



Birth Cohort

U.S. Department of Education
Institute of Education Sciences
NCES 2007-084

Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)

Psychometric Report for the 2-Year Data Collection

Methodology Report

August 2007

Carol Andreassen
Philip Fletcher
Westat

Jennifer Park
Project Officer
National Center for
Education Statistics

U.S. Department of Education

Margaret Spellings
Secretary

Institute of Education Sciences

Grover J. Whitehurst
Director

National Center for Education Statistics

Mark Schneider
Commissioner

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

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

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

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

August 2007

The NCES World Wide Web Home Page address is <http://nces.ed.gov>.

The NCES World Wide Web Electronic Catalog is <http://nces.ed.gov/pubsearch>.

This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES World Wide Web Electronic Catalog address shown above.

Suggested Citation:

Andreassen, C., and Fletcher, P. (2007). *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) Psychometric Report for the 2-Year Data Collection* (NCES 2007-084). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Content Contact:

Gail Mulligan
(202) 502-7491
gail.mulligan@ed.gov

ACKNOWLEDGMENTS

Over the past 7 years, many individuals and organizations have contributed to the design and conduct of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B). While it is not possible to name all the individuals who have made significant contributions to this study, we would like to recognize some of those who played a critical role during the development and implementation phases of the ECLS-B.

First, we would like to thank all the children and their parents who participated the study thus far. The parents of these children invited us into their homes and allowed us to work with their children. We would also like to thank the more than 200 field staff and 16 field managers and supervisors who conducted the home visits during both national data collections and the several earlier field tests of the study design and instrumentation. Also, we would like to express our appreciation to the State Vital Registration and Statistics Executives who provided the sample of 2001 birth certificates on which the study is based.

We gratefully acknowledge Marian MacDorman of the National Center for Health Statistics (NCHS) for all her efforts to make the ECLS-B a success. Her tireless work on the project's behalf to help gain the cooperation of the states and her review of the many sets of review documents submitted to the NCHS and state institutional review boards (IRBs) were invaluable.

Westat, under the direction of the National Center for Education Statistics (NCES), designed and conducted the 2-year wave of the study. NCES received day-to-day support from the Education Statistics Services Institute (ESSI), and American Institutes for Research (AIR). NCES would also like to thank Rick Dulaney and Ray Saunders of Westat for the leadership they provided to the programming, systems development, and data processing tasks associated with the study; Richard Hilpert of Westat for overseeing the field operations of the study; and Laura Branden of Westat for her tireless work with the NCHS and state IRBs.

The design of the child assessment battery also benefited from the participation of expert work groups. The design of the Bayley Short Form—Research Edition was guided by experts in child assessment and the Bayley Scales of Infant Development, Second Edition (BSID-II) in particular, and in Item Response Theory. These individuals included Don Rock and Judy Pollack of Educational Testing

Service who provided Item Response Theory analysis expertise; Barbara Wasik of the Johns Hopkins University who advised on child assessment with low-socioeconomic status and language-minority children; and Kathleen Matula who had been involved in the restandardization of the BSID-II for The Psychological Corporation and who advised on item administration and scoring and training of field staff. John Kirkland of Massey University, Palmerston North, New Zealand, and his colleagues, David Bimler and Andrew Drawneek, were vital to the creation of the TAS-45 and the development of the laptop application. Rachel Cohen (Administration on Children, Youth and Families, U.S. Department of Health and Human Services), Jeanne Brooks Gunn and Christy Brady-Smith of Teachers College were consulted about the administration and coding of the Two Bags Task and generously provided a special coding training for Westat's trainers. In addition, Howard Hoffman (National Institute on Deafness and Other Communication Disorders, National Institutes of Health) worked with Philip Dale (University of Missouri) to provide the wordlists for the English and Spanish vocabulary assessments.

We also would like to acknowledge Jerry West, formerly of NCES and currently of Mathematica Policy Research, Inc., whose dedication and extraordinary efforts brought the ECLS-B into being, as well as Jennifer Park (NCES), Kristin Denton Flanagan (AIR), Jodi Jacobson Chernoff (AIR), and Sandy Eyster (AIR), whose efforts on the ECLS-B are reflected in the study instrumentation, data processing activities, and data products. In addition, the authors would like to thank Zeyu Xu, Steve Mistler, and Rob Stillwell of AIR, as well as Chris Chapman, Larry Ogle, Emmanuel Sikali, Ralph Lee, Val Plisko, and Marilyn Seastrom of NCES for their thoughtful reviews of this document.

Sponsoring Agencies

Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)

- National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education (ED)
- National Center for Health Statistics, U.S. Department of Health and Human Services (HHS)
- National Institutes of Health (NIH), U.S. Department of Health and Human Services
 - National Institute of Child Health and Human Development
 - National Institute on Aging
 - National Institute of Mental Health
 - National Institute of Nursing Research
 - National Institute on Deafness and Other Communication Disorders
 - Office of Behavioral and Social Sciences Research
 - National Center on Minority Health and Health Disparities
- Economic Research Service, U.S. Department of Agriculture (USDA)
- Administration on Children, Youth and Families, HHS
- Maternal and Child Health Bureau, Health Resources and Services Administration, HHS
- Office of Special Education Programs, ED
- Office of the Assistant Secretary for Planning and Evaluation, HHS
- Office of Indian Education, ED
- Centers for Disease Control and Prevention, HHS
- Office of Minority Health, HHS

This page is intentionally left blank.

CONTENTS

<u>Chapter</u>		<u>Page</u>
	ACKNOWLEDGMENTS	iii
	LIST OF TABLES	xi
	LIST OF FIGURES	xvi
	LIST OF EXHIBITS	xxii
1	INTRODUCTION	1-1
	1.1 Overview of the ECLS-B	1-2
	1.2 Data Collection Instruments for the 2-Year Data Collection	1-3
	1.3 Literature Reviews	1-5
	1.4 Technical Review Panel	1-7
	1.5 Specialized Child Assessment Work Group and Expert Consultants	1-9
	1.6 Design Change: 18-Month and 30-Month Data Collections and Transition to the 2-Year Data Collection	1-9
	1.7 Organization of This Report	1-11
2	BAYLEY SHORT FORM–RESEARCH EDITION	2-1
	2.1 Decision to Use the Bayley Scales of Infant Development, Second Edition	2-1
	2.1.1 Description of the BSID-II	2-3
	2.1.2 Development of the BSF-R	2-5
	2.1.3 Overview of the 2-PL Response Model	2-6
	2.1.4 Creating the BSF–R for Round 2: Psychometric Rigor and Administrative Ease	2-9
	2.1.5 IRT Analysis and an Adaptive Testing Strategy	2-12
	2.1.6 IRT Item Calibrations of the BSID-II Standardization Dataset	2-13
	2.1.7 Constructing Item Sets for the 18-Month BSF-R Mental and Motor Scales	2-27
	2.1.8 Design of Administration Booklet, Training Materials, and Video	2-45
	2.1.9 Identification of BSF-R Problem Items in the 18-Month Field Test	2-48
3	DEVELOPING THE 2-YEAR BAYLEY SHORT FORM–RESEARCH EDITION	3-1
	3.1 Pilot Testing of 2-Year BSF-R and Results	3-1
	3.2 Psychometrics of the 2-Year BSF-R Mental Scale Core Item Set	3-3
	3.3 Creating the 2-year BSF-R Mental Scale Basal and Ceiling Sets	3-6
	3.4 Content of 2-Year BSF-R Mental Scale	3-7

