent Identifiers and Household Composition	7-17
		7.4.6	Changes in Household Composition between Fall-Kindergarten and Spring-Kindergarten	7-21
		7.4.7	The Supplemental Section (SPQ) in the Spring-Kindergarten Parent Interview	7-21
		7.4.8	Industry and Occupation Codes Used in ECLS-K	7-22
	7.5		Using Teacher Data Collected Either in Fall-kindergarten or Spring-kindergarten	7-25
	7.6		The Relationship Among Composite Variables on the Child, Teacher, and School Catalogs	7-26
	7.7		Creating a Child File for Classroom Level Analysis	7-31
	7.8		Creating a Classroom Level File Using Teacher Data	7-31
	7.9		Masked Variables	7-72
8.	GUIDE TO THE ECLS-K BASE YEAR PUBLIC-USE ELECTRONIC CODE BOOK			8-1
	8.1		Introduction	8-1
		8.1.1	Hardware/Software Requirements	8-2
		8.1.2	ECLS-K Base Year Public-Use ECB Features	8-2
	8.2		Installing, Starting, and Exiting the ECLS-K Base Year Public-Use ECB	8-3
		8.2.1	Installing the ECLS-K Base Year Public-Use ECB Program on Your PC	8-3
		8.2.2	Starting the ECLS-K Base Year Public-Use ECB	8-8

TABLE OF CONTENTS (continued)

Chapter			Page
	8.2.3	Exiting the ECLS-K Base Year Public-Use ECB	8-11
	8.2.4	Removing the ECB Program From Your PC	8-11
	8.2.5	The Title Bar	8-12
	8.2.6	The Menu Bar	8-12
	8.2.7	Using Shortcut Keys to Navigate	8-13
8.3	The Variable List		8-14
	8.3.1	Searching the Code Book for Variables	8-16
8.4	The Working Taglist		8-22
	8.4.1	Opening A Taglist	8-23
	8.4.2	Adding Variables to the Working Taglist	8-23
	8.4.3	Removing Variables from the Working Taglist	8-25
	8.4.4	Saving Taglists	8-26
	8.4.5	Using Predefined Taglists	8-28
	8.4.6	Deleting Taglists	8-29
	8.4.7	Viewing Code Book and Variable Information	8-30
8.5	Extracting Data from the ECB		8-37
	8.5.1	Reviewing the Extract Specifications	8-44
	8.5.2	Repairing and Compacting the Database	8-45
8.6	The Child Catalog		8-45
	8.6.1	Child Catalog Topical Variable Groupings	8-47
	8.6.2	Child Catalog Predefined Taglists	8-56
	8.6.3	Child Catalog Limiting Fields	8-63
8.7	The Teacher Catalog		8-64
	8.7.1	Teacher Catalog Topical Variable Groupings	8-65
	8.7.2	Teacher Catalog Predefined Taglists	8-68
	8.7.3	Teacher Catalog Limiting Fields	8-68
8.8	The School Catalog		8-69
	8.8.1	School Catalog Topical Variable Groupings	8-71
	8.8.2	School Catalog Predefined Taglists	8-73
	8.8.3	School Catalog Limiting Fields	8-73

TABLE OF CONTENTS (continued)

Chapter			Page
	8.9	Switching Catalogs	8-74
	8.10	Menu Bar Descriptions	8-76

List of Tables

Tables

2-1	Instruments used in the fall and spring ECLS-K	2-1
2-2	Flow of the fall direct child assessment	2-4
2-3	Direct child assessment	2-8
2-4	Flow of the spring direct child assessment	2-9
2-5	ECLS-K parent interview by major content topics and point of data collection	2-12
2-6	ARS response scale	2-15
2-7	Teacher questionnaire	2-18
2-8	Special education teacher questionnaire[1]	2-20
2-9	School administrator questionnaire[1]	2-22
3-1	Direct cognitive assessment: types of scores, variable names, descriptions, ranges, means, and standard deviations[1]	3-3
3-2	Reliability of IRT-based scores and routing test number correct	3-14
3-3	Split half reliability of item-cluster-based scores (proficiency level scores)	3-14

TABLE OF CONTENTS (continued)

List of Tables (continued)

Table		Page
3-4	Reliability for the Rasch-based score	3-16
3-5	ARS: variable names, descriptions, ranges, means, and standard deviations	3-17
3-6	OLDS: variable names, descriptions, ranges, means, and standard deviations	3-18
3-7	Psychomotor scales: variable names, descriptions, ranges, means, and standard deviations	3-19
3-8	SRS response scale	3-19
3-9	Split half reliability for the teacher SRS scale scores	3-20
3-10	Split half reliability for the parent SRS scale scores	3-20
3-11	Teacher social rating scores: variable names, descriptions, ranges, means, and standard deviations	3-21
3-12	Parent social rating scores: variable names, descriptions, ranges, means, and standard deviations	3-22
4-1	Stratum definitions for the 38 non-self-representing strata	4-3
4-2	Distribution of the ECLS-K PSU sample by SR status, MSA status, and census region	4-4
4-3	Characteristics of the ECLS-K original school sample	4-9
4-4	The ECLS-K: fall-kindergarten cross-sectional weights	4-12
4-5	The ECLS-K: spring-kindergarten cross-sectional weights	4-13
4-6	The ECLS-K: base year longitudinal weights	4-17
4-7	Distribution of schools by number of cases (children) with nonzero weights	4-18
4-8	Number of children who were not assessed due to special situations	4-22

TABLE OF CONTENTS (continued)

List of Tables (continued)

Table		Page
4-9	Characteristics of teacher- and school-level weights	4-24
4-10	Characteristics of child-level weights	4-25
4-11	ECLS-K Taylor Series stratum and first-stage unit identifiers	4-29
4-12	Specifications for computing standard errors	4-30
4-13	ECLS-K, fall-kindergarten: standard errors and design effects for the full sample – child level	4-35
4-14	ECLS-K, spring-kindergarten: standard errors and design effects for the full sample – child level	4-36
4-15	ECLS-K, panel: standard errors and design effects for the full sample – child level	4-37
4-16	ECLS-K: median design effects for subgroups – child level	4-38
4-17	ECLS-K, fall-kindergarten: standard errors and design effects for the full sample – teacher level	4-40
4-18	ECLS-K, spring-kindergarten: standard errors and design effects for the full sample – teacher level	4-41
4-19	ECLS-K: median design effects for subgroups – teacher level	4-42
4-20	ECLS-K, spring-kindergarten: standard errors and design effects for the full sample – school level	4-43
4-21	ECLS-K: median design effects for subgroups – school level	4-44
5-1	OLDS routing results in fall-kindergarten	5-13
5-2	Number of children excluded from or accommodated in the fall-kindergarten assessment	5-13
5-3	Number of children excluded from parts of the psychomotor assessment	5-13
5-4	The ECLS-K, fall-kindergarten: number of cooperating schools and before-substitution school-level response rates	5-16

TABLE OF CONTENTS (continued)

List of Tables (continued)

Table		Page
5-5	The ECLS-K fall-kindergarten: number of completed child-level cases and child-level completion rates, by school characteristic	5-18
5-6	The ECLS-K, fall-kindergarten: number of completed child-level cases and child-level completion rates, by child characteristic	5-20
5-7	The ECLS-K, fall-kindergarten: number of completed teacher questionnaires and teacher-level completion rates	5-23
5-8	The ECLS-K, fall-kindergarten: number of completed child-level cases and overall child-level response rates, by school characteristic	5-24
5-9	The ECLS-K, fall-kindergarten: number of completed child-level cases and overall child-level response rates, by school characteristic	5-25
5-10	OLDS routing patterns in spring-kindergarten for fall-kindergarten respondents	5-26
5-11	OLDS routing results in spring-kindergarten	5-27
5-12	Number of children excluded from or accommodated in the spring-kindergarten assessment	5-27
5-13	Number of children who moved by category and assessment result	5-30
5-14	The ECLS-K, spring-kindergarten: number of cooperating schools and before-substitution school-level response rates	5-36
5-15	The ECLS-K, spring-kindergarten: number of completed child-level cases and child-level completion rates, by school characteristic	5-37
5-16	The ECLS-K, spring-kindergarten: number of completed child-level cases and child-level completion rates, by child characteristic	5-40
5-17	The ECLS-K, spring-kindergarten: number of completed teacher questionnaires and teacher-level completion rates	5-42
5-18	The ECLS-K, spring-kindergarten: number of completed school administrator questionnaire and school-level completion rates	5-43

TABLE OF CONTENTS (continued)

List of Tables (continued)

Table		Page
5-19	The ECLS-K, spring-kindergarten: child-level sample sizes and overall response rates, by school characteristic	5-44
5-20	The ECLS-K, spring-kindergarten: number of completed teacher questionnaires and teacher-level response rates	5-47
5-21	The ECLS-K, spring-kindergarten: number of completed instruments and child-level completion rates for additional data collected	5-48
5-22	The ECLS-K: number sampled and number and percent cooperating during fall- or spring-kindergarten, with percent of fall respondents who also responded in the spring	5-48
7-1	Missing data for SES variables	7-9
7-2	Preliminary Census poverty thresholds for 1998[1]	7-12
7-3	Pointers to parent figure questions	7-19
7-4	SPQ item matches to fall-kindergarten items	7-23
7-5	Relationship among composite variables on the three ECB catalogs	7-27
7-6	Composite variables	7-34
7-7	Masked variables in the child catalog	7-74
7-8	Masked variables in the teacher catalog	7-83
7-9	Masked variables in the school catalog	7-89
8-1	Add variables buttons	8-24
8-2	Remove variables buttons	8-25
8-3	Code book information	8-30
8-4	Navigation buttons	8-32
8-5	Child catalog required variables	8-46

TABLE OF CONTENTS (continued)

List of Tables (continued)

Table		Page
8-6	Child catalog blocks	8-48
8-7	Child catalog topical variable groupings	8-50
8-8	Fall-kindergarten predefined taglist - child catalog	8-56
8-9	Spring-kindergarten predefined taglist – child catalog	8-58
8-10	Kindergarten base year predefined taglist – child catalog	8-60
8-11	Teacher catalog required variables	8-64
8-12	Teacher catalog blocks	8-66
8-13	Teacher catalog topical variable groupings	8-67
8-14	School catalog required variables	8-69
8-15	School catalog blocks	8-71
8-16	School catalog topical variable groupings	8-72

List of Figures

Figure		
1-1	ECLS-K conceptual model	1-4
8-1	Windows Run screen	8-4
8-4	Choose Destination Location	8-6
8-5	Setup Status	8-6
8-7a.	Desktop icon	8-8
8-7b	Desktop screen - click start	8-8
8-8	First time user dialog box	8-9

List of Figures (continued)

Figures		Page
8-9	The ECLS-K Base Year ECB splash screen	8-9
8-10	Select Catalog screen	8-10
8-11	Main ECB screen	8-10
8-12	Exit screen	8-11
8-13	Save Working Taglist dialog box	8-11
8-14	Title Bar	8-12
8-15	Menu Bar	8-13
8-16	The ECLS-K Base Year ECB main screen	8-14]
8-17	The Variable List	8-15
8-18	The “Go” button	8-17
8-19	Narrow Text Search dialog box	8-18
8-20	No matches found message	8-19
8-21	Example of narrowing a search	8-20
8-22	Expand Text Search dialog box	8-21
8-23	No matches found message	8-21
8-24	The ECB Working Taglist	8-22
8-25	Open Taglist dialog box	8-23
8-26	Add Taglist dialog box	8-25
8-27	Save Taglist as dialog box	8-27
8-28	Save Taglist as dialog box (#2)	8-28
8-29	Delete Taglist Selection screen	8-29

List of Figures (continued)

Figures		Page
8-30	Delete Taglist confirmation window	8-29
8-31	ECB view	8-31
8-32	Printing Status screen	8-32
8-33	Export Taglist selection screen	8-33
8-34	Variable Quick View	8-34
8-35	Add All Variables dialog box	8-35
8-36	View of the entire code book	8-36
8-37	Printer Dialog box	8-36
8-38	Teacher catalog limiting fields dialog box	8-38
8-39	Save SAS program file dialog box	8-39
8-40	Save SAS program file location browse screen	8-39
8-41	Save SAS data file dialog box	8-40
8-42	Save SPSS program file dialog box	8-41
8-43	Save SPSS program file location browse screen	8-41
8-44	Save SPSS data file dialog box	8-42
8-45	Save STATA program file dialog box	8-43
8-46	Save STATA program file location browse screen	8-43
8-47	Save STATA data file dialog box	8-44
8-48	Repair database completed screen	8-45
8-49	Child catalog limiting fields	8-63
8-50	Teacher catalog limiting fields	8-68

TABLE OF CONTENTS (continued)

List of Figures (continued)

Figures		Page
8-51	School catalog limiting fields	8-73
8-52	Save changes	8-74
8-53	Select Catalog screen	8-75

List of Exhibits

Exhibit		
5-1	Timeline of base year data collection	5-2

List of Appendixes

Appendix	
A	PDF FILES CONTAINING BASE YEAR INSTRUMENTS
B	RECORD LAYOUT OF SCHOOL, TEACHER, AND CHILD DATA FILES

1. INTRODUCTION

The Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) is a multisource, multimethod study that focuses on children's early school experiences beginning with kindergarten. 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 the Survey Research Center and the School of Education at the University of Michigan and Educational Testing Services (ETS).

The ECLS-K is following a nationally representative cohort of children from kindergarten through fifth grade. A total of 21,260 children throughout the country participated by having a child assessment or parent interview in the fall and/or spring of kindergarten. The children were in kindergarten when sampled. Base year data were collected in the fall of 1998 and spring of 1999. Four waves of data collection are planned beyond kindergarten: fall and spring first grade, and spring third and fifth grades. All data collection will be completed in the spring of 2004 when most of the children will be in fifth grade.

The ECLS-K has several major objectives and numerous potential applications. The ECLS-K combines elements of (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) a cross-sectional study 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) will follow a national sample of children, born in the year 2001, from birth through first grade. ECLS-B will focus 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 range and breadth of data required to more fully describe and understand children's health, early learning, development, and education experiences.

The ECLS-K has both descriptive and analytic purposes. It will provide descriptive data on children's status at entry into school, children's transition into school, and their progress through fifth grade. The ECLS-K also will provide a rich data set that will enable researchers to analyze how a wide range of family, school, community, and individual variables affect children's early success in school; explore school readiness and the relationship between the kindergarten experience and later elementary school performance; and 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 is reflected in an intensified recent national policy aimed at ensuring that children are capable of reading by the third grade, providing college student and adult volunteer tutors for children who are having difficulty learning to read, and increasing the number of children served by Head Start to 1 million by the year 2002.

Efforts to expand and improve early education will benefit from insights gained through analyses of data from the large-scale, nationally representative, longitudinal ECLS-K database. The ECLS-K database contains information about the types of preschool and elementary 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. This 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 policy analysts to use an ecological perspective 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 family and school environments.

Data collected during the kindergarten year can serve as baseline measures to examine how schooling shapes later individual development. 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 school experiences in kindergarten and the early grades.

A goal of the kindergarten data collection has been to describe accurately the diversity of kindergarten children and the programs they attend. For instance, national data are available for the first time on public and private kindergarten programs and the children who attend them. 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 approaches 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 interaction 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, socio-emotional and intellectual development and incorporates factors from the child's family, community, and school-classroom environments. The conceptual model is presented in figure 1-1. 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.

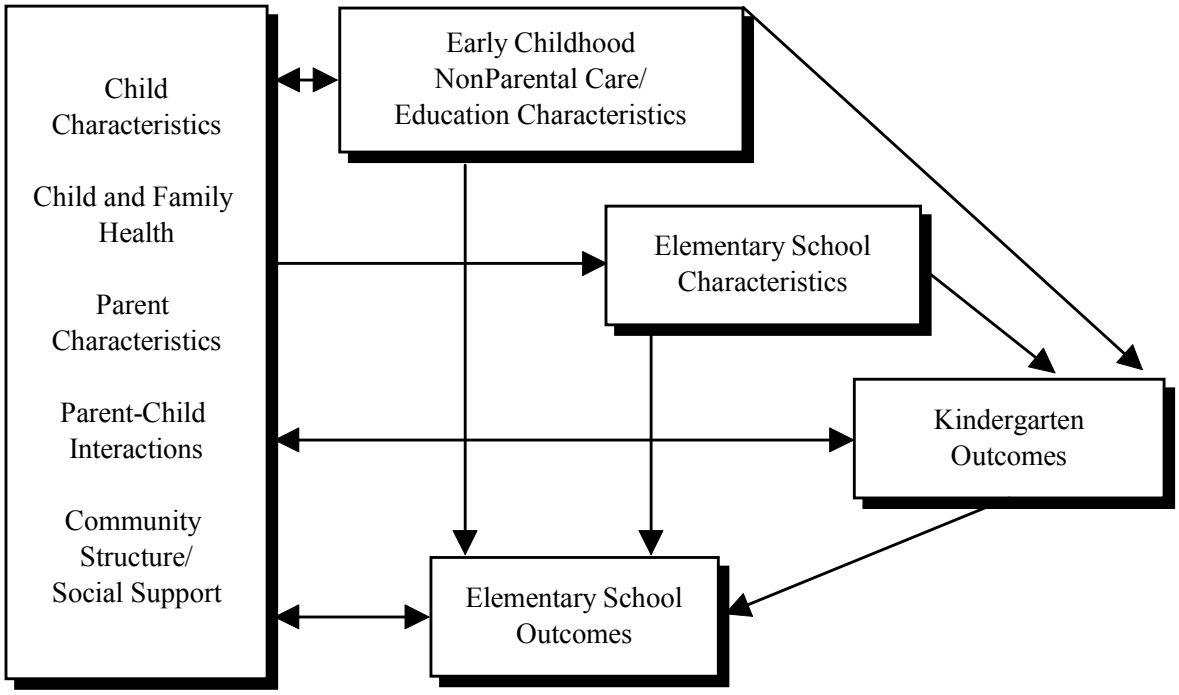

Figure 1-1. ECLS-K conceptual model

1.3 Study Components

The emphasis that is being placed on the whole of the child's environments and development has critical implications for the design of the ECLS-K. The design of the study allows for the collection of data from the child, the child's parents/guardians, teachers, and schools.

- **Children** are asked to 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 (e.g., general knowledge, literacy, and quantitative skills) and noncognitive (e.g., fine motor and gross motor coordination, socio-emotional) skills and knowledge. All measures of a child's cognitive skills are obtained through an untimed one-on-one assessment of the child. During later years of the ECLS-K, children will report on their own experiences in and out of school.

- **Parents/guardians** are an important source of information about the families of the children selected for the study and about themselves. They are asked to provide key information about their children, especially during the first years of the study. Parents are one of the important sources of 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 telephone interviewing (CATI) [or computer-assisted personal interviewing (CAPI) if they do not have a telephone].

- **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 upon 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. Teachers complete self-administered questionnaires each time children are assessed, with the exception of the fall first grade data collection.

- **School administrators** are asked to complete self-administered questionnaires during the spring data collection. They 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.

1.4 Contents of Manual

This manual provides documentation and guidance for users of the three public-use data files of the ECLS-K: the child file, teacher file, and the school file. The manual contains information about the data collection instruments (chapter 2) and the psychometric properties of these instruments (chapter 3). The manual describes the ECLS-K sample design (chapter 4); data collection procedures and response rates (chapter 5); and data processing procedures (chapter 6). In addition, the manual shows how the public-use data file is structured, provides definitions of composite variables (chapter 7), and explains how to use the Electronic Code Book (chapter 8). The Electronic Code Book contains unweighted frequencies for all variables.

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 employ 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.

1.5 Differences Between the ECLS-K Restricted-Use Base Year Files and the ECLS-K Base Year Public-Use Files

In preparing the ECLS-K Base Year Public-Use data files the National Center for Education Statistics (NCES) has taken steps to minimize the likelihood that an individual school, teacher, parent or child participating in the study can be identified. This is in compliance with the Privacy Act of 1974 and the National Education Statistics Act of 1994, both of which mandate the protection of the confidentiality of respondents. The process began with the ECLS-K Restricted-Use Base Year data files, which underwent a formal disclosure risk analysis. Variables identified as posing the greatest disclosure risk were altered, and in some instances entirely suppressed, and in this way the ECLS-K Base Year Public-Use data files were created. Every effort has been made to alter the files as little as possible, consistent with the requirement for confidentiality protection. After altering the variables the disclosure risk analysis was repeated to verify that the disclosure risk had been reduced to acceptable levels.

