ies NATIONAL CENTER FOR SPECIAL EDUCATION RESEARCH
Institute of Education Sciences

Changes in the Characteristics, Services, and Performance of Preschoolers with Disabilities from 2003-04 to 2004-05

Wave 2 Overview Report from the Pre-Elementary Education Longitudinal Study (PEELS)

NCSER 2008-3011

U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR SPECIAL EDUCATION RESEARCH
Institute of Education Sciences

Changes in the Characteristics, Services, and Performance of Preschoolers with Disabilities from 2003-04 to 2004-05

Wave 2 Overview Report from the Pre-Elementary Education Longitudinal Study (PEELS)

Authors:
Elaine Carlson
Tamara Daley
Amy Shimshak
Jarnee Riley
Brad Keller
Frank Jenkins
Joy Markowitz
Westat

Celia Rosenquist
Project Officer
Institute of Education Sciences

NCSER 2008-3011 U.S. DEPARTMENT OF EDUCATION

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Special Education Research
Lynn Okagaki
Acting Commissioner

The National Center for Special Education Research (NCSER) supports a comprehensive research program to promote the highest quality and rigor in research on special education and related services, and to address the full range of issues facing children with disabilities, parents of children with disabilities, school personnel, and others.

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 NCSER product or report, we would like to hear from you. Please direct your comments to

National Center for Special Education Research
Institute of Education Sciences
U.S. Department of Education
555 New Jersey Ave., NW
Washington, DC 20208

The Pre-Elementary Education Longitudinal Study (PEELS) has been funded by the U.S. Department of Education, Institute of Education Sciences, under contract number ED-04-CO-0059. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

This report was prepared for the Institute of Education Sciences under Contract No. ED-04-CO-0059. The project officer is Celia Rosenquist in the National Center for Special Education Research.

May 2008

The NCSER World Wide Web Home Page address is http://ncser.ed.gov
The NCSER World Wide Web Electronic Catalog is http://ncser.ed.gov/pubs
The Pre-Elementary Education Longitudinal Study website is http://www.peels.org

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

Suggested Citation

Carlson, E., Daley, T., Bitterman, A., Riley, J., Keller, B., Jenkins, F., Markowitz, J. (2008). Changes in the Characteristics, Services, and Performance of Preschoolers with Disabilities from 2003-04 to 2004-05, Wave 2 Overview Report from the Pre-Elementary Education Longitudinal Study. Rockville, MD: Westat. Available at www.peels.org.

Content Contact

Celia Rosenquist
(202) 219-2024
Celia.Rosenquist@ed.gov

Contents

List of Tables

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

List of Figures

Executive Summary

The *Education for All Handicapped Children Act (EHA)* (P.L. 94-142) has guaranteed the educational rights of children with disabilities for more than 30 years. And although it has been two decades since the 1986 amendments to *EHA* extended services to preschoolers with disabilities, the characteristics, educational needs, and growth of these children still remain largely unexamined. The Pre-Elementary Education Longitudinal Study (PEELS), funded by the U.S. Department of Education, is examining the characteristics of children receiving preschool special education, the services they receive, their transitions across educational levels, and their performance over time on assessments of academic and adaptive skills.

PEELS includes a nationally representative sample of 3,104 children with disabilities who were ages 3 through 5 when the study began in 2003-04. The sample is divided into three age cohorts (A, B, and C) based on the children's age at first data collection, ages 3, 4, and 5, respectively. The children will be followed through 2009. PEELS data were collected through several different instruments and activities, including a direct one-on-one assessment of the children, a telephone interview with their parents/guardians, and mail questionnaires to the teacher or service provider of each child.

This is the second in a series of overview reports that provide broad findings from the study. This report covers findings from the first two waves of data collection—school year 2003-04 and school year 2004-05.

Changes in Eligibility and Classification Status of Children Who Received Preschool Special Education Services

- At the time they were recruited into PEELS, all children had an active individualized education program (IEP) or individualized family service plan (IFSP). Since then, some have been *declassified*, meaning they are no longer eligible to receive special education services. Seventy-nine percent of children had an IEP/IFSP at both Wave 1 and Wave 2 data collection points; 14 percent had an IEP/IFSP in 2003-04 but not in 2004-05; 2 percent did not have an IEP/IFSP in 2003-04 but had one in 2004-05; and 5 percent did not have an IEP/IFSP at either point in time because they were declassified before the Wave 1 data collection.

- Among preschoolers with disabilities, 52 percent were identified as having primarily speech or language impairments, 27 percent developmental delays, 6 percent autism, and 3 percent mental retardation. Other disability groups accounted for fewer than 3 percent of children who received special education services when they were ages 3 through 5.

- Declassification differed significantly by primary disability category. In percentage terms, 21 percent of children identified as having a speech or language impairment and 21 percent of those identified as having an emotional disturbance were declassified. However, because very few preschoolers were identified as having an emotional disturbance, they made up 1.5 percent of those who were declassified compared with children with speech or language impairments, who accounted for 66 percent of those declassified. The percentage of children with developmental delays, learning disabilities, orthopedic impairments, and low-incidence disabilities who were declassified ranged from 2 percent to 13 percent.

- Urban, suburban, and rural districts differed significantly in the number of young children who were declassified. In all, 22 percent of children from small districts were declassified, whereas 12 percent of children from medium, 15 percent of children from large, and 13 percent of children from very large districts were declassified.

- Twenty percent of children who transitioned from preschool to kindergarten and 24 percent of children who transitioned from kindergarten to first grade between 2003-04 and 2004-05 were declassified, whereas 6 percent of non-transitioning children were declassified.

- Children who were declassified scored significantly higher on the Woodcock-Johnson III: Letter-Word Identification subtest (M = 104.5) than those who remained eligible for services (M = 99.2). The difference between these groups of children was greater on the Applied Problems subtest—declassified students had a mean score of 101.3, and students who retained eligibility had a mean score of 90.1.

- A significant difference between children who were declassified and those who remained eligible for services was also observed on the Peabody Picture Vocabulary Test (PPVT), with means of 96.1 and 87.6, respectively.

- In the year that passed between the first and second data collection, 23 percent of children who continued to receive special education services had a change in their primary disability category, that is, they were *reclassified.* Children who were reclassified scored significantly lower on measures of emerging literacy and math skills than children who were not reclassified (on the PPVT, M = 82.2 for reclassified and M = 89.1 for not reclassified, and on the Woodcock-Johnson III: Applied Problems subtest, M = 82.9 for reclassified; M = 90.0 for not reclassified).

- The opposite of reclassification could be termed *stability* of disability classification. From 2003-04 to 2004-05, the stability of disability classification varied depending on the type of disability, ranging from 89 percent for children with autism to 57 percent for children with other health impairments.

- Of the children initially identified as having a developmental delay as their primary disability, 64 percent retained that disability category.[1] Fourteen percent moved from the developmental delay to the speech or language impairment category, 8 percent moved to the autism category, and 4 percent moved to the learning disability category.

- Children also moved into the developmental delay category from other disability groups. For example, 13 percent of children identified as having an emotional disturbance, 9 percent of children identified as having an orthopedic impairment or other health impairment, and 10 percent of children identified as having a low-incidence disability (deafness, deaf-blindness, vision impairment, or traumatic brain injury) in 2003-04 were identified as having a developmental delay as their primary disability in 2004-05.

- Of children identified as having primarily a speech or language impairment in 2003-04 who continued to receive special education services, 88 percent retained the speech or language impairment category, and 12 percent were reclassified. Six percent were reclassified to

[1] At age 9 (or the age at which states and districts stop using the developmental delay category), children identified as having a developmental delay must either be reclassified into another disability group or declassified.

developmental delay, and each of the other disability categories received fewer than 2 percent.

Changes in Services for Preschoolers With Disabilities

- Teachers indicated that 89 percent of the children received speech or language therapy in the 2003-04 school year, and 86 percent of the children received it in the 2004-05 school year. Occupational therapy (32% in 2003-04 and 35% in 2004-05) and learning strategies/study skills assistance by a special educator (30% in 2003-04 and 20% in 2004-05) were also commonly reported services both years.

- From Wave 1 to Wave 2, there was a statistically significant decrease in the percentage of children receiving nine types of services. For example, service coordination/case management decreased from 25 percent to 9 percent; training, counseling, or other supports/services for the children's family decreased from 16 percent to 5 percent; and learning strategies/study skills assistance decreased from 30 percent to 20 percent. The one service showing a significant increase from 2003-04 to 2004-05 was help from a one-to-one paraeducator or assistant, which increased from 10 to 13 percent.

- The mean number of special education and related services provided to young children with disabilities decreased from 3.5 in 2003-04 to 2.8 in 2004-05.

- From Wave 1 to Wave 2, the mean hours per week children with disabilities spent in a regular education classroom increased significantly, from 8.2 hours to 15.0 hours, and time in special education settings decreased significantly, from 8.0 hours a week to 6.2 hours a week.

Parent Satisfaction With Special Education Services

- Parent satisfaction remained relatively stable from 2003-04 to 2004-05. There was a significant decline only in the percentage of parents who were *very satisfied*, from 47 to 42 percent.

One Year of Growth in the Knowledge and Skills of Preschoolers With Disabilities

- In school year 2003-04, children who received preschool special education services performed close to the average performance of their peers without disabilities on the Woodcock-Johnson III: Letter-Word Identification subtest, with an overall mean performance of 98.2. In school year 2004-05, the mean overall performance increased significantly to 100.2. Mean performance increased significantly for all three age cohorts, for both males and females, for children identified as having a developmental delay, and for children identified as having a speech or language impairment.