CONTENTS (continued)

<u>Chapter</u>		<u>Page</u>
	3.5 Projected Standard Error of Measurement for the 2-Year BSF-R Mental Scale	3-9
	3.6 Psychometrics of the 2-Year BSF-R Motor Scale Core Item Set	3-10
	3.7 Creating the 2-Year BSF-R Motor Scale Basal and Ceiling Sets	3-13
	3.8 Content of the 2-Year BSF-R Motor Scale	3-14
	3.9 Subsequent Revisions in Preparation for 2-Year National Training...	3-17
	3.10 Training Procedures and Standardized Training Video	3-18
	3.11 Certification Procedures for the 2-Year BSF-R	3-19
4	CONCURRENT CALIBRATIONS AND EQUATING DESIGN FOR THE BSF-R IN THE 2-YEAR NATIONAL DATA COLLECTION	4-1
	4.1 Results of the BSF-R Adaptive Testing Strategy	4-2
	4.2 Examining the Potential of BSF-R Item Sets	4-6
	4.3 BSF-R Compatibility	4-12
	4.4 BSF-R Conditioning	4-16
	4.5 BSF-R Equating Design	4-24
	4.6 BSF-R Developmental Growth	4-41
	4.7 BSF-R Developmental Indices	4-51
	4.8 BSF-R Precision and Reliability	4-54
	4.9 BSF-R Assessor Effects	4-58
	4.10 BSF-R Proficiency Level Probabilities	4-66
	4.11 BSF-R Differential Test and Item Functioning	4-70
5	BAYLEY SHORT FORM—RESEARCH EDITION SCORES	5-1
	5.1 BSF-R Scoring and Ability Estimates	5-1
	5.2 Expected a Posteriori Ability Estimate	5-3
	5.3 Expected a Posteriori Standard Error of Measurement	5-3
	5.4 BSF-R Scale Scores and <i>T</i> Scores	5-4
	5.5 ECLS-B Proficiency Level Probabilities	5-5
	5.6 ECLS-B Data File	5-12
	5.7 Average BSF-R Scores and Probabilities by Key Demographic Variables	5-14
6	TWO BAGS TASK IN THE ECLS-B	6-1
	6.1 Technical Review Panel Advice	6-1
	6.2 Rationale for Transition to Two Bags Task	6-4
	6.3 Two Bags Task Protocol for In-Home Administration	6-6
	6.4 Two Bags Task Field Staff Training, Trainer Training, and Coder Training	6-7
	6.4.1 Field Staff Training	6-8
	6.4.2 Trainer Training	6-9
	6.4.3 Coder Training	6-10

CONTENTS (continued)

<u>Chapter</u>		<u>Page</u>
	6.5 Two Bags Coding Quality Control Procedures and Reliability	6-11
	6.6 Two Bags Performance in the 2-Year National Data Collection	6-13
	6.7 Correlations of 2-Year Two Bags Task and 9-Month NCATS Scale Scores	6-14
	6.8 Two Bags Task Measures in the 2-Year Data Collection	6-15
7	PHYSICAL MEASUREMENTS	7-1
	7.1 Procedural Differences at 2 Years Compared to 9 Months	7-1
	7.2 Two-Year Physical Measurement Variables on the Data File	7-2
	7.3 Reliability of 2-Year Physical Measurements	7-3
	7.3.1 Training Procedures and Certification	7-3
	7.3.2 Reliability and Data Quality Control During the Data Collection Year	7-8
	7.4 Comparison of 2-Year and 9-Month Physical Measurements	7-9
8	TODDLER’S SECURITY OF ATTACHMENT STATUS	8-1
	8.1 Overview of Attachment Theory	8-1
	8.2 Development of the TAS-45 Items and Adaptation to a Laptop Application	8-6
	8.2.1 Identifying the Core Attachment Items	8-7
	8.2.2 Shortening the Q-Sort Procedure	8-9
	8.2.3 The 39-Item TAS in the 18-Month Field Test	8-10
	8.2.4 Identifying the Subset of Disorganized Items	8-11
	8.3 TAS-45 Protocol in the 2-Year Data Collection	8-13
	8.4 TAS-45 Variables: Hotspots and Traditional Classifications	8-13
	8.5 TAS-45 Training, Certification, and Quality Control	8-20
	8.6 TAS-45 Results in the 2-Year National Data Collection	8-23
9	INTERVIEWER OBSERVATIONS OF THE CHILD AND THE CHILD’S ENVIRONMENT	9-1
	9.1 Interviewer Observations of the Child’s Behavior During the BSF-R	9-1
	9.1.1 Results From the 2-Year National Data Collection	9-4
	9.1.2 Comparison of 2-Year Results With 9-Month Results	9-7
	9.1.3 Training on the Interviewer Observations of the Child’s Behavior	9-7

CONTENTS (continued)

<u>Chapter</u>		<u>Page</u>
	9.2 Interviewer Observations of the Child’s Home Environment: HOME Items.....	9-8
	9.2.1 Comparison of 2-Year Results With 9-Month Results	9-11
	9.2.2 Training for HOME Observation and Parent Interview Items.....	9-12
10	INDIRECT ASSESSMENTS OF THE CHILD IN THE PARENT INTERVIEW	10-1
	10.1 Developmental Milestones	10-1
	10.2 Self-Regulatory Skills (Infant/Toddler Symptom Checklist)	10-7
	10.3 Toddler Vocabulary	10-11
	REFERENCES	R-1
	Appendix A: Intercorrelations of Major Direct Child Assessments	A-1
	Appendix B: Toddler Attachment Sort Items	B-1