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

- Outlier values were top- or bottom- coded;
- Individual cases for which a particular variable posed an especially high risk for disclosure had the value of that variable altered (usually by no more than 5 to 10 percent) to reduce the risk;
- Some continuous variables were modified into categorical variables, and certain categorical variables had their categories collapsed; and
- Certain variables with too few cases and a sparse distribution were suppressed altogether, rather than modified.

A comprehensive list of the variables that have been altered or suppressed can be found in section 7.9.

Both the public- and restricted-use files provide data at the individual child, teacher, and school levels. The modifications that were implemented to avoid the identification of schools, teachers, 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 have been 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 and those

interested in examining the type and number of hours of kindergarten programs offered in schools 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 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.

2. DESCRIPTION OF DATA COLLECTION INSTRUMENTS

This chapter describes the instruments used to collect base year data in the fall of 1998 and spring of 1999 for the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K). In the fall, data were collected by teacher questionnaires, parent computer-assisted interviews (CAIs), and direct child assessments. Most of the fall instruments were repeated for the spring of 1999 data collection, although the parent and teacher measures varied by content between the two data collection points to ease respondent burden. This chapter also describes the Head Start verification study, where part of the data was collected in the fall parent interview and student record abstract. Table 2-1 below lists all of the instruments used in each of the two rounds of data collection.

Table 2-1. Instruments used in the fall and spring ECLS-K

Fall-Kindergarten	Spring-Kindergarten
Parent Interview	Parent Interview
Child Assessment	Child Assessment
Teacher Questionnaire-Part A	Teacher Questionnaire-Part A
Teacher Questionnaire-Part B	Teacher Questionnaire-Part B
Teacher Questionnaire-Part C	Teacher Questionnaire-Part C
	Special Education Teacher Questionnaire-Part A
	Special Education Teacher Questionnaire-Part B
	Adaptive Behavior Scale
	School Administrator Questionnaire
	Student Record Abstracts
	School Facilities Checklist
	Salary and Benefits Questionnaire

Appendix A contains a copy of the base year instruments, except for the direct child assessment, the social rating scale[1] in the parent interview and teacher questionnaire, and the adaptive

[1] Adapted with permission from Elementary Scale A ("How Often?"), Frank M. Gresham and Stephen N. Elliott ©1990, American Guidance Service, Inc.

behavior scale.[2] These latter measures contain copyright protected materials and agreements with the test publishers that restrict their distribution.

2.1 The Direct Child Assessments

Fall Child Assessments. In the fall of the base year, one-on-one child assessments were conducted with the sampled children. This assessment included cognitive, psychomotor, and physical components. The assessment took approximately 50-70 minutes and was designed to provide data on the developmental status of children in the United States at the start of their formal schooling. The ECLS-K cognitive assessment battery consisted of questions in three subject areas: language and literacy, mathematical thinking, and general knowledge. Psychomotor assessments were also included in the fall, along with assessments of the child's height and weight.

The assessment began by verifying the child's name and administering a set of warm-up exercises similar in form to the items used to administer the cognitive component. Prior to administering the cognitive assessment battery, a language-screening assessment was administered to those children identified from their school records (or by their teacher, if no school records were available) as coming from a language minority background (meaning that their primary home language was not English). This screening test was used to determine if a child was able to understand and respond to the cognitive assessment items in English.

The Language Screener. Efforts were made to include children who spoke a language other than English in the ECLS-K assessment. Field supervisors either checked the school records to determine children's home language or, if records were not available, requested this information directly from children's teachers. (See chapter 5 for a complete description of this process.) A brief language screener, the Oral Language Development Scale (OLDS), was given to those children who had a non-English language background. The screener determined if children understood English well enough to receive the direct child assessment in English. Children who passed the language screener received the full ECLS-K direct assessment battery. Children who did not pass an established cut score on the language screener received a reduced version of the ECLS-K assessments.

---

[2] Lambert, Nadine, Nihira, Kazuo, and Leland, Henry, Adaptive Behavior Scale-Second Edition, ©1993, The American Association on Mental Retardation.

The OLDS was also used to capture baseline information on children whose primary language was not English. The baseline data for these children can be used by researchers to examine English language acquisition over time.

The OLDS measured children's listening comprehension, vocabulary, and ability to understand and produce language. The OLDS consisted of three parts extrapolated from the preLAS 2000 (Duncan, S.E. and De Avila, E.A., 1998[3]). For the OLDS, children participated in Simon Says, Art Show, and Let's Tell Stories. Part one, Simon Says, measured listening comprehension of basic English instructions (i.e., asking a child to do things such as touch ear, pick up paper, or knock on table). Part two, Art Show, was a picture vocabulary assessment where children were asked to name pictures they were shown. The Art Show served as an assessment of a child's ability to produce language and measured the child's command of expressive language. The final part of the OLDS, Let's Tell Stories, was used to obtain a sample of a child's natural speech by asking a child to retell a story read by the assessor. The child was read two different stories and asked to tell what happened using pictures as prompts. The assessor recorded on paper exactly what the child said and scored the story using the established preLAS 2000 scoring rules. The scores assigned were based on the complexity of the child's sentence structure and vocabulary in his or her retelling of the story. These scores provide researchers with a direct measure of oral language performance.

Children who passed the language screener received the full English direct assessment. Certain components of the direct child assessment could also be conducted in Spanish. If a child did not pass the language screener but spoke Spanish, he or she was administered a Spanish translated form of the mathematics assessment and an alternate form of the language screener, the Spanish version of the Oral Language Development Scale (Spanish OLDS), as well as a psychomotor assessment that was conducted in Spanish. The Spanish OLDS that was administered is similar in content to the English OLDS and measures the same constructs.

A variety of steps were undertaken to confirm that the scores obtained from the Spanish mathematics assessment would be comparable to those for the English version. After the test items were translated into Spanish, a back-translation was carried out, followed by a review of the assessment instrument by two Spanish-speaking math experts. Psychometric analyses were also performed to compare the English and Spanish mathematics test results. Differential Item Functioning (DIF)

[3] Duncan, S.E. and De Avila, E. A., preLAS 2000 Cue Picture Book English Form C, CTB/McGraw-Hill Companies, Inc., 1998

procedures were carried out to determine whether the relative difficulty of each of the mathematics items was comparable in the English and Spanish version. Of the 31 mathematics items with sufficient numbers of observations for analysis, 16 appeared to be relatively easier for the Spanish speaking children and 15 for the English version. Of these 31 DIF statistics, 28 showed differences that were slight and/or not statistically significant. The three items marked as having "C" level DIF (sizeable difference in performance, and statistical significance) were split between the groups, with one item found to be relatively easier on the Spanish test and two relatively easier in English. A finding of differential item functioning does not necessarily indicate bias in the test; it may simply be an indication that achievement differences among groups may be more pronounced for certain skills than for others. A review of the "C" level test items by developmental and Spanish-language experts found no evidence of bias in either the content or the translation of the items. When Item Response Theory (IRT) procedures were carried out to calculate scaled scores, plots showing the fit of data to item parameters were examined for the Spanish compared to the English mathematics tests. The results were very similar to the DIF findings: there was an essentially identical fit for almost all of the test items; there were small differences in one direction or the other for a few items; and there was no evidence of systematic bias. All of these analyses support the conclusion that the language of administration had little or no impact on the scores obtained.

Children who did not pass the established cut score on the language screener and whose native language was not Spanish were excluded from the assessment; however, assessors collected physical measurements of these children's height and weight. Table 2-2 shows the paths of the direct child assessment by home language and scores on the English OLDS.

Table 2-2. Flow of the fall direct child assessment

Home Language	English OLDS	Spanish OLDS	Warm-up Booklet	Reading	Math	Psycho-motor	General Knowledge	Height/ Weight
English			✓	✓	✓	✓	✓	✓
Other	✓ Score **at or Above** Cutpoint		✓	✓	✓	✓	✓	✓
	✓ Score **Below** Cutpoint Speaks Spanish	✓	✓ **Spanish**		✓ **Spanish**	✓ **Spanish**		✓
	✓ Score **Below** Cutpoint Doesn't Speak Spanish							✓

Two-Stage Assessment Design

The direct cognitive assessment consisted of a set of two-stage assessments: a first-stage routing section for each of the three subject areas, followed by several alternative second-stage forms. The same reading, mathematics, and general knowledge routing sections, consisting of 12 to 20 items with a broad range of difficulty, were administered to all children. A child's performance on the routing section determined the second-stage form that was administered. The reading and mathematics assessments had low, middle, and high difficulty second-stage options, while the general knowledge assessment had two second-stage alternatives. The purpose of this adaptive assessment design was to maximize accuracy of measurement and minimize administration time.

The second-stage forms varied by level of difficulty so that a child would be administered questions appropriate to his or her current level of ability for each cognitive domain. Administering assessment items that are too hard for a particular child not only causes frustration and distress but also provides very little information on the precise level of the child's ability. Because most of the items are likely to be answered incorrectly, all that can be concluded is that the child's ability level is below the difficulty level of the questions, but there is no information on *how much* below. Similarly, giving a larger number of very *easy* items to a child of *high* ability may be boring and, again, they are not very useful in pinpointing the child's achievement level. The assessment items that provide the best information are those that are slightly too easy or slightly too hard for an individual. The pattern of right and wrong responses on such items makes it possible to estimate ability within a narrow range. The number of questions included in this assessment was limited in order to minimize the time and burden on the children. Consequently, it was important to match the difficulty of the questions to the ability level of the children, to the extent that this was possible with preselected sets of items. The routing section provided a rough estimate of each child's achievement level, so that a second-stage form with items of the appropriate difficulty for maximizing measurement accuracy could be selected.

The cognitive assessment included both multiple choice and open-ended items. For ease of administration, questions of similar format were grouped together in order of increasing difficulty within each group. When the question format changed, practice items were used to introduce children to the new format. Assessments were shortened or discontinued if the administrator perceived that the child was uncomfortable or distressed about responding to the assessment items. When a child did not respond to a question, the assessor repeated the question. If there was still no response, or the child did not know the

answer, the assessor entered a code for "don't know" and moved on to the next question, while periodically reminding the child to try. For a child who did not respond to several questions in a row, the assessor still waited ten seconds after reading each question and entered a code for "don't know," but did not continue to repeat each question. If no response was given to ten questions in a row, assessors entered a "refuse" code into the computer for the remainder of the items in that subject area, without reading the questions, until reaching the next subject area, where he or she resumed reading the questions. This procedure was used to give children who did not want to respond to questions in one subject area (e.g., reading) a chance to respond to questions in another subject area (e.g., math). Scores in each subject area were computed only if at least ten questions were answered in the combined first and second stages.

Cognitive Components. The cognitive assessment focused on three general areas of competence: (1) language use and literacy (reading); (2) mathematics; and (3) knowledge of the social and physical world, referred to as "general knowledge." The assessment did not ask the children to write anything or to explain their reasoning; rather, they used pointing or verbal responses to complete the tasks. The data were collected using a computer-assisted interviewing methodology. The assessment included the use of a small easel with pictures, letters of the alphabet, words, short sentences, numbers, or number problems. A brief description of the three components of the cognitive assessment follows.

Language and Literacy. The language and literacy (reading) assessment included questions designed to measure basic skills (print familiarity, letter recognition, beginning and ending sounds, rhyming sounds, 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 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 that comprised the earlier levels as well. These five levels were: (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 words by sight; and (5) reading word in context.

Mathematical Thinking. The mathematics assessment items were 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. Each of the mathematics assessment forms 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 mathematics assessment could also be grouped into five-level proficiency levels or progression of skills, though the math clusters were less homogeneous in content than the reading clusters. The clusters of math items included: (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 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.

General Knowledge. The general knowledge assessment consisted of science and social studies material. The science items measure two broad classes of science competencies: a) conceptual understanding of scientific facts, and b) skills and abilities to form questions about the natural world, to try to answer them on the basis of the tools and the evidence collected, and to communicate answer and how the answers were obtained. Social studies material included questions relating to history/government, culture, geography, and economics. The assessment items drew on children's experiences with their environment, and many questions related to more than one of the categories.

The subject matter of the assessment of general knowledge was too diverse and insufficiently ranked or graded to permit formation of a set of proficiency levels. A score was calculated to represent each child's breadth and depth of understanding of the world around them. This assessment 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 also measured.

Physical Components. There were two parts to the physical component of the child assessment, psychomotor and antropometric. The assessor administered a psychomotor measure during

the baseline data collection in the fall of kindergarten and recorded each child's height and weight. The physical component included a spiral bound psychomotor booklet, psychomotor instructions and score sheet, a height and weight recording form, the Shorr Board (for measuring height), and a digital bathroom scale. Materials used for measuring fine motor skills included ten wood blocks, a pencil, and two pieces of plain white paper. The psychomotor assessment was divided into measuring fine and gross motor skills. Fine motor skills were assessed by having each child use building blocks to replicate a model, copy forms (e.g., an asterisk, a square) on paper, and draw a person. The child was asked to skip, hop on one foot, walk backward, and stand on one foot to assess gross or large motor skills. To measure physical growth and development, children's height and weight were measured. Table 2-3 displays the major domains measured during the child direct assessments in the fall and spring of kindergarten.

Table 2-3. Direct child assessment

Direct child assessment	Fall-Kindergarten	Spring-Kindergarten
Language screener (OLDS)	X	/
Reading (language and literacy)	X	X
General knowledge (science and social studies)	X	X
Mathematical thinking	X	X
Psychomotor	X	
Height and weight	X	X

Note: The columns to the right of each construct correspond to the waves of questionnaire administration. Waves that included the construct are marked with an "X." A "/" indicates that the OLDS was given to students new in the spring, or who did not pass the cut score in the English version during the fall OLDS administration.

Child Assessment Spring

In spring of the base year, the children who participated in the fall data collection were assessed a second time. In addition, children who were not assessed in the fall were assessed for the first time. These children included those sampled in converted schools (see chapter 5, section 5.2 for more detail). Assessments began at least 12 weeks before the end of the school year. Table 2-4 displays the

Table 2-4. Flow of the spring direct child assessment

Home Language	English OLDS	Spanish OLDS	Warm-up Booklet	Reading	Math	General Knowledge	Height/ Weight
English			✓	✓	✓	✓	✓
Other	Score **at or Above** Cutpoint in fall-kindergarten		✓	✓	✓	✓	✓
	✓ Score at or **Above** Cutpoint		✓	✓	✓	✓	✓
	✓ Score **Below** Cutpoint Speaks Spanish	✓	✓ **Spanish**		✓ **Spanish**		✓
	✓ Score **Below** Cutpoint Doesn't Speak Spanish						✓

of the spring direct child assessments. The same assessment materials used in the fall were used with the following exceptions:

1. The psychomotor assessment was not administered in the spring-kindergarten assessment and beyond.

2. If the sampled child passed the OLDS in fall-kindergarten, he or she was automatically routed by the computer program to take the assessment in English and did not have to retake the OLDS. If the child did not score above the cut point for the OLDS in fall-kindergarten, he or she took the OLDS again in spring-kindergarten and was routed according to the new spring-kindergarten OLDS score.

3. If the child did not take the assessment at all in fall-kindergarten or was a newly sampled child in spring-kindergarten, he or she was routed according to his or her home language, just as the children were in fall-kindergarten.

2.2 Parent Interview

The majority of parents participating in the base year data collection were interviewed in the fall of 1998 and again in the spring of 1999. Because more children were added to build school response rates, a group of parents completed their first interview in the spring of 1999. (see chapter 5, section 5.2). Parents or guardians were asked to provide important information about the sampled child, the home

environment, parent behavior (e.g., interactions with the child's teacher, activities with the sampled child), and family characteristics using a computer-assisted telephone interview (CATI) or computer-assisted personal interview (CAPI) for families without a telephone (see chapter 5, sections 5.4.4 and 5.6.4 for more details). Questions regarding family structure, child care use, household income, and child rearing practices were also included in the parent interview.

Fall Parent Interview

Typically the respondent for the parent interview was the mother of the child; however, the respondent could be a father, stepparent, adoptive parent, foster parent, grandparent, another relative, or nonrelative guardian. The respondent had to be knowledgeable about the child's care and education, be 18 years of age or older, and be living in the household with the child. In fall-kindergarten, respondents for the parent interview were selected according to the following order of preference:

1. The child's mother;
2. Another parent or guardian; and
3. Another household member.

The parent interview was conducted primarily in English, but provisions were made to interview parents who spoke other languages. The questionnaire was translated into Spanish, which was then printed on hardcopy. Bilingual interviewers were trained to conduct the parent interview in either English or Spanish. If the interview was conducted in Spanish, the interviewer used the hardcopy questionnaire and then entered respondents' answers into the CAI program. The parent interview was also translated into Chinese, Lakota, and Hmong languages and administered using the same data collection procedures as were used with Spanish speaking parents.

Topics addressed in the fall parent interview included a roster of current household members, family socio-demographic characteristics, languages spoken in the child's home, child care arrangements (currently and previously used), child's physical functioning, home activities, parent education and employment, and receipt of public assistance.

Parents were also asked to tell how often a student exhibited certain social skills and behaviors. The social rating scale (SRS) has five scales: approaches to learning, self-control, social interaction, impulsive/overactive, and sad/lonely.

See chapter 3, section 3.6.2 for variable names, ranges, means, and standard deviations for these scales.

- The **Approaches to Learning** scale (Parent SRS) includes six items that rate how often a child shows eagerness to learn, interest in a variety of things, creativity, persistence, concentration, and responsibility.
- The **Self-Control** scale (Parent SRS) has five items that indicate children's ability to control their behavior. It includes items that are worded positively as well as negative behaviors that are reverse coded (e.g., frequency with which a child fights, argues, throws tantrums, or gets angry).
- The **Social Interaction** scale (Parent SRS) asks about children's interactions with peers and adults. The three items address children's ease in joining in play, ability to make and keep friends, and positively interacting (comforting, helping) with peers.
- The **Impulsive/Overactive** scale (Parent SRS) has two items that ask about children's impulsivity and activity level.
- The **Sad/Lonely** scale (Parent SRS) has four items that ask parents about children's problems with being accepted and liked by others, sadness, loneliness, and low self-esteem.

Table 2-5 shows the broad content areas addressed in the parent interview and the point of data collection.

Spring Parent Interview

The content of the parent interview in spring-kindergarten differed from the fall interview. To avoid redundancy and increased respondent burden, many questionnaire topics were split between fall and spring data collection points. In spring-kindergarten, the parent interview included updating the household roster, parent's participation in activities in the child's school, parent's attitudes toward child-rearing, parent's psychological well-being and health, and the household's food situation. In addition, items from the SRS scales were repeated in the spring parent interview.

Table 2-5 shows the overall structure of the interview and distribution of topics across the two base year data collection points.

Table 2-5. ECLS-K parent interview by major content topics and point of data collection

Parent Questionnaire	Fall-Kindergarten	Spring-Kindergarten
Family structure	X	X
Demographics	X	X
Household roster	X	X
Marital status	X	X
Immigration status		X
Primary language(s) spoken in home	X	
Parent's involvement with child's school		X
Child care	X	
Arrangements with relatives	X	
Arrangements with nonrelatives	X	
Head Start attendance year before kindergarten	X	
Year before kindergarten child care arrangements	X	
Child's health and well-being	X	X
Birth weight	X	
Physical functioning	X	
Services for children with special needs	X	
Social skills rating	X	X
Home environment and cognitive activities	X	X
Frequency of literacy activities	X	X
Computer use		X
Television viewing		X
Parental educational expectations for child	X	
Neighborhood safety		X
Parent education	X	
Parent employment	X	
Parent income		X

Table 2-5. ECLS-K parent interview by major content topics and point of data collection (continued)

Parent Questionnaire	Fall-Kindergarten	Spring-Kindergarten
Welfare and other public assistance use	X	X
Parent/child interaction		X
Parent discipline		X
Parent health and emotional well-being		X
Relationships and social support	X	X
Marital satisfaction		X
Background data for fall-kindergarten nonresponding parents		X
Child's physical functioning/birth weight of child		/
Home language		/
WIC benefits		/
Parent education/mother's employment history		/
Prekindergarten Head Start attendance		/
Services for children with special needs		/

Note: The columns to the right of each construct correspond to the waves of questionnaire administration. Waves that included the construct are marked with an "X." Content areas asked in spring only to new parent participants are marked with a "/."