- Overall, children who received preschool special education services performed similarly on the PPVT in 2003-04 and 2004-05. The only significant increase in performance on the PPVT was by children identified as having a low-incidence disability, with a mean of 85.2 in 2003-04 and of 90.0 in 2004-05.

- On the Woodcock-Johnson III: Applied Problems subtest, the mean performance of children who received preschool special education services was 90.8 in school year 2003-04. This performance increased significantly to a mean of 91.9 in 2004-05. There were significant increases for children in Cohort A and Cohort C, for males, and for children identified as having the following primary disabilities: developmental delay, learning disability, speech or language impairment, and low-incidence disability.

- Children in Cohort C had a significant increase in performance on the Woodcock-Johnson III: Quantitative Concepts subtest, from 90.9 in 2003-04 to 93.9 in 2004-05. This is the only cohort for which 2 years of data were available for this test.

- Mean teacher ratings on the Preschool and Kindergarten Behavior Scales 2 (PKBS-2) – Social Skills scale for children who received special education services increased significantly from 92.9 in school year 2003-04 to 96.0 in school year 2004-05. There were significant increases for males and females and for children in Cohort A and Cohort B. This scale of the PKBS-2 assesses skills such as "works or plays independently" and "comforts other children who are upset." Children identified as having a developmental delay and children identified as having a speech or language impairment had a significant increase.

- The mean score on the Fine Motor and Gross Motor subscales from the Vineland Adaptive Behavior Scales, Classroom Edition, increased from 94.4 in school year 2003-04 to 96.2 in 2004-05. There were significant increases for children in Cohort A and Cohort C, for males, for children identified as having a learning disability, and for children identified as having a developmental delay.

- Alternate assessments were completed for children who were not capable of participating in the direct assessment or who scored very low on the two English-language subtests that were included in the direct assessment. An alternate assessment was completed for 12 percent of the children in the sample in Wave 1 and 7 percent of the children in the sample in Wave 2. Not all of the children for whom an alternate was completed in Wave 1 also had an alternate completed for them in Wave 2 and vice versa; thus the alternate assessment data for Waves 1 and 2 reflect somewhat different samples. Overall, scores for children for whom an alternate assessment was completed were more than one standard deviation below the population mean for each of the skill areas measured by the Adaptive Behavior Assessment System-II (ABAS-II); in most cases, scores were more than two standard deviations below the population mean.

Chapter 1: Introduction

The *Education for All Handicapped Children Act (EHA)* (P.L. 94-142) has guaranteed the educational rights of children with disabilities for more than 30 years. And although it has been two decades since the 1986 amendments to *EHA* extended services to preschoolers with disabilities, the characteristics, educational needs, and growth of these children still remain largely unexamined, particularly in comparison with school-aged children with disabilities.

The 1997 amendments to the law (now referred to as the *Individuals with Disabilities Education Act (IDEA)*) called for a national assessment of this landmark legislation. The national assessment comprised seven studies: four child-based longitudinal studies covering the age range of children protected under the law as well as three topical studies:

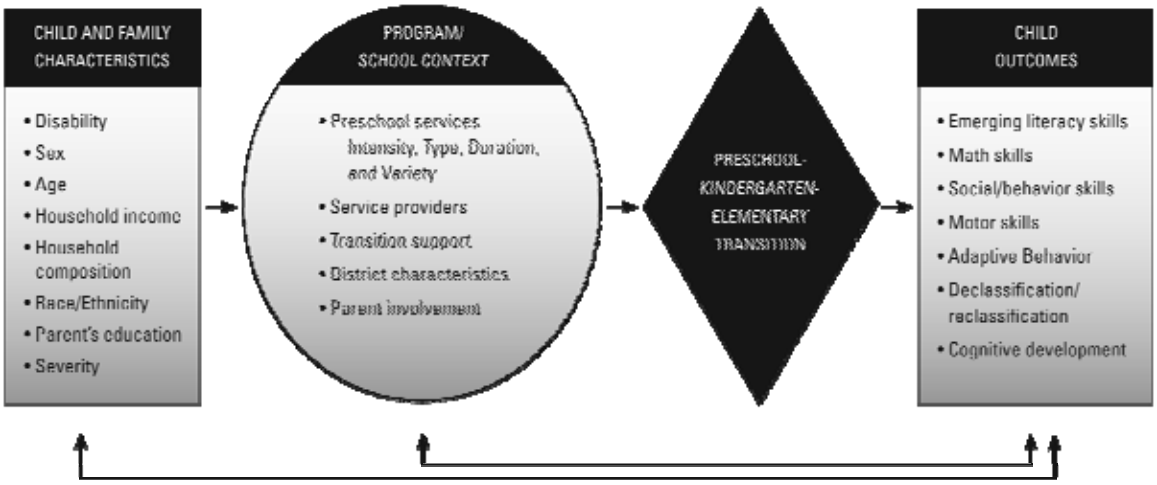
- the National Early Intervention Longitudinal Study (NEILS),
- the Pre-Elementary Education Longitudinal Study (PEELS),
- the Special Education Elementary Longitudinal Study (SEELS),
- the National Longitudinal Transition Study-2 (NLTS-2),
- the Study of Personnel Needs in Special Education (SPeNSE)
- the Study of the State and Local Implementation and Impact of IDEA (SLIIDEA), and
- the Special Education Expenditures Project (SEEP).

PEELS is uniquely suited to examine the preschool and early elementary school experiences of children with disabilities and the outcomes they achieve. The study follows a nationally representative sample of children with disabilities through 2009. Children enter the study in three age cohorts (3-, 4-, and 5-year-olds), and the sample includes children who have a history of receiving Part C (early intervention) services as well as those receiving services for the first time through Section 619 (preschool services). Five broad descriptive research questions guide the data collection, analysis, and reporting for this multiyear study.

- What are the characteristics of children receiving preschool special education?
- What preschool programs and services do they receive?
- What are their transitions like—between early intervention and preschool, and between preschool and elementary school?
- How do these children function and perform in preschool, kindergarten, and early elementary school?
- Which child, service, and program characteristics are associated with children's performance over time on assessments of academic and adaptive skills?

Figure 1 provides a broad model that has guided the PEELS analyses. Earlier PEELS results, which addressed some of the study questions and parallel the model in figure 1, were reported in the Wave 1 Overview Report (Markowitz et al. 2006).

FIGURE 1: OVERALL CONCEPTUAL MODEL FOR PEELS ANALYSIS



Highlights of those findings are as follows:

- For 40 percent of preschoolers with disabilities, concerns were raised about their health or development before 24 months of age, and for 31 percent, concerns were raised between 24 and 36 months.

- Nearly half of preschoolers with disabilities were identified as having a speech or language impairment as their primary disability, and over one-fourth were identified as having a developmental delay as their primary disability. Fewer than 10 percent of preschool children were identified as having other primary disabilities.

- On average, preschoolers with disabilities were nearly 3 years old when they started receiving special education or therapy services from a professional.

- The vast majority of children ages 3 through 5 with disabilities who received special education services received speech or language therapy (93%). Other common services included special education in school (42%), occupational therapy (34%), physical therapy (21%), and tutoring for learning problems (19%).

- Of the children ages 3 through 5 with disabilities who had an individualized family service plan (IFSP) before age 3, nearly one-third (31%) had a gap between the end of early intervention services and the beginning of preschool services.

- Overall, preschoolers with disabilities who participated in the PEELS direct assessment performed close to the population mean on the a test of letter-word identification and within one standard deviation of the population mean on several other tests of emerging literacy and early math proficiency. For children who took an alternate assessment instead of the direct

assessment, scores on a range of functional skills assessments were more than one to two standard deviations below the population mean.

This report is the second in a series of five PEELS overview reports. It supplements findings from the first report and delves more deeply into topics that could not be addressed adequately with a single wave of data. Chapter 2 describes the study design and methods. Chapter 3 presents data on declassification (i.e., children leaving special education) and reclassification (i.e., movement from one primary disability group to another). Chapter 4 describes changes over time in the special education and related services provided to preschoolers with disabilities. Chapter 5 documents the changes in children's performance on a series of direct and indirect assessments in the areas of emerging literacy, early math skills, social behavior, and motor skills from 2003-04 to 2004-05. Six appendices are included in this report. Appendix A contains a diagram of LEA sampling procedures. Appendix B provides detailed information on weighting procedures used in PEELS. Appendix C contains the results of a nonresponse bias study. Appendix D includes standard errors for data tables presented in chapters 3–5. Appendix E provides the number of children who had various test accommodations by gender, race/ethnicity, cohort, and disability category. Appendix F provides a description of all analytic variables used in this report.

A few key points are critical for understanding the data in this report.

- PEELS draws on a national sample of children ages 3 through 5 with disabilities. The sample was selected by age, not by grade, so some of the children were in kindergarten; others were in preschool, day care, or at home. The children may or may not have received early intervention services through Part C of *IDEA* before age 3.

- The data in this report were weighted.[2] Therefore, estimates apply to all children ages 3 through 5 with disabilities in the United States, not just the sample of participating children.

- A small supplemental sample of local education agencies (LEAs) and children was added to PEELS in Wave 2 to account for undercoverage in one region of the country. Wave 1 sampling weights were adjusted at the conclusion of Wave 2 data collection. As such, Wave 1 estimates presented in previous reports were preliminary.

As additional data become available, PEELS researchers will expand upon the findings in this report and address how children's characteristics, services, transitions, and outcomes change over time. For access to PEELS data collection instruments, data tables, and publications, please go to www.peels.org.