LIST OF TABLES

<u>Table</u>		<u>Page</u>
2-1	BSID-II standardization sample: Mental and motor raw scores and index scores means and standard deviations, by age group: 1993.....	2-16
2-2	18-month field test age distribution of children: 2001.....	2-48
4-1	Number of BSF-R items, administrations, observations, and dependencies, by BSF-R scale, round of data collection, and item set: 2001–02 and 2003–04	4-3
4-2	Number and percentage of children, and mean ability estimates, by BSF-R scale, item set, and round of data collection: 2001–02 and 2003–04	4-4
4-3	Descriptive statistics for 900 standardization sample observations scored with publisher calibrations using both the full BSID-II and BSF-R item sets: 2001–02 and 2003–04.....	4-7
4-4	Number and percentage distribution of 900 standardization sample observations scored with publisher calibrations using both the full BSID-II and combined 9-month and 2-year BSF-R item sets, by level of outfit and scale: 1993	4-11
4-5	Cross-classification of number of children assessed/not assessed at 9-month and 2-year round of data collection, by BSF-R scale: 2001–02 and 2003–04.....	4-13
4-6	Mean fit indices for ECLS-B observations scored directly with publisher calibrations, by BSF-R scale and round of data collection: 2001–02 and 2003–04	4-14
4-7	Number and percentage distribution of ECLS-B sample observations scored with publisher calibrations, using combined 9-month and 2-year BSF-R item sets, by level of outfit and scale: 2001–02 and 2003–04	4-15
4-8	Frequency count for ECLS-B longitudinal sample and publisher standardization dataset, by test and months of age: 2001–02 and 2003–04	4-17
4-9	First Non-Equivalent groups with Anchor Test (NEAT) design, by item sets: 2001–02 and 2003–04.....	4-19
4-10	Mean fit indices for ECLS-B observations scored after concurrent calibration, by BSF-R scale and round of data collection: 2001–02 and 2003–04.....	4-22
4-11	Number and percentage of ECLS-B sample children, by level of fit for the BSF-R scales, after concurrent calibration, by BSF-R scale and round of data collection: 2001–02 and 2003–04	4-22

LIST OF TABLES (continued)

<u>Table</u>		<u>Page</u>
4-12	Frequency count and percentage of BSF-R item parameters (using ECLS-B combined 9-month and 2-year data) for NEAT design items exhibiting low, medium, and high levels of DIF when compared with BSID-II item parameters (using publisher data): 2001–02 and 2003–04	4-32
4-13	Second Non-Equivalent groups with Anchor Test (NEAT) design: 2001–02 and 2003–04.....	4-33
4-14	Mean fit indices for ECLS-B observations scored after DIF analysis and concurrent calibration, by BSF-R scale and round of data collection: 2001–02 and 2003–04.....	4-34
4-15	Number and percentage of ECLS-B sample children, by level of fit for ECLS-B observations scored after differential item function analysis and concurrent calibration, by BSF-R scale and round of data collection: 2001–02 and 2003–04.....	4-35
4-16	Standard errors and reliability coefficients for the 9-month and 2-year BSF-R mental and motor-scales: 2001–02 and 2003–04.....	4-57
4-17	Mean proficiency level probabilities for the 10 proficiency level subscales of the BSF-R mental and motor scales at 9 months and 2 years: 2001–02 and 2003–04.....	4-69
4-18	Proficiency level subscales for the BSF-R mental and motor scales: Items in each proficiency level subscale and their IRT ability and discrimination parameters: 2001–02 and 2003–04	4-71
4-19	BSF-R differential test functioning (DTF) statistics (DTF index, RMSE, Alpha, and Beta) for focal group—reference group comparisons on the mental and motor scales: 2001–02 and 2003–04	4-79
4-20	BSF-R differential item functioning (DIF) for mental and motor items that exhibited weighted root mean squared errors (RMSEs) of 0.10 or more: 2001–02 and 2003–04.....	4-81
5-1	ECLS-B descriptive statistics for BSF-R proficiency level subscales, including proficiency level subscale labels, numbers of items in each subscale, mean scale score, and 9-month and 2-year percentages, in the data file: 2003–04	5-11
5-2	BSF-R mental scale and motor scale variable names, variable labels, and range of values, 2-year data collection: 2003–04	5-13
5-3	Recalibrated 9-month BSF-R mental scale and motor scale variable names, variable labels, and range of values, 2-year data collection: 2003–04.....	5-14

LIST OF TABLES (continued)

<u>Table</u>		<u>Page</u>
5-4	Weighted means (and standard deviations) of the BSF-R mental scale and mental probability scores by key demographic variables, 2-year data collection: 2003–04	5-15
5-5	Weighted means (and standard deviations) of the BSF-R motor scale and motor probability scores by key demographic variables, 2-year data collection: 2003–04	5-19
6-1	Average reliability (percent agreement) for subscales of the Two Bags Task for the ECLS-B 2-year data collection: 2003–04	6-13
6-2	Weighted means and standard deviations for the Two Bags Task rating scales in the ECLS-B 2-year data collection: 2003–04	6-14
6-3	Correlation (<i>r</i>) of 2-year Two Bags Task rating scales with 9-month NCATS total scale, total parent scale and total child scale, ECLS-B 9-month and 2-year data collections: 2001–02 and 2003–04	6-15
6-4	Weighted means (and standard deviations) of the Two Bags Task parent scales by key demographic variables, 2-year data collection: 2003–04	6-16
6-5	Weighted means (and standard deviations) of the Two Bags Task child scales by key demographic variables, 2-year data collection: 2003–04	6-19
7-1	Children’s average physical measurements and standard deviations for total sample and by key demographic variables, 2-year data collection: 2003–04	7-4
7-2	Reliability of sets of physical measurements, 2-year data collection: 2003–04	7-9
7-3	Average growth as measured by increases in physical measurements, 2-year and 9-month data collections: 2001–02 and 2003–04	7-10
8-1	Percentage distribution of children’s attachment classifications obtained by interviewer-completed TAS-39 sorts in the 18-month field test: 2001	8-11
8-2	Variable names and descriptions for the TAS-45 hotspots, traditional classification types, TAS classification confidence score, and traditional AQS security and dependency scores obtained in the 2-year ECLS-B data collection: 2003–04	8-16
8-3	Weighted means and standard deviations for TAS-45 hotspot scores in the 2-year ECLS-B data collection: 2003–04	8-24

LIST OF TABLES (continued)

<u>Table</u>		<u>Page</u>
8-4	Weighted percentage distribution for the four types of attachment (variable X2TASCLS [traditional classification scores]) in the 2-year data collection of the ECLS-B: 2003–04	8-24
8-5	Weighted percentages of attachment classifications for the total sample and by key demographic grouping variables in the 2-year data collection: 2003–04	8-26
9-1	Interviewer-completed observations of child behavior in the 2-year BSF-R compared with those in the 9-month data collection: 2003–04	9-3
9-2	Parent interview items on child behavior in the 2-year BSF-R compared with those in the 9-month data collection: 2003–04	9-4
9-3	Variable names, descriptions, average scores, and standard deviations for interviewer-completed observations about child behavior during the BSF-R in the 2-year data collection: 2003–04	9-5
9-4	Variable names, descriptions, average scores, and standard deviations for parent interview items about child behavior during the BSF-R in the 2-year data collection: 2003–04	9-5
9-5	Correlations of major 2-year BSF-R mental and motor scores with the interviewer observation items, 2-year data collection: 2003–04	9-6
9-6	Correlations of major 2-year BSF-R mental and motor scores with the two questions for parents, 2-year data collection: 2003–04	9-7
9-7	Intercorrelations of 9-month and 2-year interviewer observations of child behavior items, 9-month and 2-year data collections: 2001–02 and 2003–04	9-8
9-8	Variable names, variable labels, and summary statistics for interviewer observations of home environment, 2-year data collection: 2003–04	9-10
9-9	Variable names, variable labels, and summary statistics for home environment parent interview items, 2-year data collection: 2003–04	9-10
10-1	Weighted percentages of children who have reached and have not yet reached developmental milestones by the time of assessment in the 2-year data collection, with item names and descriptions, 2-year data collection: 2003–04	10-3
10-2	Average age weighted means and standard deviations for developmental milestones, by key demographic characteristics, 2-year data collection: 2003–04	10-4