In the spring parent interview, households were routed to one of two questionnaire paths, either as a new household or as a continuing household that participated in fall data collection. Parents who completed the fall-kindergarten parent interview were classified as continuing households. Parents who were selected to participate in fall-kindergarten, but either refused, did not finish the interview, could not be located, or were unavailable during the field period were also classified as new households in the spring. Parents of children in schools converted in 1999 were classified as new households. The group of parents entering the study in spring 1999 were asked the same questions in the spring interview as those who participated in the fall. These respondents were also asked an additional set of questions to replicate some of the items included in the fall interview.

The rules for respondent selection varied by questionnaire path. Respondents in continuing households were selected according to the following order of preference:

1. The fall-kindergarten respondent;
2. The child's mother;

3. Another parent or guardian; and

4. Another household member.

Respondents in new households were selected according to the following order of preference:

1. The child's mother;

2. Another parent or guardian; and

3. Another household member.

2.3 Teacher Questionnaires

Each kindergarten teacher received a self-administered questionnaire consisting of three distinct parts. The first section, part A, asked about the teacher's class and classroom characteristics. It was designed to collect data about the composition and demographics of the children in the class. Part B addressed more specific questions on class organization, typical class activities, and evaluation methods, as well as teacher views on kindergarten readiness, school environment, and overall school climate. Background questions about the teacher were also included in this section.

Part C asked teachers to report about the sampled children in their classrooms. Teachers were asked to respond to 20 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, general knowledge, and mathematical thinking. For example, some questions asked if the child used complex sentence structure, demonstrated early writing behaviors, formed explanations based on observations, or solved problems involving numbers. Part C also included questions from the SRS that collected data on five areas of children's social skills.

2.3.1 Content of the ARS

There are three scales of the ARS: language and literacy, general knowledge, and mathematics. Each of these is described below. 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 based on the teacher's past observation and experience with the child. Unless otherwise noted, the fall and spring ARS have the same number of items for each concept.

- The **Language and Literacy** section of the ARS asks teachers to rate each child's proficiency in speaking (1 item), listening (1 item), early reading (3 items fall, 4 items spring), and writing (1 item fall, 2 items spring). In addition, teachers rate the child's computer literacy (1 item).
- The **General Knowledge** section of the ARS asks teachers to rate each child's proficiency in social studies (1 item fall, 2 items spring) and science (3 items).
- In the **Mathematics** section, teachers rate each child on one item on each of five skills: concept of numbers, solving number problems, use of math strategies, data analysis (graphing), and measurement.

In all sections, the teacher rated the child's skills, knowledge and behaviors on a scale from "Not Yet" to "Proficient" (see table 2-6). If a skill, knowledge, or behavior had not been introduced into the classroom, the teacher coded that item as N/A (Not applicable).

See chapter 3, section 3.3 for scale scores, value ranges, means, and standard deviations for the ARS.

Table 2-6. ARS response scale

Not yet:	Child has not yet demonstrated skill, knowledge, or behavior.
Beginning:	Child is just beginning to demonstrate skill, knowledge, or behavior but does so very inconsistently.
In progress:	Child demonstrates skill, knowledge, or behavior with some regularity but varies in level of competence.
Intermediate:	Child demonstrates skill, knowledge, or behavior with increasing regularity and average competence but is not completely proficient.
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.

2.3.2 Teacher SRS

Teachers rated individual students as part of a self-administered 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). See chapter 3, section 3.6.1 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.

- The **Self-Control** (Teacher SRS) Scale 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 five **Interpersonal Skills** (Teacher SRS) items 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 two problem behavior scales reflect behaviors that may interfere with the learning process and with the child's ability to interact positively in the classroom.

- **Externalizing Problem Behaviors** (Teacher SRS) include acting out behaviors. Five items on this scale rate the frequency with which a child argues, fights, gets angry, acts impulsively, and disturbs ongoing activities.

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

This measure is adapted with permission from the instrument Elementary Scale A ("How Often?") (Gresham, F. M. and Elliott, S.N., 1990).

Fall-Kindergarten Teacher Questionnaires

Part A of the fall-kindergarten teacher questionnaire was given to all kindergarten teachers, regardless of whether they taught a sampled child. It included items characterizing the classroom and

students. Part B included questions regarding class organization and evaluation, typical class activities, and views on kindergarten readiness. The part C questionnaire asked teachers to complete an ARS and SRS rating for each sampled child in his or her classroom.

Spring-Kindergarten Teacher Questionnaires

Similar to the fall instrument, the spring questionnaire was also divided into three sections. The content of the items varied between fall and spring instruments to ease respondent burden. Part A was again completed by all kindergarten teachers in the school, regardless of whether they taught a sampled child. This section asked about the characteristics of the kindergarten class and the children in the class. The spring part B questionnaire included some, but not all, of the questions used in the part B fall questionnaire. Part B of the teacher questionnaire was completed only by kindergarten teachers new to the study. This questionnaire was similar to the part B questionnaire completed in fall. It was used to gather information about teacher background, instructional practices, views on kindergarten readiness, and the teaching climate of the school. The teacher part C questionnaire was completed by kindergarten teachers who taught sampled children. This questionnaire, similar to the part C questionnaire completed in fall, asked teachers to provide ratings of each sampled child in their classroom in academic and social areas. The spring part C also included a section titled "student information" that asked for overall academic and physical activity comparisons of the sampled child with the other children in the class and student support service participation. Table 2-7 shows the overall structure of the teacher questionnaire and the distribution of topics among the fall and spring data collection points.

2.4 Special Education Teacher Questionnaires

The special education teacher questionnaires were new in the spring. ECLS-K supervisors reviewed accommodation and inclusion information for children who received special education services. During the preassessment visit, the field supervisors specified primary special education teachers of sampled children and listed special education staff working with each child (e.g., speech pathologists, reading instructors, and audiologists). These questionnaires were given to special education teachers who taught sampled children. If a child received special education services from more than one special education teacher, a field supervisor determined the child's primary special education teacher.

Table 2-7. Teacher questionnaire

Teacher Questionnaire	Fall-Kindergarten	Spring-Kindergarten
Part A		
Description of class—age, race, and gender distribution	X	
Class organization	X[1]	X
Children with special needs		X
Classroom aides		X
Class assignment and grouping		X
Types of activity/ interest areas	X	
Class activities	X[1]	X
Instructional time in different subjects		X
Types of materials and activities		X
Child vs. teacher initiated activities	X[1]	
Parent involvement		X
Professional development		X
Part B		
Evaluation and grading practices of students	X	/
Sharing progress information with parents		/
Teachers' views on school readiness	X	/
Perceptions about school climate	X	/
Perception of personal influence on policies and classroom planning	X	
Teacher demographic information	X	/
Teacher experience and education	X	/
Job satisfaction	X	/
Transition to school activities	X	/

Table 2-7. Teacher questionnaire (continued)

Teacher Questionnaire	Fall-Kindergarten	Spring-Kindergarten
Part C		
Indirect child cognitive evaluation by teacher	X	X
Language and literacy, mathematics, general knowledge	X	X
Teacher evaluation of target child's social skills	X	X
Sampled child additional information		X
Participation in special services and programs		X
Target child's overall academic skills and physical activity		X
Reading group participation		X

[1] Collected in part B for fall-kindergarten.

Note: The columns to the right of each construct correspond to the waves of questionnaire administration. Waves that included the construct are marked with an "X." Content areas asked in spring only to new teacher participants are marked with a "/."

The primary special education teacher was defined as:

- The teacher who managed the child's Individualized Education Plan (IEP), or
- The teacher who spent the most amount of time providing special education services to the child, or
- The teacher who is most knowledgeable about the child's special needs and equipment.

Items in the special education teacher 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.

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. Table 2-8 provides a summary of the content areas addressed in the special education teacher questionnaire.

Table 2-8. Special education teacher questionnaire[1]

Teachers of Sampled Students with IEPs Questionnaire	Spring-Kindergarten
Part A	
Teacher's gender	X
Teacher's age	X
Teacher's race-ethnicity	X
Teaching experience	X
Educational background	X
Special education teacher background	X
Location of service provision	X
Student load/week	X
Part B	
Disability category	X
IEP goals for the school year	X
Extent of services	X
Types of services provided for the year	X
Primary placement	X
Teaching practices, methods, and materials	X
Assistive technologies used by child	X
General education goals, expectations and assessments	X
Collaboration/communication with child's general education teacher	X
Frequency of communicating with child's parents	X
Receipt of formal evaluations in the past year	X

[1] Data collected only in the spring.

2.5 Adaptive Behavior Scale

The Adaptive Behavior Scale was completed for all sampled children who were identified in spring-kindergarten as excluded from the direct child assessment. A child was excluded from the assessment if he or she needed the assessment administered in Braille, enlarged print, or sign language, or if the child's IEP specifically prohibited the child from taking standardized assessments. This questionnaire was completed by the child's primary special education teacher and asked the teacher to provide ratings of the sampled child in three domains: independent functioning, language development, and numbers and time.

2.6 School Administrator Questionnaire

The school principal, administrator, or headmaster was asked to complete the school administrator questionnaire in the spring of 1999 (shown in table 2-9). 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 requested mainly factual information about each school and the programs offered at the school. These sections could be completed by either a principal or a designee who was able to provide the requested information. The school's principal was asked to complete the remaining two sections concerning his or her background and evaluations of the school climate. This questionnaire was administered only in the spring of the base year data collection.

2.7 School Facilities Checklist

ECLS-K supervisors completed the facilities checklist. The facilities checklist collects information about the (1) availability and condition of the selected schools, (2) presence and adequacy of security measures, (3) presence of environmental factors that may affect the learning environment, and (4) overall learning climate of the school. (See chapter 5, section 5.6.5 for more detail on the collection of these data.)

2.8 School Records Abstract Form

School staff completed the student records abstract form. This instrument was used to obtain information about the child's attendance record, report card, and use of an individualized education plan. Information about the type of language or English proficiency screening that the school used and whether the child participated in Head Start prior to kindergarten was also retrieved from the students' records. (See chapter 5, section 5.6.5 for more detail on the collection of these forms.)

Table 2-9. School administrator questionnaire[1]

School Questionnaire	Spring-Kindergarten
School characteristics	X
Type of school	X
Admission requirements	X
School size	X
Student characteristics	X
Race-ethnicity of students	X
Children eligible for special services	X
Types of kindergarten programs	X
School facilities and resources	X
Equipment	X
Community characteristics and school safety	X
Teaching and other school staff characteristics	X
Range of salary paid to teachers	X
Race-ethnicity of staff	X
School policies and programs	X
Assessments, testing, and retention	X
School-family-community connections	X
Programs and activities for families	X
Parent involvement and participation	X
Programs for special populations	X
ESL and bilingual education	X
Special education	X
Gifted and talented	X
Principal characteristics	X
Gender, race-ethnicity, age of principal	X
Experience and education	X
School governance and climate	X
Goals and objectives for kindergarten teachers	X
School functioning and decisionmaking	X

[1] Collected only in the spring.

2.9 Salary and Benefits Questionnaire

The salary and benefits questionnaire collects information on the base salary, merit pay, and benefit pay of teachers and principals. The salary and benefits questionnaire was completed at the school or district level, generally by the school or district business administrator or by a private school administrator or headmaster. The teacher salary and benefits questionnaire can be used to develop child-level school resource variables that can be linked to child outcomes. The interest in payroll information stems from the fact that salaries and benefits constitute approximately 80 percent of all current expenditures in school budgets. Although instructional expenditures are 61.8 percent of total current expenditures, salaries and benefits for instruction alone constitute 56 percent of total current expenditures, and 91.6 percent of all instructional expenditures.[4] These data provide an opportunity to learn more about how resources are allocated and used in schools and how those spending decisions impact children's achievement.

2.10 Head Start Verification

The goal of this part of the study is to verify that parent and school reports of children's Head Start participation. Respondents to the fall-kindergarten parent interview were asked in a series of questions about childcare outside the home whether or not the sampled child had ever attended Head Start. If the response was "Yes," the respondent was asked whether or not the sampled child attended Head Start in the year before kindergarten. Information on the name and location of the Head Start facility was matched by the interviewer against a database of Head Start centers. Similar data were collected in the spring-kindergarten student record abstract. In the student record abstract, one item asked whether or not the sampled child had attended a Head Start center before entering kindergarten. If the answer was "Yes," then the school staff person was asked to record the name, address, and telephone number of the Head Start center and the name of the Head Start director. The next step was to verify that the centers reported by the respondents were Head Start centers and that the child did attend the center in the 1997-98 school year. Chapter 5, section 5.6.7, describes how the data were collected.

[4] U.S. Department of Education, National Center for Education Statistics. *Digest of Education Statistics*, 1999, NCES 2000-031, Washington DC 2000. Pp. 185, table 168.

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

Several types of scores are used in the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K) to describe children's cognitive and social development during their kindergarten year. These scores are for the direct cognitive assessment, the academic rating scale (ARS), the psychomotor assessment, and the social rating scale (SRS). Descriptions of the scores for each assessment or scale are provided below, along with variable names, variable descriptions, and value ranges found in the ECLS-K data files. Guidelines for when to use each cognitive test score are also provided in this chapter.

3.1 Direct Cognitive Assessment

The direct cognitive assessment contained items on reading, mathematics, and general knowledge in the fall and spring of kindergarten. In each subject area, children received a 12 to 20 item routing test. Performance on the routing items guided the selection of one of several alternative second-stage forms. The second-stage form contained items of appropriate difficulty for the level of ability indicated by the routing items. There are five different types of scores that can be used to describe children's performance on the direct cognitive assessment: (1) number right scores and (2) item response theory (IRT) scores, which measure children's performance on a set of test questions with a broad range of difficulty; (3) standardized scores, which report children's performance relative to their peers; (4) criterion-referenced proficiency level and (5) proficiency probability scores, which evaluate children's performance with respect to subsets of test items that mark specific skills. See chapter 2, section 2.1 for a description of the ECLS-K assessment battery.

Table 3-1 shows the types of scores, variable names, descriptions, ranges means, and standard deviations for the direct cognitive assessment.

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 tests that are the same for all children. However,

when these scores are for tests that vary in average difficulty, they are not comparable to each other. For example, a student who took the middle difficulty mathematics second-stage form would probably have gotten more questions correct if he or she had taken the easier low form and fewer correct if the more difficult high form had been administered. For this reason, raw number right scores are reported in the database only for the first stage (routing) tests, which were the same for all children. Each routing test consisted of sets of items spanning a wide range of skills. For example, the reading routing test had four questions each on letter recognition, recognizing beginning sounds, recognizing ending sounds, reading simple sight words, and selecting words in the context of a sentence. An analyst might use the routing test number right scores to report actual performance on this particular set of tasks.

See table 3-1 for the variable names, ranges, means, and standard deviations for the number right scores.

3.1.2 IRT Scale Scores

Scores based on the full set of test items were calculated using IRT procedures. IRT made it possible to calculate scores that could be compared regardless of which second-stage form a child took. IRT uses the pattern of right, wrong, and omitted responses to the items actually administered in a test 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 test, plus a core set of items shared among the different second-stage forms, 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 test forms had been administered.

IRT has several other advantages over raw number-right scoring. By using the overall pattern of right and wrong responses 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, a correct difficult item is, in effect, assumed to have been guessed. 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, in effect, 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 test questions. Finally, IRT scoring makes possible longitudinal measurement of gain in achievement over time, even though the tests administered are not identical at each point. The common

Table 3-1. Direct cognitive assessment: types of scores, variable names, descriptions, ranges, means, and standard deviations[1]

Type of Score	Variable	Description	Range of Values	Mean	Standard Deviation
Number Right	C1RROUNR	C1Reading Routing Test, - Number Right	0 - 20	5.7	3.9
	C1MROUNR	C1 Mathematics Routing Test, - Number Right	0 - 16	4.4	2.9
	C1GROUNR	C1 Genl Knowledge Routing Test, - # Right	0 - 12	4.7	2.9
	C2RROUNR	C2Reading Routing Test, - Number Right	0 - 20	9.8	4.2
	C2MROUNR	C2 Mathematics Routing Test, - Number Right	0 - 16	7.1	3.3
	C2GROUNR	C2 Genl Knowledge Routing Test, - # Right	0 - 12	6.2	3.0
IRT	C1RSCALE	C1 Reading IRT Scale Score	0.0 - 72.0	22.0	8.3
	C1MSCALE	C1 Math IRT Scale Score	0.0 - 64.0	19.1	7.2
	C1GSCALE	C1 General Knowledge IRT Scale Score	0.0 - 51.0	22.1	7.4
	C2RSCALE	C2 Reading IRT Scale Score	0.0 - 72.0	31.6	10.3
	C2MSCALE	C2 Math IRT Scale Score	0.0 - 64.0	27.1	8.8
	C2GSCALE	C2 General Knowledge IRT Scale Score	0.0 - 51.0	26.8	7.8
Standardized Score	C1RTSCOR	C1 Reading T-Score	0.0 - 90.0	50.0	10.0
	C1MTSCOR	C1 Math T-Score	0.0 - 90.0	50.0	10.0
	C1GTSCOR	C1 General Knowledge T-Score	0.0 - 90.0	50.0	10.0
	C2RTSCOR	C2 Reading T-Score	0.0 - 90.0	50.0	10.0
	C2MTSCOR	C2 Math T-Score	0.0 - 90.0	50.0	10.0
	C2GTSCOR	C2 General Knowledge T-Score	0.0 - 90.0	50.0	10.0
Proficiency Level Score	C1RPROF1	C1 Prof 1 - Letter Recognition	0 - 1	0.65	0.8
	C1RPROF2	C1 Prof 2 - Beginning Sounds	0 - 1	0.30	0.46
	C1RPROF3	C1 Prof 3 - Ending Sounds	0 - 1	0.18	0.38
	C1RPROF4	C1 Prof 4 - Sight Words	0 - 1	0.04	0.20

[1] See chapter 7, section 7.3 for variable naming conventions.