[2] Sample weights were used to derive population estimates from the sample. To generate the weighted estimates, sample data were multiplied by the appropriate weight, which reflected the probability of being sampled. For more complete information, see Lee, H., Carlson, E., Lo, A., Fan, J., Chen, L., and Klein, S. (2004). *Final Methodology Report* (Deliverable 13.2 under Contract # ED-01-CO-0082). Rockville, MD: Westat, available at www.peels.org.

Chapter 2: Methods

PEELS is designed to describe children ages 3 through 5 with disabilities and the services they receive; what their transitions are like from early intervention to preschool and preschool to elementary school; and their performance in preschool, kindergarten, and elementary school. This chapter provides basic information on the sample design, data collection instruments and activities, and data analyses.

Sample Design

PEELS used a two-stage sample design to obtain a nationally representative sample of 3- through 5-year-olds receiving special education services. In the first stage, a national sample of LEAs was selected. In the second stage, a sample of preschoolers with disabilities was selected from lists of eligible children provided by the participating LEAs.[3]

We refer to different samples throughout the chapter, so it may be helpful to define them clearly from the outset. The sample selected following the original sample design is called the main sample. This sample was selected by a two-stage design, LEAs at the first stage and children at the second stage. To address nonresponse bias at the LEA level, a nonresponse bias study sample was selected from the nonparticipating LEAs to examine potential differences between the respondents and nonrespondents.[4] The combined sample of the main and the nonresponse study sample is a three-phase sample, where the first phase is the same as the main sample, the second phase is a combined LEA sample comprising the main sample LEAs and the nonresponse study sample LEAs, and the third phase is the sample of children selected from the combined LEA sample. This combined sample was treated as one sample, as if it had been selected with the original sample design and is called the amalgamated sample. In Wave 2, a supplemental sample was selected from a state that was not covered in Wave 1. The amalgamated sample was augmented by adding the supplemental sample and is named the augmented sample. The results presented in this report are based on this augmented sample.

Main LEA Sample

In 2001, 2,752 LEAs were selected from the universe of LEAs serving preschoolers with disabilities, although the target sample size was 210. The universe of LEAs was stratified by four Census regions, four categories of estimated preschool special education enrollment size, and four wealth classes defined on the basis of district poverty level. This resulted in 64 cross-classified stratum cells. The sample of 2,752 LEAs was then divided into many subsamples. Releasing these subsamples one by one, the contractor recruited from the minimum number of subsamples possible to secure participation from 210 LEAs, the target number needed to generate a sufficient number of children in the second stage sample. Ultimately, 709 LEAs were contacted during recruitment, and 245 LEAs agreed to participate. However, a state that contains a considerable portion of the population for its region banned its districts from participating in the study, so they were not even contacted for recruitment. This created a serious under-coverage problem for the study population. This undercoverage was resolved in Wave 2 by randomly selecting a supplemental sample for the state. More details on the supplemental sample are given later in this chapter.

---

[3] In this report, the terms LEA and district are used interchangeably.

[4] Details about the nonresponse study can be found in appendix C.

The design contractor contacted directors of special education and superintendents to secure districts' participation. A participating LEA was required to return a signed agreement affirming that the district would complete the following tasks:

- provide one or more names and contact information for a potential site coordinator for the study;
- allow the site coordinator and other cooperating district staff to recruit families into the study;
- forward contact information from parents who consented to participate in the study;
- allow selected teachers, other service providers, and principals of sampled children to complete a mail questionnaire; and
- allow selected children to participate in a direct assessment, with parental consent.

The design contractor focused recruitment efforts on very large LEAs because a large proportion of the child sample would be selected from these districts, and smaller LEAs could be replaced. Because the initial recruitment occurred in 2001, and data collection did not begin until 2003, researchers recontacted the participating LEAs to confirm their willingness to participate.

In spring 2003, a total of 46 of the 245 LEAs recruited in 2001 dropped out of the study. The 199 remaining LEAs confirmed their participation and began to supply lists of preschool children receiving special education services.

Nonparticipation of a large state in the first phase of LEA recruitment in 2001 created serious undercoverage[5] for the region in which the state is located (We refer to this nonparticipating state as state X). Moreover, a large district in the same geographic region as state X was 1 of the 46 that dropped out in 2003.[6] By spring 2003, the state education agency (SEA) in state X lifted the ban and allowed its districts to participate in the study, and researchers tried to replace the large district in the region that dropped out by sampling four large LEAs from state X in the hope of reducing the undercoverage.[7] Only one of those four LEAs agreed to participate in PEELS, and recruitment of children from the district was very low; therefore, the undercoverage was largely unresolved. This LEA was considered part of the main sample.

To address this undercoverage so the final sample would be nationally representative, a supplemental sample of LEAs, with stratification by size, was randomly selected from state X in Wave 2 (2004-05)—it was too late to do this in Wave 1. The Wave 1 sample, despite the undercoverage problem, was weighted as if state X had been covered by the sample, in the hope of obtaining reasonable national estimates, despite the risk of possible bias. In this way, researchers produced preliminary Wave 1 data.

In Wave 2, the supplemental sample provided data for state X, and researchers used imputation to create missing Wave 1 data for the supplemental sample based on Wave 2 data. All data (child assessment, teacher questionnaire, and parent/guardian interview) except principal and program director questionnaire

---

[5] Undercoverage by a sample indicates that a certain portion of the survey population has no chance of being selected. Because of a state ban, the LEAs in one state had no chance of being selected into the PEELS sample, so it created an undercoverage problem.

[6] This drop out worsened the response rate among the selected LEAs in the region but did not aggravate the undercoverage problem.

[7] Although having some sample from the nonparticipating state would reduce the undercoverage problem, it would not eliminate the problem, because there were still many LEAs that did not have any chance of being selected.

data were imputed for the supplemental sample in Wave 1. Six percent of the augmented sample data for Wave 1 are imputed data, including assessment data. The Wave 1 sample was then reweighted. Therefore, the Wave 1 (2003-04) point estimates and standard errors presented in this report will differ from the preliminary Wave 1 results presented in a previous publication (Markowitz et al. 2006).

In Wave 1, among the contacted 709 LEAs, only 199 LEAs participated in the study. Poor response raised a concern about nonresponse bias. To address it, the U.S. Department of Education funded a comprehensive nonresponse study. In Wave 1, a random sample of 32 LEAs was selected from among the 464 nonparticipating LEAs originally contacted but unsuccessfully recruited. Note that the state ban was still in effect at the time of selection of the nonresponse bias sample, so it was not feasible to include that state in the nonresponse bias study. Because the LEA sample for the nonresponse bias study was small compared to the main LEA sample, it was not possible to use the original LEA sample design (i.e., stratified by geographic region, size category, wealth class)[8], and so only size was used to stratify the 464 nonparticipating LEAs to select the random sample of 32.[9] Twenty-five of those LEAs (78%) initially agreed to participate in the study. This nonresponse study sample was roughly 10 percent of the size of the main LEA sample. Because the results of the nonresponse bias study showed no systematic differences between the respondents and nonrespondents for the key variables we studied (see appendix C for details), the two samples (main and nonresponse bias study) were amalgamated into a single sample as if they had been selected as one based on the original sample design. Nevertheless, this amalgamation could cause some unknown bias in estimates.

This amalgamated sample was then augmented by adding the supplemental sample; this report is based on this augmented sample. Thus, Wave 1 assessment data from the supplemental sample were included in all analyses in this report. The augmented sample, although not selected using the original sample design, is nationally representative because the supplemental sample eliminated the undercoverage issue, and weighting of this sample was done to produce nationally representative estimates.

A diagram[10] in appendix A depicts the sample selection processes for the main sample, which was stratified by size, region, and wealth class, and the nonresponse bias and supplemental samples, both of which were stratified by size only. The final result of the augmented LEA sample, which includes the nonresponse bias study and supplemental samples, is shown by stratum variables (of the main sample) in table 1.

---

[8] If the original sample design was used for the nonresponse bias study, at least half of the 64 possible stratum cells would have been allocated a sample size of zero. This would have created a serious coverage problem because the strata for which no sample was allocated would have had no chance of selection. Using the same stratification is not an issue of representativeness (i.e., coverage) but of efficiency. The notion of sample representativeness is used here to mean that the sample is designed to give every unit in the survey population (represented by the sampling frame) a non-zero probability of selection.

[9] This sample (10 percent of the main LEA sample (245 districts) and with full participation in all aspects of data collection) was considered quite comprehensive to study bias due to nonresponse. To maintain the 64 initial sampling strata, the nonresponse sample would have required resources beyond those available or required for the sample's purposes.

[10] The diagram does not show the intermediary sample of 2,752 LEAs from which a random sample of 709 LEAs was used because the unused portion was simply a reserve sample, which was put back to the frame.

Table 1. Final augmented LEA sample size by three stratification variables

Size				
Total	Very Large	Large	Medium	Small
232	39	42	51	100
Region				
	Northeast	Southeast	Central	West/Southwest
232	66	56	63	47
District wealth				
	High	Medium	Low	Very Low
232	67	67	59	39

[1] Note: The supplemental sample is included only in one region. Region was not used as a stratification factor for the nonresponse bias sample, but the counts include nonresponse bias sample LEAs that happened to fall in the respective regions.
[2] Note: Wealth class was not used as a stratification factor for either the nonresponse bias sample or the supplemental sample, but the counts include their sample LEAs that happened to fall in the respective classes.