LIST OF TABLES (continued)

<u>Table</u>		<u>Page</u>
10-3	Differentiation of regulatory disordered children by items from the 19- to 24-month and 25- to 30-month versions of the Infant/Toddler Symptom Checklist in the ECLS-B: 2003–04.....	10-9
10-4	Self-regulatory item frequency distributions for the total sample, 2-year data collection: 2003–04.....	10-10
10-5	Weighted means and standard deviations for children’s self-regulatory behaviors, by key demographic characteristics, 2-year data collection: 2003–04.....	10-12
10-6	Frequency distribution of children’s vocabulary items in the Parent CAPI Instrument, 2-year data collection: 2003–04.....	10-17
10-7	Frequency distribution of parent reports of children’s syntax, 2-year data collection: 2003–04.....	10-19
10-8	Weighted means and standard deviations of children’s total vocabulary scores for total sample and by key demographic variables, 2-year data collection: 2003–04.....	10-20
10-9	Weighted means and standard deviations for children’s language use/syntax by key demographic characteristics, 2-year data collection: 2003–04.....	10-22

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
2-1	Schematic representation of publisher data, Item Response Theory ability estimates θ_i , and item parameters β_j : 1993.....	2-17
2-2	BSID-II mental scale score means by age: Item Response Theory (IRT) 2-parameter logistic item calibrations using publisher data: 1993.....	2-19
2-3	BSID-II motor scale score means by age: Item Response Theory (IRT) 2-parameter logistic item calibrations using publisher data: 1993.....	2-20
2-4	Item characteristic curve (ICC) for item MEN028 representing the probability of a correct response: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993.....	2-21
2-5	Standard error of measurement for the mental scale: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993.....	2-24
2-6	Standard error of measurement for the motor scale: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993.....	2-25
2-7	Relationship between Item Response Theory (IRT) true score and publisher raw score for the mental scale: IRT 2-parameter logistic item calibrations using publisher data: 1993.....	2-26
2-8	Relationship between Item Response Theory (IRT) true score and publisher raw score for the motor scale: IRT 2-parameter logistic item calibrations using publisher data: 1993.....	2-27
2-9	Standard error of measurement by proficiency level for the publisher-recommended 17- to 19-month BSID-II mental age item set: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993.....	2-29
2-10	Standard error of measurement by proficiency level for the 18-month BSF-R mental core item set: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993.....	2-31
2-11	Standard error of measurement by proficiency level for the BSID-II mental scale 14- to 16-month age item set: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993.....	2-32
2-12	Standard error of measurement by proficiency level for the BSID-II mental scale 23- to 25-month age item set: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993.....	2-33

LIST OF FIGURES (continued)

<u>Figure</u>		<u>Page</u>
2-13	Establishing basal and ceiling rules for the 18-month BSF-R mental core item set using true scale scores: Item Response Theory (IRT) 2-parameter logistic item calibrations using publisher data: 1993.....	2-34
2-14	Projected standard error of measurement by proficiency level for the 18-month BSF-R mental scale: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993	2-36
2-15	Standard error of measurement by proficiency level for the 18-month BSF-R motor scale core set: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993.....	2-38
2-16	Standard error of measurement by proficiency level for the BSID-II 12-month motor age item set: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993	2-39
2-17	Standard error of measurement by proficiency level for the BSID-II motor scale 26- to 28-month age item set: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993.....	2-40
2-18	Establishing basal and ceiling rules for the 18-month BSF-R motor core item set using true scale scores: Item Response Theory (IRT) 2-parameter logistic item calibrations using publisher data: 1993.....	2-41
2-19	Projected standard error of measurement by proficiency level for the 18-month BSF-R motor scale: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993	2-43
2-20	Item Response Theory (IRT) equating of 18-month BSF-R mental scale field test data and publisher standardization data: 1993	2-51
2-21	Item Response Theory (IRT) equating of 18-month BSF-R motor scale field test data and publisher standardization data: 1993	2-52
3-1	Standard error of measurement by proficiency level for the publisher-recommended 23- to 25-month age item set of the mental scale: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993.....	3-4
3-2	Standard error of measurement by proficiency level for the 2-year BSF-R mental scale core item set: Item Response Theory 2-parameter logistic item calibrations using publisher data: 1993.....	3-5

LIST OF FIGURES (continued)

<u>Figure</u>		<u>Page</u>
3-3	Establishing basal and ceiling rules for the 2-year BSF-R mental core item set using true scale scores: IRT 2-parameter logistic item calibrations using publisher data: 1993	3-7
3-4	Projected standard error of measurement by proficiency level for the 2-year BSF-R mental scale: IRT 2-parameter logistic item calibrations using publisher data: 1993	3-10
3-5	Standard error of measurement by proficiency level for the BSID-II motor scale 23- to 25-month age item set: IRT 2-parameter logistic item calibrations using publisher data: 1993.....	3-11
3-6	Projected standard error of measurement by proficiency level for the 2-year BSF-R motor scale core item set: IRT 2-parameter logistic item calibrations using publisher data: 1993	3-12
3-7	Establishing basal and ceiling rules for the 2-year BSF-R motor core item set using true scale scores: IRT 2-parameter logistic item calibrations using publisher data: 1993.....	3-14
3-8	Projected standard error of measurement by proficiency level for the 2-year BSF-R motor scale: IRT 2-parameter logistic item calibrations using publisher data: 1993.....	3-17
4-1	Kernel density estimation for age distributions of children in the 9-month and 2-year ECLS-B data collections: 2001–02 and 2003–04	4-5
4-2	Expected a posteriori ability estimates, using fixed publisher item calibrations throughout; standardization sample observations scored first with the full BSID-II mental (x axis) and then with the BSF-R mental item set (y axis): 1993.....	4-9
4-3	Expected a posteriori ability estimates, using fixed publisher item calibrations throughout; standardization sample observations scored first with the full BSID-II motor (x axis) and then with the BSF-R motor item set (y axis): 1993	4-9
4-4	Expected a posteriori ability estimates for standardization sample observations scored first with publisher item calibrations (full BSID-II mental items) and then scored with ECLS-B item calibrations (BSF-R) following concurrent item calibration: 2001–02 and 2003–04	4-23

LIST OF FIGURES (continued)