Table 3-1. Direct cognitive assessment: types of scores, variable names, descriptions, ranges, means, and standard deviations (continued)

Type of Score	Variable	Description	Range of Values	Mean	Standard Deviation
	C1RPROF5	C1 Prof 5 - Word in Context	0 - 1	0.26	0.44
	C1MPROF1	C1 Prof 1 - Count, Number, Shape	0 - 1	0.90	0.31
	C1MPROF2	C1 Prof 2- Relative size	0 - 1	0.55	0.50
	C1MPROF3	C1 Prof 3 - Ordinality, sequence	0 - 1	0.20	0.402
	C1MPROF4	C1 Prof 4 -Add/Subtract	0 - 1	0.04	0.197
	C1MPROF5	C1 Prof 5 - Multiply/Divide	0 - 1	0.02	0.131
	C2RPROF1	C2 Prof 1 - Letter Recognition	0 - 1	0.92	0.277
	C2RPROF2	C2 Prof 2 - Beginning Sounds	0 - 1	0.70	0.46
	C2RPROF3	C2 Prof 3 - Ending Sounds	0 - 1	0.50	0.50
	C2RPROF4	C2 Prof 4 - Sight Words	0 - 1	0.14	0.35
	C2RPROF5	C2 Prof 5 - Word in Context	0 - 1	0.24	0.43
	C2MPROF1	C2 Prof 1- Count, Number, Shape	0 - 1	0.95	0.21
	C2MPROF2	C2 Prof 2 - Relative Size	0 - 1	0.82	0.38
	C2MPROF3	C2 Prof 3 - Ordinality, Sequence	0 - 1	0.54	0.50
	C2MPROF4	C2 Prof 4 - Add/Subtract	0 - 1	0.17	0.38
	C2MPROF5	C2 Prof 5 - Multiply/Divide	0 - 1	0.04	0.19
Proficiency Probability Score	C1RPROB1	C1 Prob 1 - Letter Recognition	0.0 - 1.0	0.66	0.43
	C1RPROB2	C1 Prob 2 - Beginning Sounds	0.0 - 1.0	0.29	0.36
	C1RPROB3	C1 Prob 3 - Ending Sounds	0.0 - 1.0	0.16	0.27
	C1RPROB4	C1 Prob 4 - Sight Words	0.0 - 1.0	0.02	0.13
	C1RPROB5	C1 Prob 5 - Word in Context	0.0 - 1.0	0.01	0.08
	C1MPROB1	C1 Prob 1 - Count, Number, Shape	0.0 - 1.0	0.92	0.20
	C1MPROB2	C1 Prob 2 - Relative Size	0.0 - 1.0	0.54	0.38
	C1MPROB3	C1 Prob 3 - Ordinality, Sequence	0.0 - 1.0	0.20	0.31
	C1MPROB4	C1 Prob 4 - Add/Subtract	0.0 - 1.0	0.04	0.12

Table 3-1. Direct cognitive assessment: types of scores, variable names, descriptions, ranges, means, and standard deviations (continued)

Type of Score	Variable	Description	Range of Values	Mean	Standard Deviation
	C1MPROB5	C1 Prob 5 - Multiply/Divide	0.0 - 1.0	0.00	0.04
	C2RPROB1	C2 Prob 1 - Letter Recognition	0.0 - 1.0	0.93	0.23
	C2RPROB2	C2 Prob 2 - Beginning Sounds	0.0 - 1.0	0.70	0.36
	C2RPROB3	C2 Prob 3 - Ending Sounds	0.0 - 1.0	0.50	0.36
	C2RPROB4	C2 Prob 4 - Sight Words	0.0 - 1.0	0.13	0.27
	C2RPROB5	C2 Prob 5 - Word in Context	0.0 - 1.0	0.04	0.16
	C2MPROB1	C2 Prob 1 - Count, Number, Shape	0.0 - 1.0	0.99	0.08
	C2MPROB2	C2 Prob 2 - Relative Size	0.0 - 1.0	0.84	0.26
	C2MPROB3	C2 Prob 3 - Ordinality, Sequence	0.0 - 1.0	0.53	0.40
	C2MPROB4	C2 Prob 4 - Add/Subtract	0.0 - 1.0	0.16	0.26
	C2MPROB5	C2 Prob 5 - Multiply/Divide	0.0 - 1.0	0.02	0.10
	C1RPRINT	C1 Print Familiarity	0 - 3	1.8	1.1
	C2RPRINT	C2 Print Familiarity	0 - 3	2.3	0.9

items present in the routing test and in overlapping second-stage forms allow the test scores to be placed on the same scale, even as the two-stage test design adapts to children's growth over time.

The IRT scale scores in the database represent estimates of the number of items students would have answered correctly if they had taken all of the 72 questions in the first- and second-stage reading forms, the 64 questions in all of the mathematics forms, and the 51 general knowledge items. These scores are not integers because they are probabilities of correct answers, summed over all items in the pool. Gain scores may be obtained by subtracting the estimated number right at time 1 from the estimated number right at time 2. (Note that scores for different subject areas are not comparable to each other because they are based on different numbers of test questions; that is, 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.)

See table 3-1 for variable names, ranges, means, and standard deviations for the IRT scale scores.

3.1.3 Standardized Scores (T-Scores)

T-scores provide norm-referenced measurements of achievement, that is, estimates of achievement level *relative to the population as a whole*. A high T-score mean for a particular subgroup indicates that the group's performance is high in comparison to other groups. It does not mean that group members have mastered a particular set of skills, only that their mastery level is greater than a comparison group. Similarly, a change in T-score means over time reflects a change in the group's status with respect to other groups. In other words, they provide information on *status compared to children's peers*, while the IRT scale scores and proficiency scores represent *status with respect to achievement on a particular criterion set of test items*. The T-scores can 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 fall- and spring-kindergarten. For example, a T-score of 55 (C1RTSCOR) represents a

reading achievement level that is one-half of a standard deviation higher than the mean for the fall-kindergarten population represented by the tested sample of ECLS-K participants.

See table 3-1 for variable names, ranges, means, and standard deviations for the standardized (T) scores.

3.1.4 Proficiency Scores

Proficiency scores 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. Since the ECLS-K direct cognitive child assessment was a two-stage design (where not all children were administered all items), information on children's specific proficiencies are presented in two ways: proficiency scores (raw scores) and proficiency probability scores (IRT-based scores). In most situations, analysts use the proficiency probability scores in analyzing children's specific reading and mathematics knowledge and skills. Clusters of assessment questions having similar content and difficulty were included at several points along the score scale of the reading and mathematics assessments. No proficiency scores were computed for the general knowledge test, because the questions did not follow a hierarchical pattern. The following proficiencies were identified in the reading and mathematics assessments.

Reading:

- **Letter recognition:** identifying upper- and lowercase letters by name;
- **Beginning sounds:** associating letters with sounds at the beginning of words;
- **Ending sounds:** associating letters with sounds at the end of words;
- **Sight words:** recognizing common words by sight; and
- **Comprehension of words in context:** reading words in context.

Mathematics:

- **Number and Shape:** identifying some one-digit numerals, recognizing geometric shapes, and one-to-one counting of up to ten objects;
- **Relative Size:** reading all single-digit numerals, counting beyond ten, recognizing a sequence of patterns, and using nonstandard units of length to compare objects;

- **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;
- **Addition/Subtraction:** solving simple addition and subtraction problems; and
- **Multiplication/Division:** solving simple multiplication and division problems and recognizing more complex number patterns.

Clusters of items provide a more reliable test of proficiency than do single items because of the possibility of guessing; it is very unlikely that a student who has not mastered a particular skill would be able to guess enough answers correctly to pass a four-item cluster. 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 have indicated nonmastery at higher levels. Only a very small percentage of students in fall- and spring-kindergarten 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 both fall- and spring-kindergarten, 5.5 percent of the children did not follow the model for reading and 6.6 percent of the children did not follow the model for mathematics. This does not necessarily indicate a different order or learning for these children - since most of the proficiency-level items were multiple choice, many of these reversals are due to children guessing.

Proficiency level (dichotomous) scores and proficiency probability (continuous) level scores are two types of proficiency scores used in the ECLS-K. The following is a description of these scores.

Proficiency Level Scores (Dichotomous)

The proficiency level scores reflect the children's raw ECLS-K direct cognitive assessment scores. These scores are intended for very distinct kinds of analysis. Most analysts should use the proficiency probability scores.

For each proficiency level, a score of one was assigned to children who correctly answered at least three of the four items in the cluster, and a score of zero was given if at least two items were incorrect or don't know. If children did not answer enough items for pass or fail to be determined for a particular cluster, a pass/fail score was assigned only if the remaining proficiency scores indicated a level that was unambiguous. That is, a "fail" could be inferred if there were easier cluster(s) that had been failed and no higher cluster passed; and a "pass" was imputed if harder cluster(s) were passed and no

easier one failed. In the case of ambiguous (e.g., pass, blank, fail, where the blank could be either a pass or a fail) or contradictory (e.g., fail, blank, pass) patterns, missing cluster scores were left blank.

Averaging a population subgroup's zero and one scores for a particular proficiency cluster results in an estimate of the proportion of children in the subgroup who had mastered the material at that level. The difference between this average at two points in time represents the proportion of children who attained mastery during that time period. These scores are not designed to extrapolate to the entire population of kindergarten children. These scores simply show - of the children who took the items represented by the level, how many passed the level. For example, these scores would be used in an analysis involving only the population of children that comprehend words in context. The analyst would analyze data on the subset of children that received a value of 1 on reading proficiency level 5 (comprehension of words in context), and would base the analysis only on this group of children.

See table 3-1 for variable names, ranges, means, and standard deviations for the proficiency level scores.

Proficiency Probability Scores (Continuous)

The proficiency probability scores are based on the same clusters of items as the proficiency level scores but differ from them in several ways. They are continuous rather than dichotomous and can take on any value between zero and one. They are estimates based on overall performance rather than counts of actual item responses. They are also estimates for all children with scorable test data, not only for the ones who were administered the test items in the cluster.

Due to the two-stage format of the cognitive assessment battery, not all children received all items. An IRT model was employed to produce proficiency probability scores, which reflect the probability that a child would have passed a proficiency level. 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. Because the proficiency probabilities were based on overall performance, they could be calculated for all children who had scorable test data, not just those with relatively complete sets of responses to the necessary item clusters.

The proficiency probability scores can be averaged to produce estimates of mastery rates within population subgroups. These continuous measures can provide a closer 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 are making their largest gains in achievement during a particular time interval. Thus, students' school experiences can be related to improvements in specific skills.

Proficiency level scores differ from proficiency probability scores. Proficiency level scores are based on the items administered to each child. Since not all children received the same items because of the two-stage assessment design, these scores only represent those children who were administered the items. The use of proficiency level scores to estimate the total population of children mastering a specific proficiency level is not recommended, because stopping rules within the test resulted in missing data for the lower-achieving children. The proficiency probability scores are more suited for estimating the total population of children mastering specific proficiency scores.

See table 3-1 for variable names, ranges, means, and standard deviations for the proficiency probability scores.

Familiarity with Conventions of Print

Some items from the child assessment measured children's familiarity with conventions of print but were not part of the set of proficiency scores because they did not fit the hierarchical pattern. The score for these questions was obtained by counting the number of correct answers (0-3) for the following three items, administered while the child was looking at an illustrated story.

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

These items were part of the reading score calculations in the direct cognitive assessment but were not part of the hierarchical set of proficiency and proficiency probability scores because they did not fit the proficiency scoring pattern. The proficiency levels assume that mastery of a higher level usually means that the child has mastered lower levels. This was not the case with conventions of print. Some

children scored high on conventions of print but could not recognize letters, while others had the reverse pattern. Thus, the score for familiarity with conventions of print is reported separately.

See table 3-1 for variable names, ranges, means, and standard deviations for the conventions of print scores.

3.1.5 Choosing the Appropriate Score for Analysis

Each of the types of scores described above 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 vs. achievement in specific skills;
- An indicator of status at a single point in time versus growth over time; and
- A criterion-referenced vs. norm-referenced interpretation.

IRT-Based Scores

The scores derived from the IRT model (IRT scale scores, T-scores, proficiency probabilities) are 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, are used to estimate a point on an ability continuum, and this ability estimate, theta, then provides 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 may also be used as longitudinal measures of growth, but it is important to remember that gains made at different points on the score scale have qualitatively different interpretations. For example, children who make gains in recognizing letters and letter sounds are learning very different things 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 gains in scale score points is most meaningful for groups that started with similar initial status. When initial status is very different, comparisons of scale score gains may be misleading because the skills being learned are qualitatively different, and comparisons of total number of points gained may be difficult to interpret.

- 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.

- **Proficiency probability scores**, although derived from the overall IRT model, are criterion-referenced measures of proficiency in **specific skills**. Because each proficiency score targets a particular narrow set of skills, they are ideal for studying the **details of achievement**, rather than the single summary measure provided by the 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 SES kindergarten children show 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 are making big gains in basic skills, but most have not yet made major gains in reading words and sentences. 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.

Non-IRT Based Scores

The routing test number right, proficiency level, and Conventions of Print scores do **not** depend on the assumptions of the IRT model. They are counts of actual number correct for specific sets of test items, rather than estimates based on patterns of overall performance.

- **Routing test number right scores** for the reading, math, and general knowledge assessments are based on 20, 16, and 12 items 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 test items, which was the same for all children who took the test.

- **Proficiency level scores** are based on the same sets of items as the proficiency probability scores, but are dichotomous, rather than continuous, measures of proficiency. They have a somewhat more intuitive interpretation than the probability scores, since they simply report whether children were able to answer correctly on at least three out of four actual test items in a cluster. Users of the proficiency level scores should be aware of possible bias due to missing data. Stopping rules employed in the administration of the tests to minimize stress on low-performing children results in substantial numbers of missing scores for the higher proficiency levels. Estimates based on variables with substantial amounts of missing data can be assumed to generalize to the whole sample only if "missing-ness" is unrelated to what the variable is measuring. This condition is called "MAR," or Missing-At-Random. The missing level four and five scores for low-achieving children are *not* missing-at-random, they were not administered based on performance. Interpretations of results based on these scores must take this into account. Similarly, missing data for the *lowest* math proficiency level are due to items in this cluster having been taken from the low second stage test, which was not taken by *high*-achieving children. Estimates based on proficiency level scores, without adjustments for missing data, would overstate the population performance at the high proficiency levels and understate performance at math level one.

- **Conventions of print scores**, like the proficiency level scores, are based on a count of the number correct for a particular set of items. Users may wish to relate this score to process variables to get a perspective that is somewhat different from that of the hierarchical levels of reading skills.

3.1.6 Reliabilities

Reliability statistics appropriate for each type of score were computed for each subject area, for fall- and spring-kindergarten assessments. 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 0.88 to 0.95, apply to all of the scores derived from the theta estimate, namely, the IRT scale scores, T-scores, and proficiency probabilities. Alpha coefficients for the routing test number correct ranged from 0.78 to 0.88. Split half reliabilities were computed for the item clusters that made up the dichotomous proficiency level scores and the conventions of print cluster. These reliabilities were higher for the reading clusters (0.60 to 0.83) than for the math levels (0.27 to 0.66). The difference in internal consistency statistics is due to the reading items being essentially replications of the same task, while the math items had a greater diversity of content.

Note that the split half reliabilities for the low level item clusters decreased from fall- to spring-kindergarten, while the reliabilities for the clusters at the upper end increased. This is a

consequence of changes in the variance of the cluster scores as children progressed in their development of skills. By spring-kindergarten, the vast majority of children had mastered the lowest proficiency levels, so the sample variance was low, resulting in lower reliability than for the previous fall. Conversely, the sample variance for the difficult tasks was very low in the fall-kindergarten, when most children had *not* mastered these skills, and the reliability rose as some children attained high-level proficiency by the spring, increasing the total variance. This effect is more pronounced for the math than for the reading clusters for two reasons. First, the math item clusters were more heterogeneous than the reading, in terms of content and difficulty. Second, the reading item clusters were based entirely on items from the routing test, which was taken by all children, while the lowest math cluster employed items from the low level second stage test as well. By spring-kindergarten, fewer than half of the test takers were routed to the low form, and this constrained variance in ability resulted in a lower reliability for children who had this score. Tables 3-2 and 3-3 present the reliability statistics for all of the test scores.

Table 3-2. Reliability of IRT-based scores and routing test number correct

	IRT-based Scores (Reliability of Theta)		Routing Test Number Correct (Alpha Coefficient)	
	Fall-k	Spring-k	Fall-k	Spring-k
Reading	0.93	0.95	0.86	0.88
Math	0.92	0.94	0.78	0.81
General Knowledge	0.88	0.89	0.79	0.79

Table 3-3. Split half reliability of item-cluster-based scores (proficiency level scores)

	Fall-k	Spring-k
Reading Level 1	0.83	0.79
Reading Level 2	0.76	0.76
Reading Level 3	0.72	0.76
Reading Level 4	0.78	0.77
Reading Level 5	0.60	0.69
Conventions of Print	0.70	0.68
Math Level 1	0.41	0.27
Math Level 2	0.58	0.49
Math Level 3	0.63	0.66
Math Level 4	0.54	0.63
Math Level 5	0.46	0.53

3.2 Indirect Cognitive Assessment (ARS)

The ARS was developed for the ECLS-K to measure teachers' evaluations of students' academic achievement in the three domains that are also directly assessed in the cognitive battery: language and literacy (reading), general knowledge (science and social studies), and mathematical thinking. The difference between the direct and indirect cognitive assessments, and the scores available, are described below. For a discussion of the content areas and response scales of the ARS, see chapter 2, section 2.3.1.

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.

Another major difference between the ARS and direct cognitive assessment is that the skills, knowledge, and behaviors on the ARS reflect a broader sampling of the most recent national curriculum standards and guidelines from early childhood professionals and researchers. The ARS items were not limited by the constraints of a standardized testing format as were the direct cognitive items. Therefore, the scope of curricular content represented in the indirect measures is broader than the content represented on the direct assessment battery.

3.3 IRT Scores Available for the ARS

IRT was employed to calculate scores for the ARS in order to compare performance of students from fall to spring and to be able to compare students who were not rated on all items. The Rasch

Rating Scale Model uses the pattern of ratings on the items actually administered to determine an estimate of the difficulty of each item and to place each student on a continuous ability scale.

A Rasch analysis was performed on the spring ARS data. In the fall-kindergarten, a large percentage of the teachers had not introduced at least some of the items to the classroom setting, resulting in a large number of missing ratings for all but four items.

By the spring-kindergarten, teachers had introduced the skills, knowledge, and behaviors represented in the items to their classrooms. The NA category was used in less than three percent of the cases on all items except "Composes simple stories" (7.52 percent NA); "Uses the computer" (15.83 percent NA); "Recognizes ways people rely on each other" (3.26 percent NA); and "Uses instruments for measuring" (9.82 percent NA). On the majority of the items, the use of the NA category was less than one percent.

The item difficulties from the spring analysis were applied to the fall items and ability estimates were computed for each of the children based on the difficulty estimates of the items and the pattern of ratings children received on those items. The Rasch analysis of the spring data showed that the reliability of the estimates of child ability was very high for all domains in both spring and fall (see table 3-4).

Table 3-4. Reliability for the Rasch-based score

	Fall-Kindergarten	Spring-Kindergarten
ARS Language and Literacy	0.87	0.91
ARS Mathematical Thinking	0.92	0.93
ARS General Knowledge	0.92	0.94

The ARS Scores were rescaled to have a low of one and a high of five to correspond to the five-point rating scale that teachers used in rating children on these items. The ARS scores in the database represent estimates of the rating students would have received in that domain if they had been rated on all of the items in the ARS. Gain scores may be obtained by subtracting the estimated rating at fall (time 1) from the estimated rating at spring (time 2).

The variable names, descriptions, value ranges, means, and standard deviations for the fall (T1) and spring (T2) kindergarten ARS scores are shown in table 3-5.

Table 3-5. ARS: variable names, descriptions, ranges, means, and standard deviations[1]

Variable Name	Description	Range of Values	Mean	Standard Deviation
T1ARSLIT	T1 Literacy ARS Score	1 - 5	2.6	0.8
T1ARSMAT	T1 Math ARS Score	1 - 5	2.5	0.8
T1ARSGEN	T1 General Knowledge ARS Score	1 - 5	2.6	1.0
T2ARSLIT	T2 Literacy ARS Score	1 - 5	3.3	0.8
T2ARSMAT	T2 Math ARS Score	1 - 5	3.5	0.9
T2ARSGEN	T2 General Knowledge ARS Score	1 - 5	3.6	1.0

[1] See chapter 7, section 7.3 for variable naming conventions.

The majority of teachers rated more than one student on the ARS. The number of students rated by each teacher ranged from 1 to more than 20.

3.4 The Oral Language Development Scale (OLDS)

The language assessment scores (OLDS scores) for language minority children are located with the other child scores on the file. There are a total of 16 scores – 4 English and 4 Spanish for each of the two rounds, i.e., fall-kindergarten and spring-kindergarten. Children in households speaking languages other than English were first given the English OLDS. Of that group, those scoring below the cut point of the English OLDS were administered the Spanish OLDS if the child's home language was noted as Spanish by the school. (See chapter 2, section 2.1 for more detail on the content of the OLDS items).

The variable names, descriptions, value ranges, means, and standard deviations for the OLDS are shown in table 3-6.

Table 3-6. OLDS: variable names, descriptions, ranges, means, and standard deviations[1]

Variable	Description	Range of Values	Mean	Standard Deviation
C1SCTOT	C1 AIQ400 Child's Total OLDS Score	1-60	34.0	16.0
C1SCORD	C1 AIQ400 Simon Says Child Score	1-10	8.0	2.3
C1SCART	C1 AIQ400 Art Show Child Score	1-10	7.3	2.7
C1SCSTO	C1 AIQ400 Tell Stories Child Score	4-40	22.4	9.0
C1SSCTOT	C1 SAIQ400 Spanish Total OLDS Scores	2-44	22.2	6.5
C1SSCORD	C1 SAIQ400 Spanish Simon Says Child Score	1-10	9.7	1.1
C1SSCART	C1 SAIQ400 Spanish Art Show Child Score	1-10	8.0	1.8
C1SSCSTO	C1 SAIQ400 Spanish Tell Stories Score	4-24	10.6	5.3
C2SCTOT	C2 AIQ400 Child's Total OLDS Score	1-60	30.9	15.2
C2SCORD	C2 AIQ400 Simon Says Child Score	1-10	7.6	2.4
C2SCART	C2 AIQ400 Art Show Child Score	1-10	6.8	2.6
C2SCSTO	C2 AIQ400 Tell Stories Child Score	4-40	20.7	8.7
C2SSCTOT	C2 SAIQ400 Spanish Total OLDS Scores	2-44	23.7	6.4
C2SSCORD	C2 SAIQ400 Spanish Simon Says Child Score	1-10	9.8	0.6
C2SSCART	C2 SAIQ400 Spanish Art Show Child Score	1-10	8.5	1.6
C2SSCSTO	C2 SAIQ400 Spanish Tell Stories Score	4-24	10.0	5.0

[1] See chapter 7, section 7.3 for variable naming conventions.