Child Sample

In Wave 1, participating districts in the LEA sample submitted lists of eligible children, from which the sample of children was selected. The first was a historical list that asked districts to identify age-eligible children who had an individualized education program (IEP) prior to March 1, 2003 (or an individualized family service plan [IFSP] for districts using IFSPs for children ages 3 through 5)—(see table 2 for age-eligibility). The second set of lists, called ongoing lists, were submitted monthly for 1 year and asked districts to identify newly eligible children in the district by listing children who received their first IEP in the given month. Districts identified children using numbers, rather than names, to maintain confidentiality. Children who transferred from another district with an IEP already in effect were not included on the ongoing lists because they were not newly eligible children.

In Wave 1, the lists of child identification numbers submitted by the districts were checked for ineligible or duplicate cases within and across lists. Errors were corrected through communication with district site coordinators. PEELS researchers began randomly selecting children from historical and ongoing lists late in the 2002-03 school year.[11] The districts continued to send lists of children once a month as the children entered the special education system, and researchers continued to select additional children for the site coordinators to recruit. By the end of Wave 1 family recruitment in May 2004, researchers had selected a sample of 5,259 children.

There are three age cohorts in PEELS: Cohort A comprises 3-year-olds; Cohort B 4-year-olds, and Cohort C 5-year-olds, defined in table 2. Cohort A consists of children in the specified age range who were newly enrolled in the special education program during the recruitment period, and they were to be sampled as they enrolled. These children were on the "ongoing" lists. Cohort B consists of children in the eligible age range who were enrolled before the recruitment period ("historical") and children who were newly enrolled (i.e., ongoing). Cohort C also consists of historical and ongoing children. Thus, there were five combinations of age cohort and historical-ongoing status for each district. These combinations are called child sampling classes.

[11] Sampling rates were based on district-level enrollment counts for children ages 3 through 5 with disabilities.

Historical-list children were sampled using predetermined sampling rates based on the estimated list size and the target sample size, as explained below, when the participating districts provided their historical lists of 4- and 5-year-old children. Children on the ongoing lists were sampled as the districts periodically sent lists of 3-, 4-, and 5-year-olds. Each district had a predetermined sampling rate, which was typically used throughout the recruitment period. However, in some cases, the sampling rates were recalculated based on updated information on district enrollment size, if it was very different from the original estimate.

Table 2. Definition of PEELS age cohorts

Cohort	Age at entry into PEELS	Date of birth
A	3 years old	3/1/00 through 2/28/01
B	4 years old	3/1/99 through 2/29/00
C	5 years old	3/1/98 through 2/28/99

To determine the sampling rates for the five child sampling classes in the main sample, we used district-level sampling weights and district-level child counts, by cohort. The historical sampling rates were generally lower than the ongoing sampling rates within a cohort. Both rates were determined to achieve the target sample sizes for the five child sampling classes, while keeping the weights within the child sampling classes as equal as possible. We obtained district child counts from SEA personnel or websites. Most of the child counts were from December 2003; some were older. Similarly, for the nonresponse bias study sample, the cohort sampling rates were determined in order to reach the target sample sizes (10 percent of the main sample) and to obtain homogeneous child weights within the child sampling classes as much as possible.

One constraint to this procedure was a cap of 80 children for each district. This cap was set so that no individual districts would be overburdened. Although the cap was considered in determining the sampling rates, researchers nonetheless surpassed the cap in a few instances during ongoing sample selection because some large districts submitted lists that included more children than we predicted. During ongoing sample selection in each month, PEELS staff monitored the situation. When the cap was exceeded for a district by a margin of more than 5, the ongoing sample selected for the district that month was reselected so that the overall sample size did not exceed 80, and no further ongoing sample selection was performed for the district.[12]

For the supplemental sample selected in Wave 2, a similar sampling procedure was used to select a child sample, with important exceptions. The age cohort was determined based on the children's age in Wave 1 (see table 2). Furthermore, there was no need to select children on an on-going basis because, in Wave 2, every child was from a historical list. However, to mirror the child sampling process used in Wave 1, the ongoing and historical designations were assigned based on the time of the children's special education enrollment in 2003-04. An additional sample of 542 children was added to the child sample of 5,259 selected in Wave 1, totaling 5,801 sampled children, of whom 3,104 were recruited and took part in the study (2,906 beginning in Wave 1, and 198 beginning in Wave 2).

[12] The overall district sample size was allowed to exceed the cap of 80 by up to 5.

Family Recruitment

Once children were sampled from the historical or ongoing lists, Recruitment Packets were sent to the district site coordinators. Site coordinators were district employees responsible for determining if sampled children were eligible and, if so, inviting their parents or guardians to participate in PEELS. It was necessary to use district employees for this purpose because of the confidentiality of the data on sampled children (i.e., that they were children with disabilities receiving special education services). In addition, district employees had access to information about the names and addresses of parent/guardians and service providers that would not have been available to non-employees. While some family recruitment began in summer 2003, it began in earnest in fall 2003. Recruitment for the supplemental sample occurred in winter-spring 2005. Each recruitment packet included Enrollment Forms (Part 1 and Part 2), a PEELS brochure, a cover letter explaining the study, a PEELS magnet, and a postage-paid return envelope.

Each recruitment packet was arranged according to the unique PEELS identification number assigned to each sampled child. Site coordinators from each district were given a recruitment log, which listed each child's PEELS identification number along with the child's district identification number (submitted on the historical/ongoing lists). Site coordinators were asked to match the identification numbers on the log with the proper child, apply eligibility standards, then invite the eligible families to participate in PEELS. Site coordinators were also encouraged to document the recruitment process using the log.

Part 1 of the PEELS Enrollment Form was eight questions long and was typically filled out by the district's site coordinator before inviting the family to participate in the study. Five of the eight questions on the form asked site coordinators for non-identifying information for each child sampled. PEELS researchers collected these data to test for differences between families that agreed and those that declined to participate in PEELS. The remaining three questions on the Enrollment Form were used to determine the eligibility of each family selected. PEELS had three eligibility criteria:

1. There was an English- or Spanish-speaking adult or an adult who used signed communication in the household who could respond to the telephone interview or alternatively respond using a telephone relay service or interpreter for the hearing impaired.

2. This was the first child in the family sampled for PEELS.

3. The sampled child's family resided in the participating school district at the time of enrollment in PEELS.

If all three eligibility criteria were met, families were given recruitment materials, including a letter explaining the study, the PEELS brochure, and a magnet. The site coordinator informed the family that PEELS is a longitudinal study, that participation is voluntary, and that they could drop out at any time. Site coordinators stressed the study's commitment to confidentiality, ensuring the family that their identity would be protected and that only aggregate data would be reported.

Families that agreed to participate were asked to fill out the PEELS Enrollment Form, Part 2, which asked for identifying information such as names, contact information, the type of services the child received, and the name of the child's teacher or service provider. Once they submitted a signed consent form agreeing to allow PEELS staff to conduct the parent telephone interview, the child assessment, and the teacher/service provider questionnaire, parents received $15. Site coordinators were paid $30 for each family they recruited.

As site coordinators enrolled families to participate in PEELS, their cases were released for the various data collection activities, including the parent telephone interview, the child assessment, and the teacher and program administrator questionnaires.

PEELS researchers received completed enrollment forms for 4,365 children, including the supplemental sample. Based on those enrollment forms, 3,902, or 89.4 percent of families were found eligible. Of those found ineligible, 74 percent no longer lived in the district from which they were sampled; 12 percent did not have an English- or Spanish-speaking adult in the home; and 12 percent had another child sampled for PEELS. Of the eligible families, 79.5 percent agreed to participate. In all, 3,104 families took part in PEELS, which is lower than the 3,550 families anticipated, potentially leading to nonresponse bias. However, the nonresponse bias study revealed no systematic differences between respondents and nonrespondents (see appendix C for details). Also, the set of final recruited families was properly weighted to produce national estimates. Details of the weighting procedure are given in appendix B. (For information on the characteristics of recruited and nonrecruited families, see Carlson 2004).

Nine districts out of 232 that agreed to participate in the study did not recruit any families with eligible children or had no eligible children, and so the final tally of the participating districts in the child-base surveys is 223.[13] This final sample result is tabulated by stratification variables and cohort in tables 3 through 5. Tables 6 and 7 provide final child samples by disability and gender, respectively.

Table 3. The final study sample of districts and children, by size of LEA

	Total	Very Large	Large	Medium	Small
LEA Sample	223	39	42	51	91
Child sample size by cohort					
Total	736	851	729	788	3,104
Cohort A	226	257	238	265	986
Cohort B	300	325	252	248	1,125
Cohort C	210	269	239	275	993

Table 4. The final study sample of districts and children, by region

	Total	Northeast	Southeast	Central	West/Southwest
LEA Sample	223	63	55	59	46
Child sample size by cohort					
Total	3,104	756	727	658	963
Cohort A	986	287	177	210	312
Cohort B	1,125	260	288	226	351
Cohort C	993	209	262	222	300

[13] Child-base surveys are the parent interview, child assessment, and teacher questionnaires. Some of those districts, nevertheless, participated in the LEA questionnaire.

Table 5. The final study sample of districts and children, by wealth

	Total	High	Medium	Low	Very Low
LEA Sample	223	63	55	59	46
Child sample size by cohort					
Total	3,104	848	856	796	604
Cohort A	986	291	296	223	176
Cohort B	1,125	302	306	272	245
Cohort C	993	255	254	301	183

Table 6. The final study sample of children, by disability

	Total	AU	DD	ED	LD	MR	OI	OHI	SLI	LI	No current IEP
Total	3,104	188	806	44	73	86	43	56	1,562	150	96
Cohort A	986	72	328	13	9	23	15	20	443	49	13
Cohort B	1,125	75	280	12	22	30	18	16	590	52	29
Cohort C	993	41	198	19	42	33	10	20	529	49	54

NOTE: AU = Autism; DD = Developmental delay; ED = Emotional disturbance; LD = Learning disability; MR = Mental retardation; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; LI = Low incidence.