<u>Figure</u>		<u>Page</u>
4-5	Expected a posteriori ability estimates for standardization sample observations scored first with publisher item calibrations (full BSID-II motor items) and then scored with ECLS-B item calibrations (BSF-R) following concurrent item calibration: 2001–02 and 2003–04	4-24
4-6	ECLS-B mental item difficulty parameters b_j on the x axis plotted against the corresponding publisher difficulty parameter on the y axis after concurrent calibration: 2001–02 and 2003–04.....	4-26
4-7	ECLS-B motor item difficulty parameters b_j on the x axis plotted against the corresponding publisher difficulty parameter on the y axis after concurrent calibration: 2001–02 and 2003–04.....	4-27
4-8	Test characteristic curves (TCCs) for BSF-R and BSID-II mental scales after concurrent calibration: 2001–02 and 2003–04	4-29
4-9	Test characteristic curves (TCCs) for BSF-R and BSID-II motor scales after concurrent calibration: 2001–02 and 2003–04	4-30
4-10	Item characteristic curves (ICCs) for mental item MEN110 (Names one object) on BSF-R and BSID-II: 2001–02 and 2003–04.....	4-31
4-11	ECLS-B mental item difficulty parameters b_j on the x axis plotted against the corresponding publisher difficulty parameter on the y axis after differential item function analysis and concurrent calibration: 2001–02 and 2003–04.....	4-36
4-12	ECLS-B motor item difficulty parameters b_j on the x axis plotted against the corresponding publisher difficulty parameter on the y axis after differential item function analysis and concurrent calibration: 2001–02 and 2003–04.....	4-37
4-13	Test characteristic curves (TCCs) for BSF-R and BSID-II mental scales after differential item function analysis and concurrent calibration: 2001–02 and 2003–04.....	4-38
4-14	Test characteristic curves (TCCs) for BSF-R and BSID-II motor scales after differential item function analysis and concurrent calibration: 2001–02 and 2003–04.....	4-39
4-15	Expected a posteriori ability estimates for standardization sample observations are scored first with the full publisher BSID-II mental items calibrations and then with ECLS-B mental item calibrations (BSF-R) following concurrent item calibration with the second NEAT design: 2001–02 and 2003–04.....	4-40

LIST OF FIGURES (continued)

<u>Figure</u>		<u>Page</u>
4-16	Expected a posteriori ability estimates for standardization sample observations are scored first with the full publisher BSID-II mental items calibrations and then with ECLS-B motor item calibrations (BSF-R) following concurrent item calibration with the second NEAT design: 2001–02 and 2003–04.....	4-41
4-17	Kernel density estimation for ECLS-B mental latent ability distributions for 9-month and 2-year data collections, in publisher scale metric: 2001–02 and 2003–04.....	4-43
4-18	Kernel density estimation for ECLS-B motor latent ability distributions for 9-month and 2-year data collections, in publisher scale metric: 2001–02 and 2003–04	4-44
4-19	Essential summary of the data showing mental growth in relation to age as estimated in publisher standardization sample and ECLS-B sample data: 2001–02 and 2003–04.....	4-47
4-20	Essential summary of the data showing motor growth in relation to age as estimated in publisher standardization sample and ECLS-B sample data: 2001–02 and 2003–04.....	4-49
4-21	Weighted kernel density estimations for mental developmental index score distributions in the ECLS-B 9-month and 2-year data collections: 2001–02 and 2003–04.....	4-52
4-22	Weighted kernel density estimations for motor developmental index score distributions in the ECLS-B 9-month and 2-year data collections: 2001–02 and 2003–04.....	4-53
4-23	Standard errors of measurement for the BSF-R mental test used in the ECLS-B, across all levels of ability: 2001–02 and 2003–04	4-55
4-24	Standard errors of measurement for the BSF-R motor test used in the ECLS-B, across all levels of ability: 2001–02 and 2003–04	4-56
4-25	Response functions for proficiency level subscales representing 10 developmental milestones on the mental scale: 1993	4-67
4-26	Response functions for proficiency level subscales representing 10 developmental milestones on the motor scale: 1993	4-68
4-27	DTF analysis, showing mental test characteristic curves for BSF-R Asian focal group and White reference group before equating: 2001–02 and 2003–04	4-76

LIST OF FIGURES (continued)

<u>Figure</u>		<u>Page</u>
4-28	DTF analysis, showing mental test characteristic curves for BSF-R Asian focal group and White reference group after equating: 2001–02 and 2003–04	4-77
5-1	Response likelihood function for a specific examinee on the BSF-R mental scale: 2003–04	5-2
8-1	Characteristic profiles of typical attachment styles: Attachment Type B, secure (solid line), Attachment Type A, avoidant (dashed lines), Attachment Type C, ambivalent (dotted line), and Attachment Type D (small-dotted line).....	8-15

LIST OF EXHIBITS

<u>Exhibit</u>		<u>Page</u>
1-1	Sources of data and instruments in the ECLS-B 2-year data collection: 2003–04	1-4
1-2	Components and substantive domains covered in the ECLS-B 2-year direct child assessments: 2003–04	1-5
2-1	Publisher item calibrations for a sample BSID-II item (MEN073, Turns pages of book), using publisher standardization dataset: 1993.....	2-7
2-2	18-month BSF-R mental scale items, by item set: 2003–04.....	2-37
2-3	18-month BSF-R motor items, grouped by core, basal, and ceiling sets: 2001	2-44
2-4	Sample administration page, 18-month field test: 2003–04	2-47
3-1	2-year BSF-R mental scale items, materials, item descriptions, and item age ranges: 2003–04	3-8
3-2	2-year BSF-R motor scale items, materials, item age ranges, and item descriptions: 2003–04	3-15
3-3	Sample item from the BSF-R review form: 2003–04	3-20
4-1	Data structure for a selection of ECLS-B assessments, by child, data collection and interviewer assessor: 2001–02 and 2003–04	4-59
5-1	Changes to 9-month proficiency levels following recalibration using longitudinal 9-month–2-year dataset, original 9-month proficiency level variable names, new 9-month proficiency level variable names, and 2-year proficiency level variable names included within each proficiency: 2001–02 and 2003–04	5-7
8-1	Map of the 145 unique Attachment Q-Sort items	8-8

1. INTRODUCTION

The Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) is a multisource, multimethod study that focuses on the early home and educational experiences of young children. The study is following a nationally representative cohort of children born in 2001 from birth until they enter kindergarten. The first round of the study was conducted when the children were approximately 9 months old; the present phase of the study, round 2, consists of data collected when the children were approximately 2 years of age. Three additional rounds of data collection are planned: one at preschool (about 4 years of age) and two at kindergarten entry.¹

The ECLS-B includes direct and indirect assessments of children's competencies and skills to provide data on their developmental status at a given point in time, as well as growth over time. This report documents the design, construction, implementation, quality control, and psychometric characteristics of the direct and indirect child assessments in the second round of data collection. Direct assessments, as the name suggests, are obtained by directly administering assessments to the children and scoring the results. The indirect assessments are parent respondent or caregiver respondent² ratings of children's characteristics, abilities, or behaviors. This report parallels the *ECLS-B Methodology Report for the Nine-Month Data Collection, Volume 1: Psychometric Characteristics* (NCES 2005–100) (Andreassen and Fletcher 2005). After presenting a brief overview of the ECLS-B and the data collection instruments used in round 2, this chapter discusses the following topics:

- Literature reviews conducted to inform the study design;
- The role of the Technical Review Panel (TRP) in identifying appropriate developmental domains and associated instruments for research and in ensuring the collection of high quality data from the field;
- The child assessment working group;
- The original ECLS-B design and the results of field testing, with subsequent redesign; and
- The organization of the remainder of the report.