3.5 Psychomotor Assessment

The psychomotor assessment includes two scales, one measuring visual motor skills (eye-hand coordination) and the other measuring gross motor skills (balance and motor planning). The visual motor skills score is the sum of the points for seven tasks: build a gate, draw a person, and copy five simple figures. Children could receive up to two points for each of the first two tasks and one point for each of the figures. Gross motor skills consisted of balancing, hopping, skipping and walking backward—children could receive up to two points for each skill. Confirmatory factor analysis during the ECLS-K design phase (using LISREL) confirmed the two scales. The internal consistency of the scales was constrained by the limited number of items in each scale combined with the variety of motor skills measured and the limited variance in item scores (maximum score on items was 1-2). Alpha coefficients (reliabilities) were 0.57 for fine motor skills, 0.51 for gross motor skills, and 0.61 for the composite motor

skills. Variable names, descriptions, value ranges, means, and standard deviations for the three scales are shown in table 3-7 below.

Table 3-7. Psychomotor scales: variable names, descriptions, ranges, means, and standard deviations[1]

Variable	Description	Range of Values	Mean	Standard Deviation
C1FMOTOR	C1 Fine Motor Skills	0 - 9	5.7	2.1
C1GMOTOR	C1 Gross Motor Skills	0 - 8	6.3	1.9
C1CMOTOR	C1 Composite Motor Skills	0 - 17	12.1	3.1

[1] See chapter 7, section 7.3 for variable naming conventions.

3.6 SRS

The SRS asked both teachers and parents to tell how often a student exhibited certain social skills and behaviors. Teachers and parents used a frequency scale (see table 3-8) to report on how often the student demonstrated the behavior described. See chapter 2, section 2.2 and 2.3.2 for additional information on the parent and teacher SRS instruments. 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 same items were administered in the fall and in the spring so change scores may be computed by subtracting time 1 (fall) from time 2 (spring). The reliability for the teacher SRS scales is high (see table 3-9). The reliability is lower for the parent scales (see table 3-10).

Table 3-8. SRS response scale

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.

Table 3-9. Split half reliability for the teacher SRS scale scores

	Fall-Kindergarten	Spring-Kindergarten
Approaches to Learning	0.89	0.89
Self-Control	0.79	0.80
Interpersonal	0.89	0.89
Externalizing Problem Behaviors	0.90	0.90
Internalizing Problem Behaviors	0.80	0.78

Table 3-10. Split half reliability for the parent SRS scale scores

	Fall-Kindergarten	Spring-Kindergarten
Approaches to Learning	0.68	0.69
Self-Control	0.74	0.75
Social Interaction	0.70	0.68
Impulsive/Overactive	0.46	0.47
Sad/Lonely	0.60	0.61

3.6.1 Teacher SRS

Teachers rated individual students as part of a self-administered questionnaire. The five social skill teacher scales are: approaches to learning, self-control, interpersonal skills, externalizing problem behaviors, and internalizing problem behaviors. (See chapter 2, section 2.3.2 for a description of the teacher scales.)

Variable names for the teacher scores, descriptions, ranges, means, and standard deviations for these scales are shown in table 3-11.

Table 3-11. Teacher social rating scores: variable names, descriptions, ranges, means, and standard deviations[1]

Variable	Description	Range of Values	Mean	Standard Deviation
T1LEARN	T1 Approaches to Learning	1-4	3.0	0.7
T1CONTRO	T1 Self-Control	1-4	3.1	0.6
T1INTERP	T1 Interpersonal	1-4	3.0	0.6
T1EXTERN	T1 Externalizing Problem Behaviors	1-4	1.6	0.6
T1INTERN	T1 Internalizing Problem Behaviors	1-4	1.6	0.5
T2LEARN	T2 Approaches to Learning	1-4	3.1	0.7
T2CONTRO	T2 Self-Control	1-4	3.2	0.6
T2INTERP	T2 Interpersonal	1-4	3.1	0.6
T2EXTERN	T2 Externalizing Problem Behaviors	1-4	1.7	0.7
T2INTERN	T2 Internalizing Problem Behaviors	1-4	1.6	0.5

[1] See chapter 7, section 7.3 for variable naming conventions.

Care should be taken when entering these scales into the same analysis due to problems of multicollinearity. The factor intercorrelations among the scales for social skills are high. The factor intercorrelations with the internalizing problem behaviors are the lowest. The correlations between the teacher SRS factors range from 0.25 to 0.78 in fall-kindergarten and from 0.30 to 0.80 in spring-kindergarten (absolute values).

3.6.2 Parent SRS

The items on the parent SRS were administered as part of a longer telephone or in-person survey. (See chapter 2, section 2.2 for a description of the parent scales.) The factors on the parent SRS are similar to the teacher SRS; however, the items in the parent SRS are geared to the home environment and thus are not the same items. It is also important to keep in mind that parents and teachers observe the children in very different environments. The five social skill parent scales are: approaches to learning, self-control, social interaction, impulsive/overactive, and sad/lonely. The correlations between the parent SRS factors were not as high as the teacher SRS factors. They ranged from 0.05 to 0.45 in fall-kindergarten, and from 0.08 to 0.45 in spring-kindergarten (absolute values).

Variable names for the parent scores, descriptions, ranges, means, and standard deviations for these scales are shown in table 3-12.

Table 3-12. Parent social rating scores: variable names, descriptions, ranges, means, and standard deviations[1]

Variable	Description	Range of Values	Mean	Standard Deviation
P1LEARN	P1 Approaches to Learning	1-4	3.1	0.5
P1CONTRO	P1 Self-Control	1-4	2.8	0.5
P1SOCIAL	P1 Social Interaction	1-4	3.3	0.6
P1SADLON	P1 Sad/Lonely	1-4	1.5	0.4
P1IMPULS	P1 Impulsive/Overactive	1-4	2.0	0.7
P2LEARN	P2 Approaches to Learning	1-4	3.1	0.5
P2CONTRO	P2 Self-Control	1-4	2.9	0.5
P2SOCIAL	P2 Social Interaction	1-4	3.4	0.5
P2SADLON	P2 Sad/Lonely	1-4	1.6	0.4
P2IMPULS	P2 Impulsive/Overactive	1-4	2.0	0.7

[1] See chapter 7, section 7.3 for variable naming conventions.

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. 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.

4.1 Selecting the Area Sample

The point of departure for the ECLS-K area sample frame development was an existing multipurpose frame of PSUs created using 1990 county-level population data. This frame contained 1,404 PSUs that were counties or groups of contiguous counties. PSUs did not cut across census regional boundaries but were allowed to cross state boundaries. Each 1990 Metropolitan Statistical Area (MSA) constituted a single PSU except where an MSA crossed census regions, and it was split into two PSUs. The minimum size of a PSU in the multipurpose frame was 15,000 persons.

Since the focus of the ECLS-K is kindergarten students, the existing PSU frame was updated with 1994 population estimates of five-year-olds by race-ethnicity, the most up-to-date estimates available from the U.S. Census Bureau at the time. The counts of five-year-olds by race-ethnicity were used to revise PSU definitions relative to a different minimum PSU size and to construct a measure of size that facilitated the oversampling of Asians and Pacific Islanders (APIs).

Each PSU in the frame that did not have at least 320 five-year-olds was collapsed with an adjacent PSU. This minimum PSU size was developed based on assumptions concerning anticipated school response rates, the average number of schools that would be selected per PSU, and the target number of students to be sampled per school. After this collapsing, the final ECLS-K PSU frame contained 1,335 records.

The measure of size used for selecting PSUs took into account the amount of oversampling of APIs required to meet the ECLS-K precision goals. The weighted measure of size was calculated as follows:

$$MOS = 2.5 \times n_{API} + n_{other}$$

where 2.5 is the oversampling rate for APIs and n_{API} and n_{other} are the counts of five-year-old APIs and all others, respectively. The oversampling rate for APIs was calculated as the target number of completed API cases divided by the expected number of completed API cases without oversampling.

In all, 100 PSUs were selected for the ECLS-K. The 24 PSUs with the largest measures of size were designated as certainty selections or self-representing (SR) and were set aside. Once the SR PSUs were removed, the remaining PSUs were partitioned into 38 strata of roughly equal measure of size. The frame of non-SR PSUs was first sorted into eight superstrata by MSA/nonMSA status and by Census region. Within the four MSA superstrata, the variables used for further stratification were race-ethnicity (high concentration of API, Black, or Hispanic), size class (MOS >= 13,000 and MOS < 13,000) and 1988 per capita income. Within the four non-MSA superstrata, the stratification variables were race-ethnicity and per capita income. Details of the stratification of non-self-representing strata are presented in table 4-1.

Two PSUs were selected from each non-SR stratum using Durbin's Method.[1] This method selects two first-stage units per stratum without replacement, with probability proportional to size and a known joint probability of inclusion. The Durbin method was used because it allows variances to be estimated as if the units were selected with replacement. Table 4-2 summarizes the characteristics of the ECLS-K PSU sample.

4.2 Selecting the School Sample

In the second stage of sampling, public and private schools offering kindergarten programs were selected. For each ECLS-K PSU, a frame of public and private schools offering kindergarten

[1] Durbin, J. (1967). Design of Multi-Stage Surveys for the Estimation of Sampling Errors. *Journal of the Royal Statistical Society* C, 16, 152-164.

Table 4-1. Stratum definitions for the 38 non-self-representing strata

Metro status	Census region	Race-ethnicity (%)	PSU size (MOS)	Per capita income Range ($)	
MSA	Northeast	-	≥13,000	22,062	25,424
				16,342	22,030
			<13,000	18,128	29,084
				16,697	18,032
				12,279	16,616
	Midwest	-	≥13,000	17,277	18,150
				16,103	17,092
			<13,000	16,552	24,009
				15,732	16,475
				14,450	15,693
				10,185	14,433
	South	Hispanic ≥30	-	-	-
		Black ≥40	-	-	-
		26≤ Black <40	-	14,743	18,731
			-	10,892	14,573
		Black <26		16,435	16,601
			≥13,000	14,586	16,337
			<13,000	15,572	22,824
				14,194	15,432
				11,262	13,979
	West	API ≥15	-	-	-
		API <15	-	-	-
		Hispanic ≥30	-	-	-
		12≤ Hispanic <30	-	-	-
		Hispanic <12	-	15,048	21,840
			-	9,993	14,839
NonMSA	Northeast	-	-	-	-
	Midwest	-	-	14,124	17,446
				13,277	14,121
				12,169	13,272
				6,992	12,147
	South	Black ≥42	-	-	-
		25≤ Black <42	-	-	-
		Black <25		12,727	20,059
			-	11,165	12,676
				6,018	11,142
	West	-	-	12,887	23,286
				6,959	12,884

Table 4-2. Distribution of the ECLS-K PSU sample by SR status, MSA status, and census region

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

programs was constructed using existing school universe files: the 1995-96 Common Core of Data[2] (CCD) and the 1995-96 Private School Universe Survey[3] (PSS). The school frame was freshened in the spring of 1998 to include newly opened schools that were not included in the CCD and PSS and schools that were in the CCD and PSS but did not offer kindergarten according to those sources. A school sample supplement was selected from the freshened frame.

4.2.1 Frame Construction

The 1995-96 CCD Public School Universe File was the primary source for the ECLS-K public school sampling frame. Most schools run by the Bureau of Indian Affairs (BIA) and the schools run by the Department of Defense (DOD) are not included on the CCD. The 1995-96 Office of Indian Education Programs Education Directory was consulted in order to complete the list of BIA schools in the CCD file. For the DOD schools, a 1996 list of schools obtained directly from the DOD was used. The 1995-96 PSS Universe File was used as the primary source of the private school sampling frame.

The first step in frame construction involved subsetting the file to schools located in counties that constituted the ECLS-K PSU sample. Further subsetting retained only those schools that offered transitional kindergarten, kindergarten, or transitional first grade, or which were strictly ungraded, as indicated by the school's gradespan.

[2] U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public School Universe Survey, 1995-96.

[3] U.S. Department of Education, National Center for Education Statistics. *Private School Universe Survey, 1995-96*, NCES 98-229, by Stephen P. Broughman and Lenore A. Colaciello. Washington, DC: 1998.

The constructed ECLS-K school frame included 18,911 public-school records and 12,412 private-school records. The school frame was freshened in the spring of 1998 to include schools that would be operational in fall 1998 but that were not included in the frame just described. The freshening procedures are given later in this section.

4.2.2 School Measure of Size, Stratification, and Sample Selection

Within each PSU, schools with fewer than a predetermined minimum number of kindergarten students were clustered together before sampling in order to obtain a sample that is closer to self-weighting. The minimum was 24 kindergartners for public schools and 12 for private schools. For simplicity's sake, the term "schools" will be used in reference to both individual schools and clusters of schools in the remainder of this discussion.

Schools were selected with probability proportional to size. As with the PSU sample, a weighted measure of size was constructed taking into account the oversampling of APIs.

The target number of sampled schools per PSU was calculated separately for public and for private schools and was adjusted upward to offset anticipated school response and eligibility rates. The number of schools allocated to each PSU was set proportional to the weighted measure of size of the PSU. A minimum of one school per PSU was imposed for any PSU so small that it would not otherwise have been allocated a school.

Public and private schools constituted distinct sampling strata. Within each of these strata, schools were sorted to ensure good sample representation across other characteristics.

The public school sample was selected using a traditional nested two-stage design of schools within the 100 PSUs. Within each PSU, public schools were ranked by measure of size and partitioned into three classes of roughly equal aggregate measure of size. Within each size class, schools were sorted in a serpentine manner by the proportion of APIs.

The private school sample was nested within PSUs only for the 76 non-SR PSUs, where schools were grouped within PSU by affiliation (religious vs. nonsectarian) and then sorted in a serpentine manner by the measure of size. To better control the sample distribution of religious/

nonsectarian schools, schools in the 24 SR PSUs were treated as if they were sampled from a single PSU and the sort was by affiliation and measure of size alone.

The selection of both public and private schools was systematic, with probability proportional to the measure of size. A total of 1,280 schools were selected for the ECLS-K, of which 934 were public and 346 were private schools.

4.2.3 Freshening the School Sample

Each public school district having one or more schools sampled was sent a sampling frame-based list of all schools offering kindergarten and was asked whether any school expected to offer kindergarten in academic year 1998-1999 was missing from the list. For each such school identified, school name, address, telephone number, grade span, and kindergarten enrollment were obtained. Also contacted were districts that fell within the boundaries of the ECLS-K PSUs but for which the CCD file listed no schools offering kindergarten, unless it was clear from their name that they were strictly secondary school districts (e.g., Middlebury Union High School District). The information obtained from the school districts was checked against the ECLS-K public school frame to confirm that these schools were truly new or newly eligible. Bona fide new schools were given an appropriate chance of being sampled. A new school's chance of selection was conditioned on the school district's probability of selection, which had been calculated exactly, based on the sampling intervals used during the systematic selection of the main school sample, and the positions in the frame and measures of size of all schools in the frame. Overall 252 new public schools were identified, and 19 were selected using systematic sampling with probability proportional to size.

The procedure for obtaining new school information from Catholic dioceses was exactly the same as for public schools. Since a diocese may cut across county or even state lines, each school identified by a diocese had to be associated with the correct county, and hence the correct PSU, before checking to see whether it was truly new. Since dioceses may cross PSU boundaries, a new Catholic school's chance of being sampled had to be conditioned on the diocese probability of selection within the PSU where the new school was located. There were 126 new Catholic schools identified, and 6 were selected using systematic sampling with probability proportional to size.

The search for non-Catholic private schools was considerably more complicated. Three classes of PSS schools that had previously not been given a chance of selection were reconsidered. Those were schools that had an unknown grade span because they had not responded to the 1995-96 PSS, those that responded but did not report offering kindergarten, and those that appeared for the first time on the 1997-98 PSS file. All told these accounted for 2,544 potential new non-Catholic private schools. Beyond these additions from PSS, procedures similar to those used by the Census Bureau in the PSS area frame search[4] were followed. These procedures included collecting lists of schools from different sources, matching them against the PSS list frame to remove duplicates, and further screening by telephone to verify new school status. The majority of new schools found by the Census Bureau for PSS came from Yellow Pages listings. The Yellow Pages search was the main source of new non-Catholic private schools in the ECLS-K as well, yielding an additional 8,861 possible new private schools. Since the number of kindergartners enrolled in these schools was unknown, a minimum kindergarten enrollment was assumed for sampling purposes (typically 24, unless the name was suggestive of daycare in which case 12 was assumed). From the 11,405 schools, a sample of 279 schools was selected using systematic sampling with a probability proportional to these imputed enrollments. The sampled schools were screened, during which process it was ascertained whether each school was public or private; if it was private, whether it would be open in academic year 1998-1999; and if it would be open in 1999-2000, whether it would offer kindergarten. If the answer to the last question was yes and the school was not Catholic, the school was released for data collection.

Local Education Agencies (LEAs) and local government offices were contacted for information on non-Catholic private schools, only in the smallest ECLS-K PSUs, on the theory that if these PSUs had coverage problems their large weights were likely to introduce a larger bias in the estimates. All LEAs within these PSUs were contacted by telephone. For each city/town within the PSU, a list of local government offices was compiled using the Blue Pages. Successive government offices were called within a city or town until one was found that could provide information on private schools. As with the Yellow Pages, new schools identified by LEAs and local government offices were unduplicated against the PSS file before being added to the new school frame. Since kindergarten enrollment was unknown, it was imputed as described in the previous paragraph and sampling was performed using systematic sampling with probability proportional to size. The LEA search resulted in the identification of 30 new private schools after unduplication, of which 14 were sampled. The local government search yielded 19 new schools, of which 8 were sampled. Finally, three additional new

[4] ibid.

private schools were reported by field staff based on personal knowledge. Of these, two schools were sampled. All told, there were 109 new non-Catholic private schools selected.

The characteristics of the ECLS-K original school sample are presented in table 4-3. Schools that were discovered to be ineligible during recruitment have been omitted from the tabulation. For counts of responding schools and a discussion of the limited use of school substitution within the ECLS-K, see chapter 5, sections 5.5, Fall-Kindergarten Response Rates, and 5.8, Spring-Kindergarten Response Rates.

4.2.4 Sampling Children, Parents, and Teachers within Schools

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 each targeted subpopulation. As mentioned earlier, APIs were the only subgroup that needed to be oversampled to meet the sample size goals. For each sampled school, the field staff obtained a complete list of kindergartners enrolled, taking special care that no child was excluded from the list because of disability or language problems.

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 with twins being sampled as a unit rather than as individuals, i.e. if one twin was sampled, both were included. In general, the target number of children sampled at any one school (not including the second twin) was 24. In some schools the oversampling goal for API students could not be met. For example, in a school with 24 kindergartners enrolled, all students would be sampled, which meant that API students could not be sampled at 2.5 times the rate of non-API students in that school. To offset shortfalls of this kind, the oversampling rate for APIs was increased to three at this stage of sampling.