Table 7. The final study sample of children, by gender

	Total	Male	Female
Total	3,104	2,189	915
Cohort A	986	692	293
Cohort B	1,125	802	322
Cohort C	993	695	300

Data Collection Instruments and Activities

The PEELS design calls for five waves of data collection during the 6 years from 2003-04 to 2008-09, including several different instruments and activities. As shown in table 8, each of Waves 1 through 4 will include a telephone interview with the participating children's parents/guardians, direct one-on-one assessment of participating children, and mail questionnaires to the teacher or service provider of each child. A final child assessment is planned for Wave 5. Additionally, questionnaires were mailed to SEA, LEA, and program/school administrators to obtain contextual information.

Table 8. PEELS data collection schedule

	Wave 1 2003-04	Wave 2 2004-05	Wave 3 2005-06	Wave 4 2006-07	2007-08	Wave 5 2008-09
Parent/guardian interview	X	X	X	X		
Child assessment	X	X	X	X		X
SEA questionnaire	X					
LEA questionnaire	X	X				
Principal/program director questionnaire	X	X	X			
Teacher questionnaire	X	X	X	X		

NOTE: LEA questionnaires for only the supplemental sample were conducted in Wave 2. In Waves 2 and 3, principal/program director questionnaires were sent only to schools/programs enrolling PEELS children for the first time.

Parent/Guardian Interview

A parent/guardian of each child in the sample was asked to complete four computer-assisted telephone interviews (CATI), one in each of Waves 1-4. The interviews covered the participating child's health and disability, behavior, school programs and services, special education and related services, child care, and out-of-school activities. Respondents were also asked a series of questions about their household, its resources, and family background.

Parent interviews for Wave 1 were conducted between November 2003 and June 2004. Wave 2 interviews were conducted between January 2005 and June 2005. Both averaged about 60 minutes. The interviews were conducted in English or Spanish, based upon respondent preference. In Wave 1, interviews were conducted with 2,802 families, for a 96 percent response rate out of the recruited families.[14] In Wave 2, interviews were conducted with 2,893 families, for a 93 percent response rate. Because 198 families in the supplemental sample were added in Wave 2 to the families recruited in Wave 1, the number of families interviewed in Wave 2 may be higher than in Wave 1, while the percentage interviewed may have decreased or remained unchanged.

Child Assessment

The direct one-on-one assessment was designed to obtain information on the knowledge and skills of preschoolers with disabilities. Child outcome measures were selected based on a number of criteria: their ability to yield individual scores, acceptable reliability and validity studies, brevity, norms in the age ranges under consideration, and maximum opportunity for inclusion of all participating children. In several cases, priority was given to assessments being used in the Head Start National Reporting System and Head Start Impact Study (www.acf.hhs.gov/programs/opre/hs/impact_study/index.html). The Wave 1 and Wave 2 direct assessments averaged about 40 minutes and included the following subtests:

- preLAS 2000 Simon Says (De Avila and Duncan 1998);
- preLAS 2000 Art Show (De Avila and Duncan 1998);
- Peabody Picture Vocabulary Test (Dunn and Dunn 1997);

[14] The response rate for each instrument of data collection (parent telephone interview, child assessment, and teacher questionnaire) is calculated out of the recruited families in its specific wave.

- Woodcock-Johnson III: Letter-Word Identification (Woodcock, McGrew, and Mather 2001);
- Woodcock-Johnson III: Quantitative Concepts (Woodcock, McGrew, and Mather 2001);
- Woodcock-Johnson III: Applied Problems (Woodcock, McGrew, and Mather 2001);
- Leiter-R Attention Sustained Scale (Roid and Miller 1995, 1997);
- Individual Growth and Development Indicators: Picture Naming (ECRI MGD 2001);
- Individual Growth and Development Indicators: Alliteration (ECRI MGD 2001);
- Test of Early Math Skills (US HHS 2005);
- Individual Growth and Development Indicators: Rhyming (ECRI MGD 2001);
- Individual Growth and Development Indicators: Segment Blending (ECRI MGD 2004); and
- PIAT-R Reading Comprehension (Markwardt 1989).

More than 400 assessors were employed and trained to administer the one-on-one assessment with participating children. The assessors included school psychologists, teachers, administrators, and other individuals experienced in administering standardized assessments to young children with disabilities. Some were employees of participating districts. Others were retired or employed by neighboring education agencies or health care providers. The assessors were hired based on their experience in administering standardized assessments to young children with disabilities, and, in many cases, they had experience administering the PEELS assessments themselves, for example, PPVT and Woodcock-Johnson tests of achievement. While using local assessors could potentially threaten the objectivity of the test results, this staffing structure facilitated access to the children and their families, which would have been difficult to obtain using non-local assessors.

Based on specific information from a screening interview with the child's teacher, service provider, or parent/guardian, the assessors were responsible for determining which assessment the child would be given—direct or alternate—and if the child should be referred to a bilingual assessor. An alternate assessment was given if the child could not follow simple directions, had a visual impairment that would interfere with test administration, or if the child began the direct assessment but could not meaningfully participate (e.g., could not attend to the task or did not respond correctly to any items in the first few tests). Assessors also determined if test accommodations were needed based on short interviews with teachers, service providers, or parents. Arrangements for assessments were scheduled with early childhood education programs, elementary schools, teachers, special educators, and parents.

Building on their previous professional experience, PEELS assessors received an initial 1-1/2 day in-person training that was conducted at several locations around the country and was supplemented with video-based instruction on test procedures. The administrative procedures associated with PEELS assessments were explained during the in-person training, and the assessors practiced each subtest following the protocol prescribed for PEELS. Returning assessors completed only video-based training, while replacement assessors received both in-person and video-based instruction.

Assessors were supervised by one of nine Regional Supervisors, who were responsible for recruiting, hiring, and supervising PEELS assessors. During the data collection period, assessors were required to speak with their Supervisors on a bi-weekly basis. These calls were used for answering assessor's questions, conducting any necessary retraining, and case tracking.

In Wave 1, a direct or alternate assessment was completed for 96 percent of the participating children (84% direct, 12% alternate). In Wave 2, a direct or alternate assessment was completed for 94 percent of participating children (87% direct, 7% alternate). Table 9 provides a list of assessments given in each wave and to each age cohort. This report presents results for a subset of direct assessments: PPVT and Woodcock-Johnson III: Letter-Word Identification, Applied Problems, and Quantitative Concepts.

Table 9. Child assessment subtests

	Wave 1			Wave 2		
Subtest	A	B	C	A	B	C
preLAS Simon Says	x	x	x	x	x	x
preLAS Art Show	x	x	x	x	x	x
Peabody Picture Vocabulary (PPVT)	x	x	x	x	x	x
Woodcock-Johnson III: Letter-Word Identification	x	x	x	x	x	x
Woodcock-Johnson III Quantitative Concepts			x		x	x
Woodcock-Johnson III: Applied Problems	x	x	x	x	x	x
Leiter- R Attention Sustained (Test Item AS 1-4)	x					
Leiter- R Attention Sustained (Test Item AS 5-8)		x	x	x	x	
Leiter- R Attention Sustained (Test Item AS 9-12)						x
IGDI Picture Naming	x	x	x	x	x	x
IGDI Alliteration		x	x	x	x	x
IGDI Rhyming		x	x	x	x	x
IGDI Segment Blending		x	x	x	x	x
Early Math Skills	x	x	x	x	x	x
PIAT-R Reading Comprehension						x

The Peabody Picture Vocabulary Test. In this widely used test of receptive language, assessors show children a page with four pictures and ask them to point to the picture of the item that the assessor names. PEELS uses a psychometrically adapted and shortened version of the PPVT-III developed using Item Response Theory (IRT). IRT scaling estimates two aspects of a test. First, it estimates the proficiency scores of each student. Second, it estimates how well a student will do on each item if the student is at a certain level of proficiency. This latter estimate is the item response of IRT. If we know the item response functions of all items, we can predict what total score a student will get if he/she is at a given level of proficiency. These item responses are assumed to be constant from one sample to another in IRT. Because of this invariance of item responses across samples, if two groups are given the same set of items, then the proficiency scales can be linked. Following a method detailed in Stocking and Lord (1983), we can link the proficiency scales between two samples by finding a linear transformation of the proficiency scales that preserves the item responses of the items.

In PEELS, all children completed a core set of PPVT items. Based on their performance on the core, they either took an easier, basal set of items; stopped after the core set; or took a harder (ceiling) set of items. This adaptation was based on the full-length PPVT III and earlier work for the Head Start Family and Child Experiences Survey (FACES) (www.acf.hhs.gov/programs/opre/hs/faces/index.html) and Head Start Impact Study (HSIS) (www.acf.hhs.gov/programs/opre/hs/impact_study/index.html). The 32-item PEELS PPVT was developed using the same approach as the one used for the 40-item HSIS 2002 test. In selecting items for PEELS, the goal was to select a core set of items so 67 percent of the PEELS children (i.e., those scoring within one standard deviation of the mean) would only need to be administered that core set of items (i.e., the core set alone would provide a good estimation of their skills). Easier items on the PPVT were used in the basal set and harder items in the ceiling set. With these adjustments, PEELS Form A (for Wave 1) was constructed with 32 items, 14 core items, 8 basal items, and 10 ceiling items. Children's scores on the various parts of the test were transformed into a single score and placed on a standardized scale with a mean of 100 and a standard deviation of 15.