¹ There will be two kindergarten rounds of data collection – one in fall 2006 and one in fall 2007. In the fall of 2006, data will be collected from *all* participating sample children, about 75 percent of whom are expected to be enrolled in kindergarten. In the fall of 2007, data will be collected only from the children who are entering kindergarten for the first time, which is expected to be about 25 percent of the sample children.

² The person who provided this information was usually, but not always, the child's parent.

1.1 Overview of the ECLS-B

The ECLS-B is part of the Early Childhood Longitudinal Study (ECLS) program, a longitudinal studies program comprising two cohorts—a birth cohort of children born in 2001 (ECLS-B), and a kindergarten cohort of children who were kindergartners in the 1998-99 school year (ECLS-K). The goal of the ECLS program is to provide high quality data on children’s development and growth in the early childhood years that are useful for researchers, policymakers, practitioners, and parents. Together, the ECLS-B and the ECLS-K provide the breadth and depth of data required to more fully describe children’s health, early learning, development, and education experiences. See <http://nces.ed.gov/ecls> for information about the ECLS program.

The central goal of the ECLS-B is to provide a comprehensive and reliable set of data that may be used to describe and to better understand children’s early development; their health care, nutrition, and physical well-being; their preparation for school; key transitions during the early childhood years; their experiences in early care and education programs and kindergarten; and how their early experiences relate to their later development, learning, and experiences in school. The design of the ECLS-B was guided by three principles. The first principle was to obtain measures of growth through repeated measures at multiple time points. The second was to obtain, wherever possible, direct measures of child functioning rather than to rely on parental reporting in order to reduce potential response bias. The third guiding principle was to obtain information about a broad spectrum of children’s early experiences in order to understand their relationship to children’s development over time.

The first round of data for the ECLS-B was collected when the children born in 2001 were approximately 9 months old. In that round, about 10,700 children and their parents participated; about 9,850 of these children and their parents participated in the second round of data collection when the children were approximately 2 years old. Child assessments were conducted on a total of about 10,200 of these children at 9 months and on about 9,200 of these children at 2 years. Of these 9,200 children, about 8,950 have assessment data for both rounds of data collection. Details on the sample design, sample selection, and data collection can be found in the *ECLS-B Methodology Report for the Nine-Month Data Collection, Volume 2: Sampling* (NCES 2005-147) (Bethel et al. 2005) and in the *User’s Manual for the ECLS-B Longitudinal 9-Month-2-Year Data File and Electronic Codebook* (NCES 2007-046) (Nord et al. 2006). Users who want more detailed information about the 9-month data collection should consult the *ECLS-B User’s Manual for the Nine-Month Restricted-Use Data File and Electronic Codebook* (NCES 2004-092) (Nord et al. 2004).

The ECLS-B is sponsored by the U.S. Department of Education, National Center for Education Statistics (NCES) in the Institute of Education Sciences, in collaboration with several health, education, and human services agencies, including the National Center for Health Statistics (NCHS), the National Institutes of Health (NIH), the Administration on Children, Youth and Families (ACYF), and the U.S. Department of Agriculture (USDA). Westat, a social science research organization, conducted the first two rounds of the study for NCES.

1.2 Data Collection Instruments for the 2-Year Data Collection

The ECLS-B 2-year data collection took place between January 2003 and April 2004. Data were collected by computer-assisted personal interviews (CAPI) with parent respondents,³ self-administered questionnaires given to parent respondents and resident and nonresident fathers (if appropriate),⁴ direct child assessments during an in-person home visit, and from field staff observation of the children's behavior and home setting during the home visit. For children with regular child care arrangements, data were also obtained by computer-assisted telephone interview (CATI) with the child care provider; for a randomly selected sample of children in child care, a direct observation of the child care setting was conducted and, if the setting was center-based, a self-administered questionnaire was given to the director of the center. Exhibit 1-1 lists the sources of data in the 2-year data collection.

³ The respondent to the parent CAPI was usually, but not always, the child's parent. For 1.1 percent of the 9,850 round 2 cases, the respondent was a non-parent relative, such as a grandparent, or a nonrelative.

⁴ If the resident father was not present during the home visit, the father questionnaire was left with the respondent to give to the father to complete. If the father was present, the interviewer gave the questionnaire directly to him. Most resident fathers were the child's biological, adoptive, step-, or foster father. However, for fewer than 100 cases, the Resident Father Questionnaire was completed by someone other than the child's father (e.g., by a grandfather). Interviewers sought the mothers' permission to contact the nonresident father for an interview. If the nonresident father was present, permission still had to be obtained from the mothers before giving the Nonresident Father Questionnaire to him. Only biological fathers were contacted to complete the nonresident father questionnaire.

Exhibit 1-1. Sources of data and instruments in the ECLS-B 2-year data collection: 2003–04

Instruments

Parent computer-assisted personal interviewing (CAPI) Instrument

Direct child assessments (using CAPI, paper and pencil, and videotapes)

Parent Self-Administered Questionnaire

Resident Father Self-Administered Questionnaire

Nonresident Father Self-Administered Questionnaire

Child Observations and Interviewer Remarks Questionnaire

Child Care Provider telephone interview

Child Care Observation for a subset of the children interviewed;
Center Director Self-Administered Questionnaire for center directors

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), 2-year data collection, 2003–04.

The ECLS-B 2-year direct child assessments consisted of four components: the Bayley Short Form–Research Edition (BSF-R), the Two Bags Task, the Toddler Attachment Sort (TAS-45), and physical measurements. Exhibit 1-2 displays the major domains measured during the direct child assessments by each component. Interviewers administered the components using a hard-copy booklet called the Child Activity Booklet, which was available in both English and Spanish. BSF-R item scores and physical measurements were recorded in the Child Activity Booklet. The instructions for the Two Bags Task were also included in the Child Activity Booklet.

Exhibit 1-2. Components and substantive domains covered in the ECLS-B 2-year direct child assessments: 2003–04

Child assessment component	Domain coverage
Bayley Short Form–Research Edition (BSF-R)	Cognitive (mental), physical (motor)
Two Bags Task	Socioemotional functioning, cognitive stimulation fostered by parent, and child’s engagement with parent and willingness to learn
Toddler Attachment Sort	Security of attachment
Physical measurements (height, weight, middle upper arm circumference, head circumference ¹)	Physical growth and development

¹ Head circumference was measured only for ECLS-B sampled children who were very low birth weight.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), 2-year data collection, 2003–04.

The 2-year data collection includes the same three direct assessments of children’s developmental status, socioemotional functioning, and physical growth and development that were included in the 9-month data collection. It is important to recognize that the ECLS-B assessments do not provide information on children’s global mental ability or intelligence quotient (IQ). Instead, they provide descriptive information on children’s early cognitive, motor, and social competencies, which are skills that are important for school readiness and early school success. Assessment in the ECLS-B serves three purposes: (1) to describe children’s developmental status at particular time points, (2) to examine growth in children’s development over time, and (3) to explore the relationship of early experiences to children’s development (where assessment data are the outcome).