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. Characteristics of the ECLS-K original school sample

	Public	Private	Total
Total	914	363	1,277
Region			
Northeast	161	82	243
Midwest	210	88	298
South	306	112	418
West	237	81	318
Urbanicity			
Large central city	168	77	245
Mid-size central city	172	76	248
Urban fringe of large city	265	117	382
Urban fringe of mid-size city	78	21	99
Large town	24	9	33
Small town	76	36	112
Rural	131	27	158
KG enrollment			
1 – 24	51	187	238
25 – 49	95	110	205
50 – 99	402	59	461
100 – 149	226	7	233
150 – 169	49	0	49
170 +	91	0	91
Religious affiliation			
Catholic	-	120	120
Other religious	-	149	149
Nonreligious, private	-	94	94
Free Lunch Program			
Low (<=25% eligible students)	268	-	268
Medium low (>25% and <=50%)	157	-	157
Medium high (>50% and <=75%)	129	-	129
High (>75%)	114	-	114
Missing	246	-	246

During the fall 1998 data collection, a census of kindergarten teachers was taken at each school. Each sampled child was linked to his or her kindergarten teacher. A child could be linked to only one teacher. If a child was taught by more than one teacher, a "primary" teacher was identified for the child. For the specifics of how this was done, see section 5.4.2, Fall-Kindergarten Data Collection, Distribution of Teacher Questionnaires. In spring 1999, 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 of kindergarten data collection. If a sampled child received special education services from such a teacher, the teacher was linked to that child. As with regular teachers, a child could be linked to only one special education teacher.

4.3 Calculation and Use of Sample Weights

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. In general, there are three types of weights: child, teacher, and school-level weights. The use of these weights is essential to produce estimates that are representative of the population of kindergarten children, kindergarten teachers, and schools offering kindergarten programs. Data collected from different sources can be used to produce estimates at these three levels. For example, data collected from parents are used to produce estimates of characteristics of children as reported by parents.

Several sets of weights were computed for each of the two rounds of data collection (fall- and spring-kindergarten). Longitudinal weights were also computed for children with complete data from both rounds of the study. Unlike surveys that have only one type of survey instrument aimed at one type of sampling unit, the ECLS-K is a complex study with multiple types of sampling units, each having its own survey instrument. Each type of unit was selected into the sample through a different mechanism: children were sampled directly through a sample of schools; parents of the sampled children were automatically included in the survey; all kindergarten teachers in the sampled schools were included; special education teachers were in the sample if they taught any of the sampled children. Each sampled unit had its own survey instrument: children were assessed directly using a series of cognitive and physical assessments; parents were interviewed with a parent instrument; teachers filled out at least two different types of questionnaires depending on the round of data collection and on whether they were regular or special education teachers; school principals reported their school characteristics using the school administrator questionnaire. The stages of sampling in conjunction with the different nonresponse level at each stage and the diversity of survey instruments require that multiple sampling weights be computed for use in analyzing the ECLS-K data.

This section describes the different types of sample weights computed for the ECLS-K, how they were calculated, how they should be used, and their statistical characteristics.

4.3.1 Types of Sample Weights

Weighting was carried out in stages to produce child, teacher, and school weights. Several sets of child-level weights were computed for each round of data collection and for children with complete data from both rounds. While it is straightforward to use school- and teacher-level weights to produce school- and teacher-level estimates, careful consideration should be given to the choice of a child-level weight since it depends on the type of data analyzed. Each set of child-level weights is appropriate for a different set of data or combination of sets of data. Teacher-level weights were computed for each round of data collection, but there are no longitudinal teacher-level weights. School level weights were computed for use with data collected in spring-kindergarten through the school administrator questionnaires. These weights can also be used with any school-level data such as data from the school facilities checklists.

Tables 4-4 and 4-5 summarize the different types of cross-sectional weights and how they should be used. Cross-sectional weights provide an accurate estimate for the specific round of data collection. Table 4-4 describes weights for fall-kindergarten estimates, and table 4-5 describes weights for spring-kindergarten estimates, and table 4-6 describes weights for base-year or longitudinal estimates.

These tables are designed to help users choose appropriate weights for their analysis. Answers to the following three questions can help in the selection of the correct weight.

1. Is the analysis concerned with one point in time or two?
2. What is the population of interest or unit of analysis (i.e. child, teacher or school)?
3. What instruments do the data to be used in the analysis come from?

1. Is the analysis concerned with one point in time or two? If the analysis pertains only to fall kindergarten (single point in time) then table 4-4 guides the selection of weights, spring kindergarten (single point in time) then go to table 4-5, and both fall- and spring-kindergarten (two points in time) then go to table 4-6.

Table 4-4. The ECLS-K: fall-kindergarten cross-sectional weights

Fall-kindergarten cross-sectional weights	
Child-level weight	to be used for analysis of ...
C1CW0	fall-kindergarten 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 fall-kindergarten teacher questionnaire A, B or C data, and c) data from the school administrator questionnaire or facilities checklist
C1PW0	fall-kindergarten parent interview data alone or in combination with a) fall child assessment data, b) fall-kindergarten teacher questionnaire A, B, or C data, and c) data from the school administrator questionnaire or facilities checklist. *Exception*: If data from the parent AND child assessment AND teacher questionnaire A or B (not C) are used then C1CPTW0 should be used.
C1CPTW0	fall-kindergarten direct child assessment data combined with fall-kindergarten parent interview data AND fall-kindergarten teacher questionnaire A or B (not C) data alone or conjunction with data from the school administrator questionnaire or facilities checklist
Teacher-level weight	to be used for analysis of …
B1TW0	fall-kindergarten teacher data, questionnaire part A or B alone or with data from the school administrator questionnaire or facilities checklist. *Exception*: When using items that were in the spring-kindergarten teacher questionnaire B (i.e. questions asked of teachers who were not present during fall-kindergarten data collection) B2TW0 weight should be used.

Table 4-5. The ECLS-K: spring-kindergarten cross-sectional weights

Spring-kindergarten cross-sectional weights	
Child-level weight	to be used for analysis of ...
C2CW0	spring-kindergarten 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 spring-kindergarten teacher questionnaire A, B or C data, and c) data from the school administrator questionnaire or facilities checklist
C2PW0	spring-kindergarten parent interview data alone or in combination with a) spring child assessment data, b) spring-kindergarten teacher questionnaire A, B, or C data, and c) data from the school administrator questionnaire or facilities checklist. *Exception*: If data from the parent AND child assessment AND teacher questionnaire A or B (not C) are used then C2CPTW0 should be used
C2CPTW0	spring-kindergarten direct child assessment data combined with spring-kindergarten parent interview data AND spring-kindergarten teacher data alone or in conjunction with data from the school administrator or facilities checklist
Teacher-level weight	to be used for analysis of ...
B2TW0	spring-kindergarten data from questionnaire part A; fall- or spring-kindergarten data from questionnaire part B; or combination of data from fall- and spring-kindergarten teacher questionnaires A and/or B (there is no longitudinal teacher weight) alone or in conjunction with data from the school administrator questionnaire or facilities checklist
School administrator weight	to be used for analysis of ...
S2SAQW0	school administrator data or facility checklist data

2. What is the population of interest or unit of analysis (i.e. child, teacher or school)? After identifying the appropriate table based on the response to question 1, the next step involves whether the analysis requires a child-, teacher- or school-level weight. If the population of inference is kindergarten children, then the child-level weights will be appropriate. If generalizing to kindergarten teachers or classrooms, then the teacher-level weights should be used, and if generalizing to the population of schools with kindergartens, the school-level weight will be appropriate.

3. What instruments do the data to be used in the analysis come from? There are several options when deciding on which child-level weights to use, and the source of the data affect which weight to use. In each of the tables details under "to be used in the analysis of . . ." provide guidance based on whether the data were collected through the child assessments, parent interviews or teacher questionnaires.

Weight C1CW0 is used to estimate child-level characteristics or assessment scores for fall-kindergarten, and C2CW0 is for spring-kindergarten. Examples of such estimates are the percent of kindergarten children who are male, the percent of children who are API, the percent of children who are 6 when they enter kindergarten, the mean reading score of children, and the mean math score of children. These weights are also used for estimates of child characteristics for language minority (LM)/not Spanish children and children with disabilities. Some of these children were not assessed but their background characteristics such as age, gender, race-ethnicity, and characteristics of parents, teachers, and classrooms are available from the parent interview and the teacher questionnaires. The social rating scores (see chapter 3, section 3.5) from parents and teachers are also available for LM/not Spanish children and children with disabilities, regardless of whether they were assessed. In this chapter the terminology "LM/ not Spanish" refers to those children who spoke a non-English and non-Spanish language at home and did not pass the cut score in the OLDS. For these children, only height and weight measurements were taken. Similarly, in this chapter, the terminology "children with disabilities" refer to those children who were not able to participate in the assessment due to reasons of disability as specified in their IEP. Both these groups of children were assigned child weights even though they did not have cognitive assessment data.

When analyzing child data in conjunction with teacher data (e.g., teacher characteristics from teacher questionnaires A or B or social rating scores reported by teachers from teacher questionnaire C) and classroom data from teacher questionnaire A, weights C1CW0 (for fall-kindergarten) and C2CW0 (for spring-kindergarten) should be used. An example for the use of C1CW0 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 this type of analysis. However, different weights should be used for analysis of child data in conjunction with both parent and teacher data (CPTW0).

C1PW0 (for fall-kindergarten) and C2PW0 (for spring-kindergarten) are 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 at least 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.

C1CPTW0 (for fall-kindergarten) and C2CPTW0 (for spring-kindergarten) are 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. Also, any analysis of the subgroup of children who moved to a different school between fall- and spring-kindergarten should be done using C2CW0 or C2PW0 because these movers were treated separately in the nonresponse adjustment of C2CW0 and C2PW0.

B1TW0 (for fall-kindergarten) and B2TW0 (for spring-kindergarten) are used for teacher- and classroom-level estimates. For example, these weights would be used to estimate the percent of kindergarten teachers who teach a particular type of kindergarten program, the percent of teachers who use a language other than English in their classroom, or the percent of teachers who do not have any teacher certification. These weights would also be used in the estimation of classroom characteristics such as the percent of kindergarten classrooms with computer areas. Weights for the corresponding round should be used to produce round specific estimates; for example, B1TW0 should be used to estimate the number of teachers teaching kindergarten programs in fall 1998; B2TW0 should be used to estimate the number of kindergarten teachers in spring 1999. Teachers who were new to the study in the spring were asked a subset of questions from the fall-kindergarten teacher questionnaires A and B (spring-kindergarten teacher questionnaire B). When analyzing data from these items B2TW0 should be used (even though the variables start with the B1 preface, see chapter 7 for details). A panel weight was not created as there were very few repeated measurements between fall- and spring-kindergarten. Most of the

data collected from the teachers about themselves and their classrooms were meant to pertain to the school year.

S2SAQW0 is used in the estimation of school characteristics such as the percent of schools that offer programs for children with special needs, or the percent of schools that use a standardized achievement tests as a requirement for admission, or the percent of school administrators who believe that their school computer resources are inadequate.

The longitudinal or panel weights (table 4-6) are used for estimates of differences at two points in time. Examples of analysis using longitudinal weights include:

- Kindergarten fall-spring difference in mean child assessment scores (BYCW0);
- Kindergarten fall-spring difference in mean social skills as rated by children's teachers (BYCW0) (Data collected using the teacher questionnaire C are at a child-level and are considered as part of the child assessment data.);
- Kindergarten fall-spring difference in mean social skills as rated by children's parents (BYPW0);
- The relationship between the gains children make in their reading knowledge and skills, how often their parents read to them, how often their parents take them to the library, teacher certification, and how much class time teachers spend on reading (BYCPTW0) (This weight is used when the analysis includes data from all six components- fall and spring child assessment, teacher and parent data.); and
- The relationship between the gains children make in their reading knowledge and skills and parent and teacher beliefs on kindergarten readiness (BYCOMW0).

The difference between BYCPTW0 and BYCOMW0 is that BYCPTW0 is used for analysis of both rounds of child direct assessment data in conjunction with both rounds of parent interview data and both rounds of teacher data, while BYCOMW0 is used when analyzing a *single round* of parent interview or teacher data in conjunction with *both rounds* of child direct assessment data.

Table 4-6. The ECLS-K: base year longitudinal weights

Base year longitudinal (panel) weights	
Child-level weight	to be used for analysis of ...
BYCW0	child direct assessment data and child characteristics from both fall- and spring-kindergarten, alone or in conjunction with any combination of a) a limited set of child characteristics (e.g. age, sex, race-ethnicity), b) fall- and/or spring- kindergarten teacher questionnaires A, B or C data, and c) data from the school administrator questionnaire or facilities checklist
BYPW0	parent interview data from BOTH fall- and spring-kindergarten alone or in combination with a) fall- and/or spring-kindergarten child assessment data, b) fall- and/or spring-kindergarten teacher questionnaire A, B, or C data, and c) data from the school administrator questionnaire or facilities checklist. *Exception*: If data from the fall- AND spring-kindergarten parent, child assessment, AND teacher questionnaire A or B (not C) are used then BYCPTW0 should be used
BYCPTW0	fall- AND spring-kindergarten parent, child assessment, AND teacher questionnaire A or B (not C) data
BYCOMW0	both rounds of child assessment data in conjunction with at least one or more rounds (fall- and/or spring-kindergarten) of parent and/or teacher questionnaires A and B (not C) data. This may or may not be in conjunction with the school administrator questionnaire and facilities checklist data. *Exception*: Whenever BOTH rounds of parent data are used in the analysis either BYPW0 or BYCPTW0 is used.

Minimal teacher data were collected repeatedly in both rounds that would allow analysis of change over time for those data. In the absence of the teacher panel weights, the spring-kindergarten teacher weights can be used to cross-tabulate the fall- and spring-kindergarten teacher data from questionnaire part A (TQA).

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 C1CPTW0 were used in an analysis of child and teacher/classroom data only, then the resulting estimates will be inefficient compared to estimates using C1CW0. The lower parent response causes C1CPTW0 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. For example, no parent-level weights were calculated but parents are linked one-to-one to children so that

the child-level weights can be used for parent-level analysis, e.g., education levels of parents of kindergartners. No child-teacher weights were computed for analyzing child data in conjunction with teacher data because the response rates for the teachers are high; for the analysis of child assessment data in conjunction with teacher data, the child-level weights should be used. For further advice on which weights to use when analyzing a complex combination of data, contact NCES at ECLS@ed.gov.

For each type of weight, table 4-7 gives the distribution of schools by number of sampled students with nonzero weights and the mean number of sampled students with nonzero weights per school. This is useful in analysis using hierarchical linear modeling. For spring-kindergarten, the increase in the count of schools with one to five sampled students is due to students transferring to other schools from the original sampled schools. For the longitudinal weights, schools are classified on the basis of the number of students who did not transfer schools between rounds of data collection.

Table 4-7. Distribution of schools by number of cases (children) with nonzero weights

	1 – 5	6 – 10	11 – 15	16 – 20	21 - 27	Mean cases per school
Fall-kindergarten						
C1CW0	32	42	52	187	636	20
C1PW0	39	45	78	301	486	19
C1CPTW0	41	53	109	328	401	18
Spring-kindergarten						
C2CW0	398	45	63	348	520	15
C2PW0	439	51	114	421	387	13
C2CPTW0	85	59	163	428	300	17
Longitudinal for base year						
BYCW0	36	40	74	373	423	19
BYPW0	41	57	149	431	267	17
BYCPTW0	46	80	190	414	198	17
BYCOMW0	42	48	112	404	324	18

4.3.2 Weighting Procedures

In general, weights were computed in two stages. In the first stage, base weights were computed. They are the inverse of the probability of selecting the unit—if units were sampled at a rate of 1 in 100, sampled units must be weighted by 100 to represent the entire population. In the second stage,

base weights were adjusted for nonresponse. Nonresponse leads to bias in the survey estimates when the characteristics of the nonrespondents are very different from those of the respondents. Adjusting for nonresponse is intended to reduce the bias.

Nonresponse adjustment cells were generated using variables with known values for both respondents and nonrespondents. Analyses using the Chi-squared Automatic Interaction Detector (CHAID) were conducted to identify variables most highly related to nonresponse. At the school level, school characteristics used for constructing nonresponse cells were the type of school (public, Catholic private, non-Catholic private, or nonsectarian private), the school locale (large city, mid-size city, suburb of large city, suburb of mid-size 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. At the child level, the variables used for constructing nonresponse cells were the type of school, the locale and the geographic region where the school is located; the size classification of the school; and child characteristics such as age group, gender and race-ethnicity, and whether the child moved from the original sampled school (spring-kindergarten only). For the teachers, nonresponse cells were constructed using the type of school, the school locale, region, and the school size classification.

Once the nonresponse cells were determined, the nonresponse adjustment factors are the reciprocals of the response rates within the selected nonresponse cells. A detailed technical description of the nonresponse adjustment procedure can be found in the ECLS-K methodology report (forthcoming).

Response rates are presented in chapter 5 for the different populations and different types of instruments. A detailed analysis of response rates is available that includes a study of nonresponse bias. In this study, the ECLS-K survey estimates are compared with estimates from the sampling frames; they are also compared with estimates from other surveys such as the Current Population Survey (CPS) and the National Household Education Survey (NHES). The study also includes comparison of estimates using the nonresponse adjusted weights with estimates using unadjusted weights. Finally, a nonresponse simulation study is also provided to estimate the potential nonresponse bias.

4.3.3 Computation of School and Teacher Weights

School Base Weights

School base weights were used in calculating teacher and child weights for teacher- and child-level estimates and school administrator weights for school-level estimates. 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.

If schools were selected through the freshening procedure, as described in section 3.2.3, an additional factor equal to the inverse of the selection probability of the district or diocese was included in the base weight. This factor is necessary because new public and Catholic schools were identified through the freshening procedure with the district/diocese, and their selection probability must be conditioned on the probability of selecting that district/diocese within the stratum. This additional factor did not apply to non-Catholic private schools; these were selected directly from lists, and the school base weights were the simple inverse of the school selection probability.

School Administrator Weights (S2SAQW0)

School administrator weights were computed for schools sampled at the beginning of the study (fall-kindergarten) that completed the school administrator questionnaire in spring-kindergarten. The school administrator weight is the school base weight adjusted for school administrator nonresponse.

Teacher Weights (B1TW0 and B2TW0)

At each school sampled at the beginning of the study, all kindergarten teachers were included in the study. Each of these schools was considered a respondent if it had at least one completed teacher questionnaire, part B. The teacher weights were computed in two stages. First, the school base weights were adjusted for school nonresponse. Then, the teacher weights were computed as the school nonresponse adjusted weights adjusted for teacher nonresponse.

4.3.4 Computation of Child Weights

Child Weights

In general, child weights were computed in two stages. In the first stage, school base weights were adjusted for school nonresponse and then multiplied by the poststratified within-school child weights. In the second stage, the resulting weights were adjusted for child nonresponse. The poststratified, within-school child weight is equal to the total number of children in the school divided by the number of children sampled in the school. This is calculated separately for API and non-API children because different sampling rates were used for these two groups of children. Within a school, all API children have the same base weights and all non-API children have the same base weights.

A school was classified as responding using different criteria for fall-kindergarten, spring-kindergarten, and the base year. In fall-kindergarten, responding schools were eligible schools that agreed to cooperate. A school was considered cooperating if it agreed to provide lists of students (for sampling) and teachers (for distributing the teacher questionnaires) and certain information on students that would be used to plan for the assessment. In spring-kindergarten, schools that satisfied at least one of the following conditions were considered respondents: (1) have at least one child assessed in spring-kindergarten, or (2) have at least one sampled LM/not Spanish child who did not pass the Oral Language Development Scale (OLDS) cut score, or (3) have at least one sampled child with disabilities who could not be assessed according to the child's IEP, or (4) have at least one parent interviewed in spring-kindergarten. For the base year, a responding school was one that satisfied at least one of the following conditions: (1) have at least one child assessed in either round, or (2) have at least one sampled LM/not Spanish child in either round, or (3) have at least one sampled child with disabilities in either round, or (4) have at least one parent interviewed in either round. For each set of first-stage child weights, the appropriate school base weights were adjusted for school nonresponse, and then used in the computation of the final child weights.

Child Weights To Be Used with Direct Child Assessment Data (C1CW0, C2CW0, BYCW0)

In fall-kindergarten, responding children for this type of weight were eligible children who had fall-kindergarten scorable cognitive assessment data, or LM/not Spanish children who did not score at

or above the OLDS cut score, but height and weight measurements were collected from them, 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 in kindergarten during fall 1998. A child who transferred to kindergarten in another school between sampling and assessment was considered to be a nonrespondent. In contrast, children who moved to first grade between sampling and assessment were considered ineligible. The fall-kindergarten child weights C1CW0 are the fall-kindergarten first-stage child weights adjusted for fall-kindergarten child nonresponse.