The IRT true-score for the items in the Form A core set was used to derive basal and ceiling decision rules appropriate for the PEELS target population. The IRT true-score was a model-based estimate of the number-right raw score, which assessors could calculate in the field by adding up the number of correct responses on the core set. We expected about 67 percent of the population to be found between –2.419 and –0.393. These values roughly correspond to 6 correct responses at the low end and 12 correct responses at the high end. Consequently, the basal decision rule stated that six or fewer correct responses required administration of the basal items. In planning the assessment, we expected approximately 16 percent of the children to receive 14 core plus 8 basal items, for a total of 22 items. The ceiling decision rule stated that 12 or more correct responses required administration of the ceiling items. We expected approximately 16 percent of the children to receive 14 core plus 10 ceiling items for a total of 24 items. We expected the remaining 67 percent to receive only the 14 core items, reducing substantially the average time required for completing the subtest. The IRT estimate of test reliability for a population having distribution parameters equal to those of the PEELS latent ability distribution is r_{xx} = 0.781. The sample-based IRT reliability obtained from ability estimates and standard errors of measurement in PEELS is r_{xx} = 0.861.

Since the PEELS adapted version of the PPVT and the full PPVT have a common subset of items, it was possible to apply a Stocking Lord transformation to the proficiencies of the PEELS assessment so that proficiencies were comparable to the national norming sample. Therefore, the PPVT short forms yield the same expected score values as the full PPVT, making the publisher's norms appropriate. The expected score values on the shortened form have somewhat larger standard errors, due to the smaller number of items.[15] The standard version of the PPVT-III had high alternate form reliability for the standardized scores (.88 to .96). Split-half reliability coefficients were also high (.86 to .97). Test-retest reliability coefficients on the PPVT standard form were in the .90s (Dunn and Dunn 1997). Standard form PPVT-III scores were significantly correlated with age; the steepest part of the growth curve occurred from age 2 ½ to 12. Dunn and Dunn (1997) reported that the PPVT-III correlated with the Wechsler Intelligence Scale for Children–Third Edition (Wechsler 1991; r = .82 to .92), Kaufman Adolescent and Adult Intelligence Test (Kaufman and Kaufman 1993; r = .76 to .91), Kaufman Brief Intelligence Test (Kaufman and Kaufman 1990; r = .62 to .82), and the Oral and Written Language Scales (Carrow-Woolfolk 1995; r = .63 to .83). PPVT standard scores were generated for 2,352 PEELS participants in Wave 1 and 2,669 in Wave 2. The estimated reliability of the PPVT short form was .86, meaning that about 86 percent of what the test measured reflected the true underlying construct.

[15] The larger standard errors could be problematic in a clinical setting in which decisions are being made about individuals. However, in a research setting, standardized scores are used for population and subgroup estimates, and the norms allow comparisons with the overall national population of identical age.

The Woodcock-Johnson III: Letter-Word Identification subtest. The Letter-Word Identification test measures the child's word identification skills. Initial items require the child to identify letters that appear in large type, and the remaining items require the child to pronounce words correctly. The child is not required to know the meaning of any word. Test items progress in difficulty from common to uncommon words in written English. Each item is given a score of 1 for a correct response and 0 for an incorrect response or no response. Administration of the test continues until the child either finishes all of the items or misses the last six consecutive items at the end of a test page. The standard score scale used in the Woodcock-Johnson III: Letter-Word Identification is based on a mean of 100 and a standard deviation of 15.

McGrew and Woodcock (2001) reported a 1-year test-retest correlation of .92 for children ages 4 to 7. Test scores were correlated with age (McGrew and Woodcock 2001). They also reported that the complete Woodcock-Johnson III achievement battery was correlated with the Wechsler Individual Achievement Test (Wechsler 1992; r = .79) and the Kaufman Test of Educational Achievement (Kaufman and Kaufman 1985; r = .79). In Wave 1 and Wave 2 of PEELS, 2,434 and 2,711 children, respectively, had standard scores for the Woodcock-Johnson III: Letter-Word Identification subtest.

The Woodcock-Johnson III: Quantitative Concepts subtest. This test measures knowledge of mathematical concepts, symbols, and vocabulary. The subtest is divided into two parts. Part A, Concepts, requires the child to count and identify numbers, shapes, and sequences. Part B, Number Series, requires the child to look at a series of numbers, figure out the pattern, and then provide the missing number in the series. For example, the series 9, 8, 7, __ is presented on a page. The child must tell the assessor what number goes in the blank space. In both sections of the test, children are presented with a picture and asked a question that requires the child to demonstrate familiarity with a particular concept or mathematical operation. For example, children are presented with a picture of five stars of varying sizes. The child is asked to point to the largest star, then to the smallest star. Correct answers are totaled within sections to yield a single Quantitative Concepts score. Finally, the single score is converted to a standardized norm-referenced score with a mean of 100 and a standard deviation of 15.

In order to shorten the length of the assessment for younger children, the Quantitative Concepts test was given only to children in Cohort C in 2003-04. In school year 2004-05, the test was given to children in Cohorts B and C. Test developers reported that scores on the Quantitative Concepts subtest were correlated with age (McGrew and Woodcock 2001). In Wave 1, there were 866 PEELS children who had Woodcock-Johnson III: Quantitative Concepts standard scores and, in Wave 2, there were 1,900 who had standard scores.

The Woodcock-Johnson III: Applied Problems subtest. This test requires the child to analyze and solve math problems. In this test, the assessor presents the child with a picture and asks the child a question, such as "How many dogs are in this picture?" The child must recognize (understand) the request, then perform the correct operation. In this case, the child must count the number of dogs in the picture. The math problems are ordered with increasing difficulty either in the operation the child is required to perform (addition as opposed to subtraction) or in the age-appropriate experience with the particular concept, such as coin identification, telling time, reading temperature, etc. Children were awarded 1 point for each correct answer and 0 for each incorrect answer. The test was terminated when the child either finished all items or missed six consecutive items at the end of a test page. Scores were totaled and converted to a norm-referenced scale with a mean of 100 and a standard deviation of 15. Test developers reported a 1-year test-retest correlation of .92 for children ages 4 to 7 (McGrew and Woodcock 2001). Standard scores on the Applied Problems subtest were available for 2,437 PEELS children in Wave 1 and 2,711 in Wave 2.

Spanish Assessment: A Spanish version of the direct assessment was available for children who had limited comprehension of English as demonstrated by 1) answering fewer than five items correctly on the English version of the preLAS Simon Says and Art Show (combined) and 2) answering five or more items correctly on the Spanish preLAS Simón Dice and Muestra de Arte (combined). The Spanish assessment included subtests from preLAS 2000 Simón Dice and Muestra de Arte; Woodcock-Muñoz Letras y Palabras; Problemas Aplicados, and Conceptos Cuantitativos (Woodcock and Muñoz 1996); Leiter-R Attention Sustained Scale; IGDI: Picture Naming; and Test of Early Math Skills. Because of the small number of children completing the direct assessment in Spanish (25 in Wave 1 and 4 in Wave 2), Spanish direct assessment results are not presented in this report. Children who completed a direct assessment in Spanish were excluded from analyses of scores from the English direct assessment.

Alternate Assessment: For children who could not complete the direct assessment in English, the Adaptive Behavior Assessment System II (ABAS-II) was used as an alternate assessment. The ABAS-II is a checklist of the child's functional knowledge and skills and is completed by a teacher or other service provider. It assesses children's functional performance in several areas: communication, community use, functional (pre) academics, school living, health and safety, leisure, self-care, self-direction, social, and work. It also can be used to produce composite scores in conceptual, social, and practical domains. The scaled scores for each of the skill areas are based on a mean of 10 and a standard deviation of 3.

The ABAS-II has two versions. The first version, the Teacher/Daycare Provider Form, is for children not yet in kindergarten and measures the adaptive skills that have primary relevance for toddlers' and preschoolers' functioning in a daycare center, home daycare, or preschool. The second version, the Teacher Form, is for children in kindergarten or higher grades and measures the adaptive skills that have primary relevance for children's functioning in a school setting. In Wave 1, there were 338 PEELS children requiring an alternate assessment who had standard scores on the entire Teacher/Daycare Provider Form and 17 who had standard scores on the entire Teacher Form. In Wave 2, there were 152 who had scores on the Teacher/Daycare Provider Form, and 72 who had standard scores on the entire Teacher Form. These counts include children who took a direct assessment in Spanish and also had an alternate assessment completed for them.

Harrison and Oakland (2003) reported coefficient alpha reliabilities for the ABAS-II subtests on the Teacher/Daycare Provider Form ranging from .72 to .94, depending on the age group and subtest, with higher reliabilities for composite domain scores (r = .92 to .97). On the Teacher Form, they reported coefficient alphas ranging from .84 to .97, with composite domain coefficients in the .96 to .98 range. Test-retest reliabilities for periods of 2 days to 6 weeks ranged from .66 to .98, depending on age level and subtest. The correlation between the overall composite scores on the ABAS-II, Teacher/Daycare Provider Form, and Vineland Adaptive Behavior Scales, Classroom Edition was r = .75. The correlation between the ABAS-II Teacher Form overall composite and Vineland overall composites was r = .84 (Harrison and Oakland 2003).