1.3 Literature Reviews

To plan the design of the ECLS-B, three literature reviews were prepared for NCES and are available as working papers on the NCES website at <http://nces.ed.gov/ecls/>. These working papers included: (1) *Formulating a Design for the ECLS: Review of Longitudinal Studies* (NCES Working Paper Series, Working Paper No. 97–24) (Green et al. 1997); (2) *A Birth Cohort Study: Conceptual and Design Considerations and Rationale* (NCES Working Paper Series, Working Paper No 1999–01) (Moore et al. 1999); and (3) *Assessment of Social Competence, Adaptive Behaviors, and Approaches to Learning with Young Children* (NCES Working Paper Series, Working Paper No. 96–18) (Meisels, Atkins-Burnett, and

Nicholson 1996). Please refer to chapter 2 of the *User's Manual for the ECLS-B Nine-Month Restricted-Use Data File and Electronic Code Book* (NCES 2004–092) (Nord et al. 2004) for further details about these literature reviews. In addition, Child Trends (a nonprofit research firm located in Washington, DC) reviewed available measures in eight domains considered important for the ECLS-B, including the child assessments (Moore et al. 1999). This review of available measures was particularly important for the design of the indirect and direct child assessments because it summarized specific assessments of children's cognitive and psychomotor development, socioemotional functioning, and physical growth and development. This review was the starting point for the design of the child assessment portion of the ECLS-B. With the scope of the child assessment outlined by these literature reviews, Westat then gathered information to guide the selection of the specific measures to be used.

In addition to the above literature reviews, Westat staff reviewed the questionnaires and child assessments that have been used in comparable large-scale, nationally representative studies to determine their operational feasibility for inclusion in the ECLS-B. At the same time, Westat staff reviewed published direct assessments of children's developmental status (e.g., the Bayley Neurodevelopmental Screener, the Bayley Scales of Infant Development, Second Edition (BSID-II), the Mullen Scales of Early Learning, and the Denver Developmental Screening Test) to evaluate whether any of these were operationally feasible for administration by interviewers in the ECLS-B field setting. The main emphasis was to identify the most likely candidates for a standardized measure of children's developmental status in the ECLS-B and evaluate their respective psychometric and administrative strengths and weaknesses.

In recent decades, the literature on parent-child interaction has accumulated reliable findings of positive associations between maternal sensitivity and responsiveness and children's developmental outcomes. Methodologies described in the literature, as well as in the above-cited literature reviews, were examined to identify feasible methods for obtaining accurate, reliable, and valid information about aspects of parenting and parent-child interaction known to be predictive of children's later adjustment to, and achievement in, formal schooling. This review of available measures of parenting and parent-child interaction also was used to identify emerging areas of interest in developmental and educational psychology to ensure that the ECLS-B included assessments that were relevant to future research needs. For example, in developmental psychology, self-regulation is increasingly regarded as an important aspect of temperament and, therefore, a set of questions was included in the parent interview to obtain information about infant's self-regulatory behaviors (Aksan and Kochanska 2004; Kochanska, Coy, and Murray 2001; Raver 2004; Bornstein and Suess 2000).

1.4 Technical Review Panel

The design of the content of the child assessment component of the ECLS-B has been guided by a Technical Review Panel (TRP). The TRP is a panel of advisors from the research, policymaking, and educational communities who contribute to the ECLS-B by ensuring that it meets the diverse needs of the represented groups. As expert advisors and reviewers, the TRP members help to ensure the success of the ECLS-B in a number of ways, including commenting on overall research priorities, and reviewing and commenting on technical issues. These issues include designing and implementing the ECLS-B; providing information about emerging policy and research topics appropriate for the ECLS-B to address; reviewing questionnaires and assessment instrument content; reviewing draft reports; and reviewing operational practices. An important responsibility of the TRP is to ensure that the plans for conducting the ECLS-B are well thought out and complete, and this responsibility requires a broad range of expertise.

TRP members reviewed the quality of both the design plans and the data collection procedures for the child assessments, and discussed these plans and procedures, as well as alternatives, at the TRP meetings. The members performed the following tasks:

- Verified that the chosen assessments addressed aspects of child development that were determined to be integral to the ECLS-B purpose;
- Assessed whether the chosen instruments were reliable and valid measures of the constructs they were intended to measure;
- Introduced emerging policy issues and research topics to ensure that the information needed to address them was being collected; and
- Reviewed the plans for collecting the child assessments to make sure that their implementation would ensure the collection of high quality data and valid results.

During the early design phases of the 9-month data collection, the TRP meetings took place twice a year in Washington, DC. The TRP members met in a plenary session on the first day, along with representatives from NCES, interagency partners, and Westat staff. After the plenary group discussed general issues, the TRP members then divided into four smaller work groups that covered four content areas: (1) maternal and child health; (2) cognitive and language development and home environment; (3) socioemotional development; and (4) the family's community, father involvement, and child care. The TRP work groups then reported back to the plenary group, and their comments, suggestions, and recommendations were discussed and taken into consideration by NCES, the interagency partners, and Westat.

For the 2-year collection, the TRP, NCES, and interagency partners felt that direct assessment was an important and unique feature of the study and should be continued. With respect to the domains that should be assessed, all were in agreement that children's language use was an important milestone to capture. During children's second year, their language use is burgeoning. Language acquisition is a critical developmental milestone because it is the foundation for mental representation and continued cognitive growth. Language acquisition is also sensitive to environmental and experiential influences, such as mother-child interaction and family literacy. Therefore, it was determined that it was important to include strong measures of toddlers' developmental status, language abilities, socioemotional functioning, physical growth, and well-being, which are most accurately measured through direct assessment of the child.

To measure developmental status, the consensus was to continue to use both the mental and motor scales of the shortened version of the BSID-II, called the Bayley Short Form—Research Edition (BSF-R) that was used in the 9-month collection. The advisors also recommended that children's socioemotional functioning continue to be assessed with an observational measure of parent-child interaction. Noting that the construct of security of attachment is a key developmental milestone for children at this particular age, the advisors recommended that a direct assessment of children's attachment status be included. Given the design of the 2-year ECLS-B home visit, the prime candidate for assessing attachment status was the Attachment Q-Sort (Waters and Deane 1985). However, this tool is too complex for administration in the field by interviewers untrained in attachment theory. (Please see chapter 8 for a discussion of attachment theory.) Therefore, at the suggestion of the TRP members, a simplified and shortened version of the Attachment Q-Sort was developed, the Toddler Attachment Sort-45 item (TAS-45), as described in chapter 8.