In spring-kindergarten, responding children were classified using rules similar to those used in fall-kindergarten. A child who transferred to another school between rounds and was not followed was considered a nonrespondent; children who moved outside the country were considered ineligible. The spring-kindergarten child weights C2CW0 are the spring-kindergarten first-stage child weights adjusted for spring-kindergarten child nonresponse. The child longitudinal weights BYCW0 were computed as the base year first-stage child weights adjusted for nonresponse. A respondent is defined as a child for whom both C1CW0 and C2CW0 were nonzero.

Table 4-8 shows the number of children who were not assessed due to the following special situations: children who were LM/not Spanish, children with disabilities, children who moved to another school between fall- and spring-kindergarten and who could not be located or because the new school was in a nonsampled county, and children who moved outside of the country or who were deceased. Only the LM/not Spanish and children with disabilities had child weights.

Table 4-8. Number of children who were not assessed due to special situations

	Number of children	
	Unweighted	Weighted
Fall-kindergarten		
With disabilities	88	18,106
LM/Not Spanish	415	39,148
Spring-kindergarten		
With disabilities	70	13,693
LM/Not Spanish	229	20,211
Moved schools in spring	606	129,562
Became ineligible in spring	67	13,340

Child Weights To Be Used with Parent Data (C1PW0, C2PW0, BYPW0)

The child weights C1PW0 (fall-kindergarten) and C2PW0 (spring-kindergarten) to be used with parent interview data are the corresponding first-stage child weights adjusted for nonresponse to the parent interview. In both fall- and spring-kindergarten, 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. The child longitudinal weights BYPW0 were computed as the base year first-stage child weights adjusted for nonresponse. A respondent is defined as a child for whom both C1PW0 and C2PW0 are nonzero. Note that these weights are at the child level even though the data were collected from the parents; they sum to all kindergarten children.

Child Weights To Be Used for Any Cross-Round Combination of Child Direct Assessment Data and Parent Interview Data and Teacher Data (C1CPTW0, C2CPTW0, BYCPTW0)

The child weights C1CPTW0 (fall-kindergarten) and C2CPTW0 (spring-kindergarten) to be used for analysis involving child, parent, and teacher data are the corresponding first-stage child weights adjusted for nonresponse. In both fall- and spring-kindergarten, a respondent for this type of weight was defined as a child who had scorable cognitive assessment data for the corresponding round (or LM/not Spanish children or children with disabilities), whose parent completed the FSQ section of the parent interview for the corresponding round, and whose teacher completed part B of the teacher questionnaire. The child longitudinal weights BYCPTW0 are the first-stage child weights for the base year adjusted for nonresponse. A respondent is defined as a child for whom both C1CPTW0 and C2CPTW0 are nonzero. Again, these weights are used to produce estimates of children even though the source of the data may be parent or teacher.

Child Weights To Be Used With a Single Round of Parent Interview or Teacher Data in Conjunction with Both Rounds of Child Direct Assessment Data (BYCOMW0)

These child longitudinal weights are the base year first-stage child weights adjusted for nonresponse. For this type of weight, a respondent is defined as a child whose (a) fall- and spring-kindergarten cross-sectional weights C1CW0 and C2CW0 are nonzero, and (b) either fall- or spring-

kindergarten cross-sectional weight C1PW0 or C2PW0 is nonzero, and (c) either fall- or spring-kindergarten cross-sectional weight B1TW0 or B2TW0 is nonzero.

4.3.5 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. Any adjustments done 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. For example, the replicate weights for C1CW0 are C1CW1 through C1CW90. The method used to compute the replicate weights and how they can be used to compute the sampling errors of the estimates are described in the section on variance estimation.

4.3.6 Characteristics of Sample Weights

The statistical characteristics of the sample weights are presented in table 4-9 (teacher- and school-level weights) and in table 4-10 (child-level weights). For each type of weight, the number of cases with nonzero weights is presented together with the mean weight, the standard deviation, the coefficient variation (i.e., the standard deviation as a percentage of the mean weight), the minimum weight, the maximum weight, the skewness, the kurtosis, and the sum of weights.

Table 4-9. Characteristics of teacher- and school-level weights

	Number of cases	Mean	Standard deviation	CV (× 100)	Minimum	Maximum	Skewness	Kurtosis	Sum
Teacher									
Fall-kindergarten									
B1TW0	3,047	62.47	44.04	70.50	1.61	506.40	2.59	11.13	190,337
Spring-kindergarten									
B2TW0	3,243	58.64	39.67	67.64	1.60	453.44	2.43	10.09	190,166
School									
Spring-kindergarten									
S2SAQW0	866	83.44	53.07	63.60	6.42	484.64	2.24	8.32	72,260

Table 4-10. Characteristics of child-level weights

	Number of cases	Mean	Standard deviation	CV (× 100)	Minimum	Maximum	Skewness	Kurtosis	Sum
Fall-kindergarten									
C1CW0	19,173	201.63	91.94	45.60	1.64	755.65	1.35	4.85	3,865,946
C1PW0	18,097	213.62	96.19	45.03	2.03	832.40	1.47	5.71	3,865,946
C1CPTW0	17,124	225.76	104.57	46.32	2.17	1,018.25	1.45	5.49	3,865,946
Spring-kindergarten									
C2CW0	19,967	193.49	104.72	54.12	1.60	900.00	2.16	8.20	3,863,512
C2PW0	18,950	203.88	98.75	48.44	1.98	900.00	1.62	5.91	3,863,512
C2CPTW0	17,454	221.35	107.58	48.60	2.17	918.89	1.47	5.43	3,863,512
Longitudinal for base year									
BYCW0	18,211	212.14	119.54	56.35	1.59	900.00	2.45	9.82	3,863,204
BYPW0	16,906	228.51	109.75	48.03	2.22	900.00	1.62	5.63	3,863,204
BYCPTW0	15,420	250.53	121.33	48.43	2.54	1,146.11	1.58	6.07	3,863,204
BYCOMW0	17,060	226.45	126.48	55.85	1.59	900.00	2.33	8.82	3,863,204

The difference in the estimate of the population of teachers or students (sum of weights) between rounds of data collection is due to a combination of factors, among them: (1) the increase in the number of responding schools in spring-kindergarten that resulted from the refusal conversion efforts, and (2) the number of teachers and students who became ineligible after fall-kindergarten. The population of inference for all child-level weights is always the population of kindergartners in the school year 1998-99.

4.4 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 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.

4.4.1 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:

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

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 3.5 were produced using WesVar and the paired jackknife replication method.

Replicate Weights

Replicate weights were created to be used in the calculation of replicate 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 ECLS-K, replicate weights were created taking into account the Durbin method of PSU selection.[5] As mentioned in section 4.1, the Durbin method selects two first-stage units per stratum without replacement, with probability proportional to size and a known joint probability of inclusion.

[5] Durbin, J. (1967). Design of Multi-Stage Surveys for the Estimation of Sampling Errors. *Journal of the Royal Statistical Society C*, 16, 152-164.

In the ECLS-K PSU sample design, there were 24 SR strata and 38 non-self-representing (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 above, 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, type of school (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. Any adjustments done 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. For example, the replicate weights for child-level weights C1CW0 are C1CW1 through C1CW90.

4.4.2 Taylor Series Method

The Taylor Series method produces a linear approximation to 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 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 weight in each round of data collection and in the panel, while the combining of strata due to inadequate sample size and the sequential numbering of strata and first-stage units were done separately. 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 and STATA. They are described in table 4-11.

4.4.3 Specifications for Computing Standard Errors

Specifications for computing standard errors are given in table 4-12. For each type of analysis described in the table, users can choose between the replication method or the Taylor Series method for computing standard errors.

For the replication method using WesVar, the case identification (ID), the full sample weight, the replicate weights, and the method of replication are required parameters. All analysis of the ECLS-K data should be done using the paired jackknife method (JK2). As an example, to compute 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, C1CW0 as the full sample weight, C1CW1 to C1CW90 as the replicate weights, and JK2 as the method of replication.

Table 4-11. ECLS-K Taylor Series stratum and first-stage unit identifiers

	Variable name	Description
Child level	C1TCWSTR	Sampling stratum – fall-kindergarten C-weights
	C1TCWPSU	First-stage sampling unit within stratum – fall-kindergarten C-weights
	C1TPWSTR	Sampling stratum – fall-kindergarten P-weights
	C1TPWPSU	First-stage sampling unit within stratum – fall-kindergarten P-weights
	CICPTSTR	Sampling stratum – fall-kindergarten CPT-weights
	C1CPTPSU	First-stage sampling unit within stratum – fall-kindergarten CPT-weights
	C2TCWSTR	Sampling stratum – spring-kindergarten C-weights
	C2TCWPSU	First-stage sampling unit within stratum – spring-kindergarten C-weights
	C2TPWSTR	Sampling stratum – spring-kindergarten P-weights
	C2TPWPSU	First-stage sampling unit within stratum – spring-kindergarten P-weights
	C2CPTSTR	Sampling stratum – spring-kindergarten CPT-weights
	C2CPTPSU	First-stage sampling unit within stratum – spring-kindergarten CPT-weights
	BYCWSTR	Sampling stratum – base year panel C-weights
	BYCWPSU	First-stage sampling unit within stratum – base year panel C-weights
	BYPWSTR	Sampling stratum – base year panel P-weights
	BYPWPSU	First-stage sampling unit within stratum – base year P-weights
	BYCPTSTR	Sampling stratum – base year panel CPT-weights
	BYCPTPSU	First-stage sampling unit within stratum – base year CPT-weights
	BYCOMSTR	Sampling stratum – base year panel COM-weights
	BYCOMPSU	First-stage sampling unit within stratum – base year COM-weights
Teacher level	B1TTWSTR	Sampling stratum – fall-kindergarten weights
	B1TTWPSU	First-stage sampling unit within stratum – fall-kindergarten
	B2TTWSTR	Sampling stratum – spring-kindergarten
	B2TTWPSU	First-stage sampling unit within stratum – spring-kindergarten
School level	S2SAQSTR	Sampling stratum
	S2SAQPSU	First-stage sampling unit within stratum

Table 4-12. Specifications for computing standard errors

Type of Analysis	Full Sample Weight	Computing Standard Errors					Approximating Sampling Errors
		Replication Method (WesVarPC)			Taylor Series Method (SUDAAN & STATA)		DEFT (Average Root Design Effect)
		ID	Replicate Weights	Jacknife Method	Sample Design	Nesting Variables	
Fall-Kindergarten Cross-sectional							
# Child-level							2.154
	C1CW0	CHILDID	C1CW1 - C1CW90	JK2	WR	C1TCWSTR C1TCWPSU	
	C1PW0	PARENTID	C1PW1 - C1PW90	JK2	WR	C1TPWSTR C1TPWPSU	
	C1CPTW0	CHILDID	C1CPTW1 - C1CPTW90	JK2	WR	C1CPTSTR C1CPTPSU	
# Teacher-level	B1TW0	T_ID	B1TW1 - B1TW90	JK2	WR	B1TTWSTR B1TTWPSU	1.629
Spring-Kindergarten Cross-sectional							
# Child-level							2.096
	C2CW0	CHILDID	C2CW1 - C2CW90	JK2	WR	C2TCWSTR C2TCWPSU	
	C2PW0	PARENTID	C2PW1 - C2PW90	JK2	WR	C2TPWSTR C2TPWPSU	
	C2CPTW0	CHILDID	C2CPTW1 - C2CPTW90	JK2	WR	C2CPTSTR C2CPTPSU	
# Teacher-level	B2TW0	T_ID	B2TW1 - B2TW90	JK2	WR	B2TTWSTR B2TTWPSU	1.612
# School-level	S2SAQW0	S_ID	S2SAQW1 - S2SAQW90	JK2	WR	S2SAQSTR S2SAQPSU	1.279
Base Year Panel							
# Child-level							1.884
	BYCW0	CHILDID	BYCW1 - BYCW90	JK2	WR	BYCWSTR BYCWPSU	
	BYPW0	PARENTID	BYPW1 - BYPW90	JK2	WR	BYPWSTR BYPWPSU	
	BYCPTW0	CHILDID	BYCPTW1 - BYCPTW90	JK2	WR	BYCPTSTR BYCPTPSU	
	BYCOMW0	CHILDID	BYCOMW1 - BYCOMW90	JK2	WR	BYCOMSTR BYCOMPSU	

For the Taylor Series method using either SUDAAN or STATA, the full sample weight, the sample design, the nesting stratum and PSU variables are required. For the same example above, the full sample weight (C1CW0), the without replacement sample design (WR), the stratum variable (C1TCWSTR) and the PSU variable (C1TCWPSU) must be specified.

The last column in table 4-12 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 4.5.

4.5 Design Effects

An important analytic device is to compare the statistical efficiency of survey estimates 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 called 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:

$$DEFT = \frac{SE_{DESIGN}}{SE_{SRS}}$$

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

4.5.1 Use of Design Effects

One method of computing standard errors for the ECLS-K is the paired jackknife method, as described in section 4.4, using programs designed specifically for analyzing complex survey data such as WesVar. Another approach, Taylor Series linearization (and software designed for it), is also discussed in the same section. If a statistical analysis software package such as SPSS (Statistical Program for the Social Sciences) and SAS (Statistical Analysis System) 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 calculated 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

$$Var_{SRS} = \frac{1}{n}\frac{\sum_{1}^{n} w_i \left(x_i - \overline{x_w}\right)^2}{\sum_{1}^{n} w_i} = SE^2_{SRS}$$

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

$$Var_{srs} = \frac{p(1-p)}{n} = SE^2_{SRS}$$

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 standard error of the estimate under the actual design.

4.5.2 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 child, parent, teacher, and school data. For each survey item, the tables present the number of cases, 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 using the paired jackknife replication method (JK2).

For each survey estimate, the variable name as it appears in the ECLS-K Base Year Public-Use Electronic Code Book (ECB) is also provided in the table. If multiple variables were combined to arrive at the estimate, then the names of all the variables used are provided. For example, the estimate of the proportion of fall-kindergarten children whose home language was not English was computed using two different survey items, P1ANYLNG (parent questionnaire item PLQ020, whether another language was used at home) and P1PRMLNG (parent questionnaire item PLQ060, what was the primary language used at home). The first letter of the variable name indicates the source of the item: C – child assessment, P – parent instrument, A – teacher instrument part A, B – teacher instrument part B, T – teacher instrument part C, and S – school administrator questionnaire. The second letter of the variable name indicates when the data were collected: 1 – round 1, fall-kindergarten; 2 – round 2, spring-kindergarten. 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.

Child-Level Design Effects

Standard errors and design effects for the child-level items are presented in tables 4-13 to 4-15 for fall-kindergarten, spring-kindergarten, and for the base year. The survey items were selected so that there was a mix of items common to both fall- and spring-kindergarten and items that were specific to each round of data collection. For fall- and spring-kindergarten, the student-level items include the different scores from the assessment data, the social rating scores as provided by the parents and teachers, some characteristics of the parents, and some characteristics of the students as reported by the parents. For a small number of estimates, the data were subset to cases where the estimate is applicable; for example, the score for the Spanish math assessment applies only to students who were assessed in Spanish; the type of primary child care is only for children who had regular scheduled child care; the number of hours that the mothers work is only for women in the labor force; and the question on whether the parents have a happy relationship is only for parents in current relationships. For the base year student panel, design effects were calculated for the difference in scores between the two rounds of data collection, and also for some spring-kindergarten items.

The median design effect is 4.7 for fall-kindergarten (compared with 2.2 for the NELS:88 base year student questionnaire data) and 4.1 for spring-kindergarten (compared with 3.4 for the NELS:88 first followup). The size of the ECLS-K design effects is largely a function of the number of children sampled per school. With about 20 children sampled per school, an intraclass correlation of 0.2 might result in a design effect of about 5. The median design effect is 3.4 for the panel of students common to both fall- and spring-kindergarten, and the lower median design effect is due to the smaller cluster size in the panel. The ECLS-K design effects are slightly higher than the average of 3.8 that was anticipated during the design phase of the study, both for estimates for proportions and for score estimates.

Table 4-16 presents the median design effects for subgroups based on school type, child's gender and race-ethnicity, geographic region, and level of urbanicity. For fall-kindergarten, the median design effects vary from 2.0 (Hispanic) to 5.9 (children in small towns and rural areas). For spring-kindergarten and the panel, the range of variability of the median design effects is similar to that for fall-kindergarten; that is, 2.0 for Hispanic children to 6.7 for children in small towns and rural areas for spring-kindergarten and 1.6 for children of other race-ethnicity to 5.3 for children in small towns and rural areas for the panel. Once again, the variation in the design effects is largely a function of the sample size as well as the homogeneity of the children within schools.

Table 4-13. ECLS-K, fall-kindergarten: standard errors and design effects for the full sample – child level

Survey item	Variable name	Number of cases	Estimate	Design SE	SRS SE	DEFT	DEFF
Child scores (Mean)							
Reading score (English)	C1RSCALE	17,625	22.02	0.156	0.063	2.480	6.148
Math score (English)	C1MSCALE	17,615	19.52	0.141	0.054	2.606	6.789
Math score (Spanish)	C1MSCALE	1,021	12.88	0.279	0.132	2.115	4.472
General knowledge (English)	C1GSCALE	17,566	22.09	0.171	0.056	3.077	9.471
Composite motor skills	C1CMOTOR	18,422	12.08	0.050	0.023	2.176	4.734
Approaches to learning-Parent	P1LEARN	18,029	3.11	0.006	0.004	1.678	2.816
Self-control-Parent	P1CONTRO	18,023	2.83	0.006	0.004	1.430	2.045
Social interaction-Parent	P1SOCIAL	18,026	3.32	0.007	0.004	1.702	2.896
Withdrawn-Parent	P1SADLON	18,010	1.54	0.006	0.003	2.131	4.542
Impulsive/overactive-Parent	P1IMPULS	17,902	1.98	0.011	0.005	2.156	4.647
Approaches to learning-Teacher	T1LEARN	18,839	2.96	0.010	0.005	2.110	4.452
Self-control-Teacher	T1CONTRO	18,135	3.07	0.011	0.005	2.371	5.623
Interpersonal-Teacher	T1INTERP	17,923	2.96	0.010	0.005	2.125	4.516
Externalizing problems-Teacher	T1EXTERN	18,609	1.64	0.010	0.005	2.006	4.025
Internalizing problems-Teacher	T1INTERN	18,356	1.55	0.007	0.004	1.867	3.484
Child characteristics (Percent)							
Lived in single-parent family	P1HFAMIL	18,097	23.57	0.728	0.315	2.309	5.331
Lived in two-parent family	P1HFAMIL	18,097	74.43	0.790	0.324	2.438	5.942
Home language is non-English	P1ANYLNG, P1PRMLNG	18,059	12.09	0.724	0.243	2.983	8.896
Primary care is center-based	P1PRIMNW	8,173	36.82	1.195	0.533	2.240	5.019
Primary care is home-based	P1PRIMNW	8,173	63.18	1.195	0.533	2.240	5.019
Expected to graduate from college	P1EXPECT	17,968	74.74	0.703	0.324	2.168	4.702
Being read to every day	P1READBO	18,068	44.44	0.656	0.370	1.774	3.146
Was in excellent/very good/good health	P1HSCALE	18,055	96.92	0.164	0.129	1.274	1.624
Parents had high school or less	WKPARED	17,754	37.99	0.912	0.364	2.503	6.267
Mom worked 35 hours+/week	P1HMEMP	12,519	64.17	0.701	0.429	1.635	2.672
Child characteristics (Mean)							
Age of child in months	R1_KAGE	19,073	68.50	0.077	0.032	2.378	5.656
Child's household size	P1HTOTAL	18,097	4.52	0.022	0.010	2.106	4.436
Number of children <18 in child's HH	P1LESS18	18,097	2.50	0.020	0.009	2.247	5.047
Median						2.162	4.675
Mean						2.154	4.801
Standard deviation						0.406	1.775
Coefficient of variation						0.188	0.370
Minimum						1.274	1.624
Maximum						3.077	9.471

Table 4-14. ECLS-K, spring-kindergarten: standard errors and design effects for the full sample – child level