Assessment Procedures

When a case was assigned to an assessor, the assessor received a scoring booklet that was specific to the child. A label on the cover indicated the child's first name, last initial, and date of birth. The scoring booklet included instructions for administering the assessments as well as a place for recording children's responses to each item for each subtest. The scoring booklet also included a place to record information from a screening interview the assessor conducted with the child's teacher, service provider, or parent. The screening interview was designed to prepare the assessor for the test session. It helped identify any needed test accommodations, whether the child could participate in the standard assessment or required an alternate assessment, and whether the child should be referred to a bilingual assessor. Before returning

the completed scoring booklet, assessors completed a child assessment summary, which captured contact information for the child's current teacher or service provider, whether the direct or alternate assessment was used, the date the assessment was completed, the location where it was completed, accommodations used, and the assessor's certification that he/she assessed the child and the scores were an accurate representation of the child's performance.

If an alternate assessment was required, the assessor gave the ABAS-II to the appropriate respondent (i.e., child's teacher or other service provider) and documented the reason for the alternate assessment in the child assessment summary. The assessor received $50, and the respondent completing the alternate assessment received $50.

Assessors were instructed to offer a variety of test accommodations so participating children could demonstrate what they know and what they can do. In order to assist with decisions regarding accommodations, the PEELS Assessors' Manual included 21 pages from the following document: *Making Assessment Accommodations: A Toolkit for Educators* (Council for Exceptional Children 2000). These pages contain references to accommodations in the *IDEA*, guiding principles for making assessment accommodations, a description of types of accommodations (e.g., scheduling, setting, presentation, and response), and questions and answers about making accommodations. As noted previously, assessors determined what test accommodations were needed for individual children based on information gathered during the Screening Interview.

The following accommodations were made available without prior approval from PEELS home-office staff:

- enlarged print,
- assessments given by someone familiar with the child,
- assessments given in the presence of someone familiar with the child,
- someone to help the child respond,
- specialized scheduling,
- adaptive furniture,
- special lighting,
- abacus,
- communication device, and
- multiple testing sessions.

The above accommodations are among those permitted on the Woodcock-Johnson III: Achievement Battery (McGrew and Woodcock 2001). Prior approval from PEELS home office staff was required for using sign language interpreters because of procedures established for their remuneration.

The number of children who received various accommodations in Wave 1 and Wave 2 are presented in table 10.

Table 10. Frequency of specific test accommodations used in Waves 1 and 2

	Wave 1 2003-04	Wave 2 2004-05
Abacus	‡	‡
Adapted furniture	19	12
Communication device	11	4
Enlarged print test easel	‡	‡
Familiar person administered assessment[a]	3	‡
Familiar person present during testing	186	82
Multiple test sessions	113	85
Person to help child respond[a]	40	8
Sign language interpreter	4	3
Other accommodation (e.g., parent present and quiet location)	21	18

‡ Reporting standards not met.

NOTE: These counts include children receiving accommodations on the Spanish assessment but not children in the alternate assessment group.

[a]As an accommodation, assessments were occasionally given by a service provider familiar with the child or a service provider helped the child respond to assessment items (e.g., clarified responses the assessor could not understand because of the child's articulation difficulties). In each case, the trained PEELS assessor was present, managed use of the accommodation, and scored the test.

In Wave 1, there were 350 children who had one or more accommodations, which is 14 percent of the children who completed the English or Spanish direct assessment and for whom accommodation data were available. Because children could receive more than one accommodation, the total number of accommodations received in Wave 1 was 399. In Wave 2, eight percent of children had one or more accommodations ($n = 214$). Appendix E provides information on the number of children who had various test accommodations by gender, race/ethnicity, cohort, and primary disability. With regard to having one or more test accommodations, there were no statistically significant differences in Wave 1 or Wave 2 by gender, race/ethnicity, or age cohort. There was a statistically significant difference across disability categories in the number of children receiving accommodations ($\chi = 346.678$, $p < .0001$).

Children who completed English direct assessments with accommodations (14% of the children with completed assessments in Wave 1 and 8% of those in Wave 2) were included in direct assessment analyses. Their scores were analyzed in the same way as scores for children who did not require accommodations.

Mail Questionnaires

The Elementary School Principal Questionnaire or Early Childhood Program Director Questionnaire was sent to principals or program directors, as appropriate, of the children's schools/programs. These questionnaires ask about school/program and community characteristics; student characteristics; staff, programs, and resources; special education programs and practices; and parent involvement. Only one Elementary School Principal Questionnaire or Early Childhood Program Director Questionnaire was sent to each school/program, regardless of the number of PEELS-participating children. The Wave 1 response rate was 72 percent, and the Wave 2 response rate was 77 percent.

Two versions of the teacher questionnaire were used in Wave 1, the Early Childhood Teacher Questionnaire (for children not yet in kindergarten) and the Kindergarten Teacher Questionnaire. An Elementary Teacher Questionnaire for children in grades 1 and higher was added in Wave 2. All three

teacher questionnaires ask about the specific child named on the inside cover and the child's experiences in the class or program. Questionnaire items address classroom staffing and materials, interaction with nondisabled peers, teachers' philosophies of early childhood education, and children's transitions in and out of their current programs. A pull-out section of the teacher questionnaires addresses the children's special education programs and related services. The questionnaires were completed by either the classroom teacher or the special education service provider, as appropriate. In Wave 1, a total of 2,018 Early Childhood Teacher Questionnaires and 269 Kindergarten Teacher Questionnaires were completed. In Wave 2, a total of 1,320 Early Childhood Teacher Questionnaires, 957 Kindergarten Teacher Questionnaires, and 314 Elementary Teacher Questionnaires were completed.

The teacher questionnaires also include the following teacher rating scales (indirect assessments): three subtests of the ABAS-II—Functional (Pre) Academics, Self-Care, and Self-Direction; the Vineland Adaptive Behavior Scales Classroom Edition, Motor Skills Domain; and the Preschool and Kindergarten Behavior Scales, Second Edition.

ABAS-II: Two forms from the ABAS-II were used. The Teacher/Daycare Provider (TDP) Form is designed for children ages 2 to 5, while the Teacher Form (TF) is designed for elementary students. The TDP Form measures the adaptive skills that are relevant for toddlers' and preschoolers' functioning in a daycare center, home daycare, or preschool. The TF measures the adaptive skills that are most relevant for children's functioning in a school setting. The TDP Form was used for non-kindergarteners and the TF for children in kindergarten or elementary school.

The Functional Pre-Academics subscale (for Cohorts A and B) measures basic pre-academic skills that form the foundations for reading, writing, mathematics, and other skills needed for daily, independent functioning. The TDP Form includes such items as "reads his/her own name," "prints at least two letters from his/her name," and "tells what day comes before another, for example Wednesday comes before Thursday." The Functional Academics subscale (for Cohort C) measures basic reading, writing, mathematics, and other academic skills needed for daily, independent functioning. The TF includes such items as "writes his/her own name," "reads and obeys common signs, for example, Do Not Enter, Exit, or Stop," and "states the days of the week in order." For each subscale, teachers rated the frequency with which the identified child exhibits specific behaviors, using a 4-point scale, labeled as: 0, *is not able*; 1, *never when needed*; 2, *sometimes when needed*; and 3, *always when needed*. Teachers also are provided a box to check whether their rating is based on an educated guess or estimate. The scaled scores presented in this report are based on a mean of 10 and a standard deviation of 3.

In Wave 1, there were 2,018 children who had scores on the Pre-Academics, Self-Care, and Self-Direction scales from the TDP Form. Another 269 had Functional Academics, Self-Care, and Self-Direction scores from the TF generated from responses to the Kindergarten Teacher Questionnaire.[16] In Wave 2, a total of 1,297 children had scores from the TDP Form, and 1,255 had scores from the TF. Note that two different teachers typically rated an individual child in Waves 1 and 2.

Vineland Adaptive Behavior Scales: The Fine Motor and Gross Motor subscales from the Vineland Adaptive Behavior Scales Classroom Edition provide a measure of each child's motor skills. Teachers are asked to rate the child's performance on a series of behaviors using a 3-point scale: 1, *usually*, 2, *sometimes or partially*, and 3, *never*. Teachers select their responses from one of two columns depending on whether they observe the child performing the behavior or if their ratings are based on an estimate. The children's scores on the two subscales are summed and converted into one standardized motor skills score. The standard scores are based on a distribution with a mean of 100 and a standard

[16] Information on the reliability of the ABAS-II was provided previously in the section on the alternate assessment.

deviation of 15. Norms for the Vineland Motor Skills domain were developed only for children ages 3 to 6 because motor development levels drop off after that age (Sparrow, Balla, and Cicchetti 1985).

The coefficient alpha reliability for the Motor Skills domain was .80 (Sparrow et al. 1985). Correlations between the Vineland Motor Skills standard scores and Adaptive Behavior Composite standard scores on the Kaufman Assessment Battery for Children (K-ABC) (Lambert and Windmiller 1981), ranged from .23 to .60 in a sample of children without disabilities, with the highest correlation for nonverbal skills (Mealor in Sparrow et al. 1985). Correlations were moderate between the Vineland Motor Skills standard scores and PPVT-Revised standard scores (r = .20; Sparrow et al. 1985). In all, 2,192 PEELS children received motor skills composite scores in Wave 1, and 2,540 received them in Wave 2.