A recommendation was also made for the direct assessment of children's physical growth and well-being using standard measurements of physical growth commonly used in health studies, such as the National Health and Nutrition Examination Survey (NHANES; further information is available online at <http://www.cdc.gov/nchs/nhanes.htm>), as well as middle upper arm circumference (MUAC), which is used in health studies conducted by the World Health Organization (WHO). These standard measurements include children's height, weight, MUAC and, for very low birthweight babies (i.e., 1,500 grams or less), head circumference. In addition, since 2-year-old children are able to stand on a scale independently, it was, therefore, possible to assess body mass index (BMI), a commonly used measure of body fat based on height and weight that applies to individuals 2 years of age and older. BMI for children, also referred to as BMI-for-age, is sex and age specific. BMI-for-age is plotted on sex-specific growth

charts. These charts are used for children and teens 2 to 20 years of age. For the 2000 Centers for Disease Control and Prevention (CDC) Growth Charts, please refer to the CDC's NCHS website at <http://www.cdc.gov/growthcharts/>.

The recommended indirect child assessments in the ECLS-B obtained two types of information. The first type was a continuation of the parent-reported assessments of children's developmental functioning used at 9 months, such as the subset of items from the Infant/Toddler Symptom Checklist (DeGangi et al. 1995) and questions about developmental milestones, making age-appropriate modifications. A checklist of words spoken by the child and level of grammar use was also added, to be completed as part of the parent computer-assisted personal interview (CAPI) instrument. This checklist obtained information about the child's language use and complexity of syntax. The second type of indirect measurement obtained information about the child's home experiences and home environment. This information was obtained through interviewer observations and questions asked of the parent respondent.

1.5 Specialized Child Assessment Work Group and Expert Consultants

As design work progressed, a specialized assessment work group was formed to consult on the development of the 2-year BSF-R. In addition, appropriate experts were consulted to guide decisionmaking and the design and implementation of other measures including the Two Bags Task and the TAS-45. The guidance of the assessment work group is discussed in chapter 2. Expert guidance for the development of the coding workshop for the coding of the Two Bags Task is discussed in chapter 6. The development of the TAS-45 is discussed in chapter 8.

1.6 Design Change: 18-Month and 30-Month Data Collections and Transition to the 2-Year Data Collection

Originally there were to be two data collections during the toddler-to-preschool period: one at 18 months and one at 30 months. Planning for the direct assessments, therefore, focused on the development of an 18-month BSF-R and on a 30-month BSF-R. The plan for the direct assessments at 18 months was to develop an age-appropriate version of the BSF-R (analogous to the 9-month BSF-R) to obtain measures of children's cognitive and psychomotor development. The Nursing Child Assessment

Teaching Scale (NCATS), a semi-structured parent-child interaction activity that had been used at 9 months, would again be used as an observational measure of characteristics of the mother-child relationship. The physical measurements would be obtained again using the same procedures as at 9 months.

At the 30-month data collection, an age-appropriate version of the BSF-R would be developed to obtain measures of children's cognitive and psychomotor development. The observational measurement of mother-child interaction would be changed to the Three Bags Task in order to build a foundation for continuity of measurement with the preschool observational measure. Because continuity of measurement was an important guiding principle for the design of the ECLS-B, the plan was to use the NCATS at 9 and 18 months and the Three Bags Task at 30 months and preschool so that the same measure would be used at two data collection points. By 30 months, BMI could also be obtained.

While the 18-month field test was in progress, design work on the 30-month BSF-R and Three Bags Task was conducted. Near the end of the 18-month field test, NCES and the interagency partners decided to combine the 18- and 30-month data collections into a single data collection at 2 years. At this point, the 18-month field test, which included the BSF-R, had been completed. Also, the 30-month BSF-R had been designed and the items pilot tested by child development staff at Westat. In addition, testing the format of a 30-month BSF-R administration booklet and the Three Bags Task had begun. The pilot test of the 30-month BSF-R items gave Westat child development staff direct experience with the items. They could then eliminate items that were not feasible for administration in the field.

When the decision was made to combine the data collections into a single 2-year collection, Westat was able to take the results of the 18-month field test and the 30-month pilot work to produce an age-appropriate 2-year BSF-R and to select age-appropriate mother-child activities for the Three Bags Task at 2 years. The NCATS was not appropriate for 2-year olds so it was not carried into the 2-year design process. The Three Bags Task was selected as an age-appropriate 2-year old mother-child activity. The pilot test of the Three Bags Task for the 30-month data collection indicated that it would take too much time to administer, so a recommendation was made to reduce it to two activities; the Three Bags Task became the Two Bags Task. In addition, the transition to a 2-year data collection also meant that children's BMI could be obtained as discussed earlier. These design changes are discussed in greater detail in chapter 2 for the BSF-R, chapter 6 for the Two Bags Task, and chapter 7 for the physical measurements.

1.7 Organization of This Report

Subsequent chapters describe the specific assessment instruments and sets of questions included in the 2-year ECLS-B data collection and summarize how each performed in the field. Chapter 2 presents a discussion of the decision to include a direct assessment of children’s developmental status, the BSID-II, and the adaptation of the 18-month and 30-month shortened versions of that assessment for use at 2 years. Chapters 3 and 4 describe the work that was done to develop the BSF-R and the Item Response Theory (IRT) analyses conducted with the longitudinal dataset (i.e., combined 9-month and 2-year BSF-R scores). Chapter 5 describes the BSF-R scores included on the longitudinal 9-month–2-year data file. The Two Bags Task—the observational measure of the videotaped parent-child interaction—is summarized in chapter 6, and children’s physical measurements are summarized in chapter 7. Chapter 8 discusses the work that was done to develop the TAS-45—a shortened version of the Attachment Q-Sort—and summarizes scores obtained, and the procedures for training interviewers and maintaining reliability during the year of data collection. Chapter 9 summarizes the remaining observation items that were completed by the interviewer in CAPI after the home visit was completed, including the interviewer observations of child behavior during the BSF-R, and the child’s home environment. Chapter 10 summarizes the indirect assessments of the child in the parent CAPI instrument, including the toddler word list, developmental milestones, and children’s self-regulatory skills. Finally, a table of the direct child assessment intercorrelations is presented in appendix A and the Toddler Attachment Sort items are presented in appendix B. A list of references is provided at the end of this document. Throughout this report, a brief review of key features of the 9-month assessments relevant to the 2-year measures is included when warranted.

Please note that two longitudinal weights that can be found on the longitudinal 9-month–2-year data file were used to obtain the estimates reported in this document. Please see section 4.5.1.2 of the user’s manual for more information on these weights. These two weights are W2R0 and W2C0. Weight W2R0 represents cases with completed parent interviews at both rounds 1 and 2. It was used to estimate child-level characteristics associated with data collected through the parent interview or birth certificate, or both. Examples relevant to child assessment include sets of questions in the parent interview addressing children’s ages when developmental milestones were reached, children’s self-regulation behaviors, and children’s home environments. Weight W2C0 represents cases with completed parent interviews and completed child assessments at both rounds 1 and 2. It was used to estimate child-level characteristics associated with data collected through the child assessments—either alone or in combination with data collected through the parent interview or birth certificate, or both. Examples

include children's scores on the direct assessment of cognitive functioning and psychomotor functioning, children's and primary caregivers' scores on the observational measurement of socioemotional functioning, and children's physical measurements.