Survey item	Variable name	Number of cases	Estimate	Design SE	SRS SE	DEFT	DEFF
Child scores (Mean)							
Reading score (English)	C2RSCALE	18,937	31.64	0.213	0.075	2.855	8.151
Math score (English)	C2MSCALE	18,925	27.43	0.173	0.063	2.728	7.443
Math score (Spanish)	C2MSCALE	724	18.76	0.454	0.242	1.874	3.513
General knowledge (English)	C2GSCALE	18,903	26.81	0.158	0.056	2.801	7.844
Approaches to learning-Parent	P2LEARN	18,252	3.12	0.005	0.004	1.279	1.637
Self-control-Parent	P2CONTRO	18,251	2.87	0.006	0.004	1.685	2.839
Social interaction-Parent	P2SOCIAL	18,270	3.42	0.006	0.004	1.579	2.494
Withdrawn-Parent	P2SADLON	18,232	1.55	0.006	0.003	1.966	3.867
Impulsive/overactive-Parent	P2IMPULS	18,091	1.96	0.010	0.005	1.856	3.445
Approaches to learning-Teacher	T2LEARN	18,979	3.08	0.010	0.005	1.992	3.967
Self-control-Teacher	T2CONTRO	18,847	3.15	0.011	0.005	2.294	5.261
Interpersonal-Teacher	T2INTERP	18,767	3.09	0.009	0.005	1.999	3.997
Externalizing problems-Teacher	T2EXTERN	18,907	1.69	0.010	0.005	2.173	4.723
Internalizing problems-Teacher	T2INTERN	18,806	1.59	0.008	0.004	2.062	4.251
Child characteristics (Percent)							
Lived in single-parent family	P2HFAMIL	18,906	23.51	0.584	0.308	1.895	3.591
Lived in two-parent family	P2HFAMIL	18,906	74.33	0.645	0.318	2.030	4.120
Home language is non-English	P2ANYLNG, P2PRMLNG	18,862	12.53	0.739	0.241	3.064	9.391
Child used home computer	P2HOMECM	18,910	52.61	0.818	0.363	2.253	5.078
Child read outside school everyday	P2CHREAD	18,877	39.82	0.649	0.356	1.823	3.322
Parents had high school or less	WKPARED	17,607	37.14	0.910	0.364	2.499	6.244
Parent attended PTA	P2ATTENP	18,914	33.12	0.902	0.342	2.635	6.941
Parent thinks not safe for child to play outside	P2SAFEPL	18,898	96.17	0.220	0.140	1.574	2.479
Parents had happy relationship	P2MARRIG	14,291	97.77	0.164	0.123	1.333	1.776
Parent too busy to play with child	P2TOOBUS	18,600	45.55	0.591	0.365	1.618	2.617
Child characteristics (Mean)							
Age of child in months	R2_KAGE	19,890	74.76	0.070	0.032	2.209	4.881
Child's household size	P2HTOTAL	18,906	4.54	0.023	0.010	2.218	4.919
Number of children <18 in child's HH	P2LESS18	18,906	2.52	0.020	0.009	2.302	5.298
Median						2.030	4.120
Mean						2.096	4.596
Standard deviation						0.458	1.993
Coefficient of variation						0.218	0.434
Minimum						1.279	1.637
Maximum						3.064	9.391

Table 4-15. ECLS-K, panel: standard errors and design effects for the full sample – child level

Survey item	Variable name	Number of cases	Estimate	Design SE	SRS SE	DEFT	DEFF
Child scores							
(Difference between fall- and spring-kindergarten scores)							
Reading score (English)	C2RSCALE – C1RSCALE	16,751	9.88	0.123	0.047	2.603	6.778
Math score (English)	C2MSCALE – C1MSCALE	16,748	8.18	0.084	0.040	2.126	4.520
Math score (Spanish)	C2MSCALE – C1MSCALE	637	6.28	0.165	0.175	0.944	0.892
General knowledge (English)	C2GSCALE – C1GSCALE	16,697	5.16	0.070	0.031	2.255	5.084
Approaches to learning-Parent	P2LEARN - P1LEARN	16,326	0.01	0.005	0.003	1.467	2.151
Self-control-Parent	P2CONTRO - P1CONTRO	16,326	0.05	0.005	0.003	1.461	2.135
Social interaction-Parent	P2SOCIAL - P1SOCIAL	16,331	0.10	0.006	0.005	1.315	1.729
Withdrawn-Parent	P2SADLON – PISADLON	16,298	0.01	0.005	0.003	1.652	2.730
Impulsive/overactive-Parent	P2IMPULS - P1IMPULS	16,109	-0.01	0.009	0.005	1.670	2.790
Approaches to learning-Teacher	T2LEARN - T1LEARN	17,208	0.11	0.007	0.004	1.861	3.462
Self-control-Teacher	T2CONTRO - T1CONTRO	16,538	0.07	0.008	0.004	1.840	3.387
Interpersonal-Teacher	T2INTERP - T1INTERP	16,296	0.12	0.008	0.004	1.919	3.681
Externalizing problems-Teacher	T2EXTERN - T1EXTERN	16,944	0.06	0.005	0.003	1.481	2.193
Internalizing problems-Teacher	T2INTERN - T1INTERN	16,681	0.04	0.007	0.004	1.707	2.915
Child characteristics (Percent)							
Lived in single-parent family	P2HFAMIL	16,870	23.11	0.603	0.325	1.857	3.447
Lived in two-parent family	P2HFAMIL	16,870	74.80	0.671	0.334	2.007	4.029
Home language is non-English	P2ANYLNG, P2PRMLNG	16,906	11.65	0.706	0.247	2.860	8.182
Child used home computer	P2HOMECM	16,881	53.54	0.860	0.384	2.242	5.025
Child read outside school everyday	P2CHREAD	16,855	39.73	0.663	0.377	1.760	3.099
Parents had high school or less	WKPARED	15,733	36.19	0.939	0.383	2.451	6.006
Parent attended PTA	P2ATTENP	16,880	33.24	0.940	0.362	2.594	6.727
Parent thinks not safe for child to play outside	P2SAFEPL	16,866	96.19	0.219	0.147	1.487	2.212
Parents had happy relationship	P2MARRIG	12,823	97.83	0.168	0.129	1.305	1.702
Parent too busy to play with child	P2TOOBUS	16,642	45.73	0.590	0.386	1.528	2.335
Child characteristics (Mean)							
Age of child in months	R2_KAGE	18,146	74.76	0.072	0.033	2.171	4.713
Child's household size	P2HTOTAL	16,870	4.55	0.023	0.011	2.141	4.585
Number of children <18 in child's HH	P2LESS18	16,870	2.52	0.020	0.009	2.167	4.695
Median						1.857	3.447
Mean						1.884	3.748
Standard deviation						0.454	1.767
Coefficient of variation						0.241	0.471
Minimum						0.944	0.892
Maximum						2.860	8.182

Table 4-16. ECLS-K: median design effects for subgroups – child level

Subgroups	Fall-kindergarten[a]		Spring-kindergarten[b]		Panel[c]	
	DEFT	DEFF	DEFT	DEFF	DEFT	DEFF
All students	2.162	4.675	2.030	4.120	1.857	3.447
Type of school						
Public	2.064	4.258	1.932	3.734	1.781	3.171
Private	1.995	3.979	1.954	3.817	1.782	3.174
Catholic private	1.771	3.136	1.738	3.022	1.654	2.736
Other private	1.937	3.754	1.706	2.910	1.709	2.920
Gender						
Male	1.771	3.135	1.735	3.011	1.533	2.349
Female	1.645	2.704	1.656	2.741	1.572	2.471
Race-ethnicity						
White	1.777	3.159	1.802	3.246	1.654	2.736
Black	1.594	2.546	1.462	2.137	1.417	2.009
Hispanic	1.397	1.952	1.406	1.977	1.366	1.865
Asian/Pacific Islander	1.971	3.883	2.107	4.438	1.422	2.021
Other	1.629	2.654	1.503	2.260	1.279	1.635
Region						
Northeast	1.760	3.099	1.824	3.328	1.541	2.374
Midwest	2.366	5.599	2.306	5.319	2.102	4.418
South	2.122	4.502	1.969	3.876	1.945	3.784
West	1.647	2.712	1.666	2.775	1.532	2.347
Urbanicity						
Central city	2.136	4.563	1.952	3.812	1.752	3.068
Urban fringe and large town	1.814	3.291	1.775	3.151	1.586	2.516
Small town and rural area	2.421	5.861	2.594	6.727	2.306	5.319

[a]Each median is based on 28 items.
[b]Each median is based on 27 items.
[c]Each median is based on 27 items.

Items with the highest design effects are those related to teacher data. For example, in spring-kindergarten, the estimate of the percent of children whose teachers have a master's degree or a higher degree is 35.6 percent with a design effect of 16.814; the estimate of the percent of children whose teachers spoke only English in class is 89.5 percent with a design effect of 17.871; the estimate of the mean number of years that these children's teachers taught in schools is nine years with a design effect of 12.157. The median design effect for these three items is about 14.8 for fall-kindergarten and 16.8 for

spring-kindergarten. The high design effects are reasonable for this type of data because children in the same class have the same teacher, and the intraclass correlation is thus high.

Teacher-Level Design Effects

Standard errors and design effects for the teacher-level items are presented in table 4-17 for fall-kindergarten and table 4-18 for spring-kindergarten. Survey items were selected from both teacher instruments, part A and part B. In part A of the teacher instrument, teachers were asked to report about their children's and class's characteristics, for classes that they taught, whether they were morning, afternoon, or all day classes. In part B, they were asked about class organization, class activities, evaluation methods and also about their views on kindergarten readiness, school environment, and overall school climate. The topics covered in part B are not class-specific. Based on data collected in part A, teachers were classified as full-day (if they taught all day classes) or part-day teachers (if they taught morning, or afternoon, or both morning and afternoon classes). For both fall- and spring-kindergarten, a small number of teachers who filled out information for morning and all day classes or afternoon and all day classes could not be classified as part-day or full-day teachers and were excluded from the computation of design effects. This affects items such as the language that the teachers spoke in class and the class size since these estimates were computed over all classes taught, whether they were morning, afternoon, or all day classes.

The median design effect is 2.5 for both fall- and spring-kindergarten. These are lower than the child-level design effects because the number of responding teachers per school is relatively small. The design effect for teachers is largely a result of selecting a sample using a design most effective for child-level statistics.

Table 4-19 presents the median design effects for subgroups based on school type, geographic region, level of urbanicity, teacher type, and percent of minority enrollment in the school. For fall-kindergarten, the median design effect varies from 1.4 (teachers in Catholic private schools) to 3.2 (teachers in schools with 75 percent minority enrollment or more). The median design effects are generally lower for spring-kindergarten, but the range of variability is similar to that for fall-kindergarten (from 1.3 for teachers in Catholic private schools to 3.0 for teachers in schools with 25 to 49 percent minority enrollment).

Table 4-17. ECLS-K, fall-kindergarten: standard errors and design effects for the full sample – teacher level

Survey item	Variable name	Number of cases	Estimate	Design SE	SRS SE	DEFT	DEFF
Teacher characteristics (Percent)							
Used only English in class	A1ATNOOT, A1PTNOOT, A1DTNOOT	2,828	84.97	1.008	0.672	1.501	2.252
Used Spanish in class	A1ATSPNH, A1PTSPNH, A1DTSPNH	2,828	13.83	0.925	0.649	1.425	2.031
Had math area in class	B1MATHAR	3,037	94.71	0.712	0.406	1.753	3.072
Had computer area in class	B1COMPAR	3,031	83.31	1.095	0.677	1.617	2.615
Used 5-9 unpaid prep hours	B1NOPAYP	3,032	37.14	1.022	0.877	1.165	1.357
Had preschoolers in kindergarten	B1INKNDR	2,975	43.00	1.629	0.908	1.795	3.221
Teacher is Hispanic	B1HISP	2,973	6.53	0.766	0.453	1.692	2.862
Had at least a bachelor's degree	B1HGHSTD	2,919	95.20	0.627	0.395	1.587	2.517
Had no teaching certification	B1TYPCER	2,923	3.20	0.401	0.325	1.233	1.520
Had highest teaching certification	B1TYPCER	2,923	62.05	1.409	0.897	1.570	2.464
Certified in early childhood	B1ERLYCT	2,941	52.95	1.607	0.920	1.746	3.048
Taught all day class only	A1ACLASS, A1PCLASS, A1DCLASS	2,860	62.38	2.231	0.906	2.463	6.068
Teacher characteristics (Mean)							
Age of teacher	B1AGE	2,923	41.29	0.249	0.191	1.301	1.693
Class size	A1ATOTAG, A1PTOTAG, A1DTOTAG	2,398	19.70	0.305	0.121	2.511	6.303
Had control of teaching techniques and discipline (scale 1 to 5)	B1CNTRLC	3,023	4.43	0.021	0.015	1.423	2.026
Number of years teaching kindergarten	B1YRSKIN	3,024	8.11	0.173	0.135	1.281	1.642
Median						1.579	2.491
Mean						1.629	2.793
Standard deviation						0.386	1.444
Coefficient of variation						0.237	0.517
Minimum						1.165	1.357
Maximum						2.511	6.303

Table 4-18. ECLS-K, spring-kindergarten: standard errors and design effects for the full sample – teacher level

Survey item	Variable name	Number of cases	Estimate	Design SE	SRS SE	DEFT	DEFF
Teacher characteristics (Percent)							
Used only English in class	A2AENGLS A2PENGLS A2DENGLS	3,037	89.22	0.925	0.563	1.643	2.700
Used Spanish in class	A2ACSPNH A2PCSPNH A2DCSPNH	3,037	11.68	0.935	0.583	1.604	2.573
Directed 2 hours of whole class activities	A2WHLCLS	3,032	33.80	1.132	0.859	1.318	1.737
Had daily reading and language arts	A2OFTRDL	3,063	94.82	0.720	0.400	1.798	3.234
Had math 3-4 times per week	A2OFTMTH	3,051	15.30	0.926	0.652	1.421	2.018
Had physical education 1-2 times per week	A2TXPE	3,060	54.61	2.301	0.900	2.557	6.539
Had adequate child size furniture	A2FURNIT	3,061	72.30	1.137	0.809	1.406	1.977
Attended 3 or more in-service training days	A2INSRVC	3,044	87.32	0.975	0.603	1.617	2.615
Parents see child's work 15 times or more	A2SHARED	3,046	22.36	1.226	0.755	1.624	2.637
Had math area in class	B1MATHAR	3,225	94.67	0.749	0.395	1.894	3.588
Had computer area in class	B1COMPAR	3,220	82.85	1.039	0.664	1.565	2.448
Used 5-9 unpaid prep hours	B1NOPAYP	2,970	37.13	1.054	0.887	1.188	1.412
Had preschoolers in kindergarten	B1INKNDR	3,170	43.42	1.582	0.880	1.797	3.230
Teacher is Hispanic	B1HISP	3,167	6.43	0.754	0.436	1.729	2.991
Had at least a bachelor's degree	B1HGHSTD	3,113	94.76	0.606	0.399	1.517	2.302
Had no teaching certification	B1TYPCER	3,114	3.44	0.445	0.326	1.363	1.858
Had highest teaching certification	B1TYPCER	3,114	61.50	1.290	0.872	1.479	2.187
Certified in early childhood	B1ERLYCT	3,137	52.50	1.565	0.892	1.755	3.079
Teacher characteristics (Mean)							
Age of teacher	B1AGE	3,111	41.11	0.253	0.187	1.354	1.834
Number of paid aides	A2PDAIDE	3,053	0.90	0.030	0.014	2.148	4.615
Had control on teaching techniques and discipline (scale 1 to 5)	B1CNTRLC	2,962	4.43	0.022	0.015	1.461	2.134
Number of years teaching kindergarten	B1YRSKIN	3,219	8.09	0.162	0.131	1.234	1.522
Median						1.585	2.511
Mean						1.612	2.692
Standard deviation						0.311	1.139
Coefficient of variation						0.193	0.423
Minimum						1.188	1.412
Maximum						2.557	6.539

Table 4-19. ECLS-K: median design effects for subgroups – teacher level

Subgroups	Fall-kindergarten[a]		Spring-kindergarten[b]	
	DEFT	DEFF	DEFT	DEFF
All teachers	1.579	2.491	1.585	2.511
Type of school				
Public	1.542	2.379	1.548	2.397
Private	1.319	1.738	1.243	1.544
Catholic private	1.162	1.360	1.147	1.315
Other private	1.226	1.503	1.212	1.467
Region				
Northeast	1.410	1.987	1.415	2.005
Midwest	1.518	2.305	1.512	2.284
South	1.561	2.437	1.559	2.429
West	1.532	2.349	1.549	2.398
Urbanicity				
Central city	1.682	2.830	1.547	2.393
Urban fringe and large town	1.534	2.356	1.484	2.202
Small town and rural area	1.617	2.616	1.700	2.893
Type of teacher				
Full day	1.513	2.290	1.638	2.683
Part day	1.411	1.990	1.339	1.793
Minority enrollment				
0 – 25%	1.368	1.871	1.367	1.869
25 – 49%	1.765	3.113	1.738	3.022
50 – 74%	1.424	2.027	1.406	1.977
75 – 100%	1.776	3.166	1.453	2.110

[a] Each median is based on 16 items.
[b] Each median is based on 22 items.

School-Level Design Effects

Standard errors and design effects for the school-level items are presented in table 4-20. Survey items are selected from the school administrator questionnaire. For items having to do with children with limited English proficiency (LEP), the data were subset to schools with LEP children. The median design effect is 1.6. Table 4-21 presents the median design effects for subgroups based on school type, geographic region, level of urbanicity, and percent of minority enrollment in the school. They vary from 1.1 for schools in the Northeast region to 2.1 for schools in small towns and rural areas.

Table 4-20. ECLS-K, spring-kindergarten: standard errors and design effects for the full sample – school level

Survey item	Variable name	Number of cases	Estimate	Design SE	SRS SE	DEFT	DEFF
School characteristics (Percent)							
Had a particular focus or emphasis	S2FOCUS	859	19.94	1.736	1.363	1.274	1.622
Used standardized achievement assessment as requirement for admission	S2STNDTE	856	10.84	1.392	1.063	1.310	1.715
Funding levels decreased significantly	S2FUNDLV	854	12.50	1.382	1.132	1.221	1.491
Received federal Title 1 funds this year	S2TT1	862	51.51	1.450	1.704	0.851	0.725
Required kindergartners to wear uniform	S2UNIFRM	858	18.92	1.936	1.337	1.448	2.096
Gave children readiness or placement assessment	S2RDITST	860	61.74	2.391	1.657	1.443	2.082
Tested kindergartners with standardized assessments	S2TESTK	838	33.40	2.030	1.629	1.246	1.552
Offered after- school child care	S2AFTSCH	856	56.61	2.187	1.694	1.291	1.666
Offered pre-kindergarten	S2PREKIN	856	45.47	2.149	1.702	1.263	1.594
Had LEP children	S2LIMENG	857	39.07	1.980	1.667	1.188	1.412
Had translators for LM-LEP families	S2TRANSL	414	77.69	2.928	2.046	1.431	2.047
Offered IEP to disabled children	S2ONIEP	853	80.77	1.714	1.350	1.270	1.613
Principal is male	S2GNDER	848	39.16	2.186	1.676	1.304	1.701
Principal is black	S2RACE3	820	6.94	0.991	0.888	1.116	1.246
Principal has master's degree	S2EDLVL	806	87.49	1.665	1.166	1.428	2.040
School characteristics (Mean)							
Percent LEP students	S2LEPSCH	387	11.99	1.542	1.002	1.539	2.370
Total years as principal	S2TOTPRI	840	10.38	0.300	0.268	1.118	1.249
Median						1.274	1.622
Mean						1.279	1.660
Standard deviation						0.162	0.395
Coefficient of variation						0.127	0.238
Minimum						0.851	0.725
Maximum						1.539	2.370

Table 4-21. ECLS-K: median design effects for subgroups – school level

Subgroups	Spring-kindergarten[a]	
	DEFT	DEFF
All schools	1.274	1.622
Type of school		
Public	1.312	1.721
Private	1.150	1.323
Catholic private	1.105	1.220
Other private	1.079	1.165
Region		
Northeast	1.045	1.092
Midwest	1.202	1.445
South	1.374	1.888
West	1.414	2.000
Urbanicity		
Central city	1.279	1.635
Urban fringe and large town	1.223	1.496
Small town and rural area	1.445	2.088
Minority enrollment		
0 – 25%	1.213	1.471
25 – 49%	1.246	1.552
50 – 74%	1.182	1.396
75 – 100%	1.422	2.023

[a] Each median is based on 17 items.