PKBS-2: The PKBS-2, which was included in the Early Childhood Teacher and Kindergarten Teacher Questionnaires, is specifically designed to evaluate the social skills and problem behaviors of children ages 3 to 6. It is a norm-referenced, standardized instrument that includes two scales, a social skills scale (34 items) and a problem behavior scale (42 items). It comprises five subscales—Social Cooperation, Social Interaction, Social Independence, Externalizing Problems, and Internalizing Problems. Teachers are asked to rate how frequently the identified child exhibited a series of skills or behaviors during the previous 3 months. The measurement scale consists of four points, labeled *never*, *rarely*, *sometimes*, and *often*. The standard scores for the Social Cooperation, Social Interaction, and Social Independence subscales are summed to create the Social Skills composite score, and Externalizing and Internalizing Problems are summed to create the Problem Behaviors scale. The composite scores are then converted to composite standard scores. PKBS-2 standard scores are based on a distribution with a mean of 100 and a standard deviation of 15 (Merrell 2002). PEELS used the school-rater form of this measure.

Test developers reported Cronbach alpha coefficients of .96 to .97 for ages 3 to 6 on the Social Skills scale and .93 to .95 on the Problem Behavior scale. Three-week test-retest reliability for subscales of the Social Skills scale ranged from .58 to .66. For subscales of the Problem Behavior scale, test-retest reliability was in the .70 to .78 range.

Merrell (1995) reported correlations between the PKBS-2 Social Skills scale and the Social Skills Rating System (SSRS, Gresham and Elliott 1990) that ranged from .32 to .76 and correlations between the PKBS-2 Problem Behaviors scale and the SSRS that ranged from .25 to .83. Correlations between the Matson Evaluation of Social Skills with Youngsters (MESSY, Matson, Esvelt-Dawson, and Kazdin 1983) and the PKBS-2 Social Skills scale were moderate to strong (.62 to .85), while correlations between the PKBS-2 Problem Behaviors scale and MESSY were relatively weak to moderately strong (.22 to .72). The Conduct Problem section of the Connors Teacher Rating Scales (CTRS-39, Connors 1990) and the PKBS-2 Externalizing Problems subscale were highly correlated (.87), as were the CTRS-39 Emotional-Overindulgent scale and the PKBS-2 Internalizing Problems subscale (.78). The PKBS-2 Social Skills and School Social Behavior scales (Merrell 1993) were also highly correlated (.86). In Wave 1, 2,192 children had PKBS scale scores; in Wave 2, 2,540 had scores.

Overall, 79 percent of children's teachers in Wave 1 returned questionnaires, and 84 percent of children's teachers in Wave 2 returned questionnaires. One teacher questionnaire will be sent to the teachers of participating children in Waves 3 and 4.

Table 11. Total unweighted number of respondents and response rate for each PEELS instrument

	Wave 1		Wave 2	
Instrument type	Frequency	Response rate	Frequency	Response rate
Parent interview	2,802	96%	2,893	93%
LEA questionnaire	207	84%	†	†
SEA questionnaire	51	100%	†	†
Principal/program director questionnaire	852	72%	665	77%
Teacher mail questionnaire	2,287	79%	2,591	84%
Early childhood teacher questionnaire	2,018	79%	1,320	86%
Kindergarten teacher questionnaire	269	73%	957	79%
Elementary teacher questionnaire	†	†	314	86%
Child assessment	2,794	96%	2,932	94%
English/Spanish direct assessment	2,463	97%	2,704	96%
Alternate assessment only	331	93%	228	79%

NOTE: Wave 1 frequencies do not include cases in the supplemental sample for which data were imputed.
† Not applicable

Data Preparation and Analysis

In data preparation, imputation was conducted for selected items on the child assessment, teacher questionnaire, and parent interview data. In general, the item missing rate was fairly low. For the Wave 1 parent interview, the item missing rates for the augmented sample were less than 10 percent for 702 variables and 10 to 14 percent for 34 variables; three variables had rates between 15 and 17 percent. For Wave 2, there were no missing values for 235 parent interview variables and a missing rate of less than 9 percent for 265 variables; only two variables had higher rates, 15 percent for one and 24 percent for another. For teacher questionnaire data in Wave 1, item missing rates were under 10 percent for 94 percent of variables, 10 to 15 percent for 4 percent of variables, and 15 to 22 percent for 2 percent of variables. In Wave 2, fewer than 5 percent of cases were missing for 99 percent of teacher questionnaire variables, and 5 to 10 percent of cases were missing for 1 percent of variables. For the Wave 1 assessment data, less than 15 percent of the cases were missing for less than 80 percent of the variables. In Wave 2, less than 2 percent of the cases were missing for 95 percent of the variables, and 2 to 3 percent of the cases were missing for 5 percent of the variables. The item missing rate prior to imputation was higher in Wave 1 because data for the supplemental sample were missing.

Imputed values may have two undesirable features. The first is that they may cause bias in an estimate calculated from the post-imputed data. The second is that the variance of such estimates may increase. If the imputed values are treated as real values and an ordinary variance estimator is used, this increased variance is not reflected and the variance is underestimated, which can lead to an erroneous inference. These potential problems become more serious if the percentage of imputed cases in the analysis sample is high (for example, over 20 percent). However, the percentage of imputation for the supplemental sample was between 6.6 and 8.7 percent of the augmented sample, depending on the instrument. Therefore, the risk of imputation-related bias was judged to be minimal. The variance inflation due to imputation was also contained because the imputation rate was below 10 percent. Imputation for the supplemental sample increased the amount of data usable for analysis, offsetting the potential risk of bias.

Researchers used different methods of imputation depending on the nature of missing and available information for imputation. The methods included hot-deck imputation, regression, external data source, and deterministic or derivation method, based on the internal consistency principle of inter-related variables. In some cases, a postulated value was imputed after analyzing missing patterns. Whenever a value of a variable was imputed, an imputation flag for the variable was created in the data set to record the change.

The data presented in the report have been weighted to generate national estimates. Different weights have been used depending on the sources of data. These weights adjust the child base weights given to the 3,104 recruited families to account for nonresponse on specific data collections in specific waves or groups of waves. Appendix B includes complete information on the weights available at the end of the second wave of data collection.

It is extremely difficult to obtain an unbiased variance estimator for a complex sample like the one used in PEELS. The jackknife variance estimator was used; it takes account of clustering effects and other weighting adjustments for nonresponse and post-stratification. The variance estimator is usually slightly conservative and tends to lead to a slightly smaller chance of type I error than indicated by the significance level of the test. PEELS researchers performed post-stratification whenever possible to enhance the precision of the survey estimates.

All standard errors and significance tests were conducted using WesVar Version 4.2 (Westat 2002) to account for the complex probability sampling and weighting used in PEELS. Two independent variables used throughout the report require some description. Parents provided information on the children's race/ethnicity. Because of the small number of American Indian or Alaska Native and Asian children in the study, data for those subgroups were considered unreliable and were not included in the analyses of race/ethnicity. Children of all races/ethnicities were included in the remainder of the analyses. A three-group race/ethnicity variable was used with the following definitions:

- Hispanic—children who were Hispanic and of any race;
- Black—children who were Black or African American only and not Hispanic; and
- White—children who were White only and not Hispanic.

The disability categories used in data collection are those specified in *IDEA*. Children's primary disability category in Wave 1 and Wave 2 was obtained from their teachers or service providers; however, if service provider data were missing, disability information was obtained from the children's parents or enrollment form. Because of the small sample sizes for some disability categories, a "low-incidence" category was created that included deaf/blindness, deafness, hearing impairment, traumatic brain injury, visual impairment, and other disabilities identified by parents (e.g., comprehension problems; hand-eye coordination). A complete list of analysis variables is provided in appendix G.

In all data displays, if the number of cases in a cell dropped below 3, data were suppressed, and a footnote was added to indicate that "Reporting standards were not met." This convention was used to maintain data confidentiality. If the coefficient of variation (CV) for an estimate was more than .20, a footnote was added that indicated that the estimate is unstable and the reader should use caution in interpreting the data. The CV is the ratio of the standard error to the mean or percentage. It is designed to convey the variability of the measure, independent of the scale. Analysis results that are unreliable estimates are marked with ! in data displays. These unstable estimates are discussed in this report, however, when statistically significant differences are found, despite the large standard errors.

Chi-squares and *t* tests for dependent samples were performed to examine statistically significant differences across subgroups and over time within subgroups. The *t* tests for dependent samples took into account the correlations between the Wave 1 and Wave 2 samples. The sample sizes were too small to use repeated measures ANOVAs by disability group. Individual tests were performed using a 5 percent significance level. Sometimes, however, related *t* test results were discussed as a group or family of tests. In those cases, researchers controlled the family-wise error rate to avoid making false positive claims. The Benjamini-Hochberg procedure (Benjamini and Hochberg 1995), known to be less conservative than Bonferroni correction, was used for multiple testing situations. This procedure controls the false discovery rate (FDR) at a set level instead of the family-wise error rate.[17]

[17] False discovery rate is the expected error rate of making false positives among all positive claims. If this is set to be small, say 0.05, the Benjamini-Hochberg procedure controls FDR at 0.05.

Chapter 3: Changes in Eligibility and Classification Status of Children Who Received Preschool Special Education Services

Early data from PEELS suggest that preschoolers identified as having disabilities differed from preschoolers overall in a number of ways. They were disproportionately male (70 percent versus 30 percent female), were more likely than their peers without disabilities to live in low-income households, and had a greater likelihood of being born 3 or more weeks premature. In other ways, they were much like other children ages 3 through 5; they had roughly the same racial/ethnic mix, similar levels of parental education, and similar household compositions (Markowitz et al. 2006).

This chapter moves beyond cross-sectional analyses of preschoolers with disabilities to describe changes in children's disability status. Many children with moderate to severe disabilities will receive special education services throughout their school years, yet other youth may receive services for a period of time and subsequently return to the general program or become "declassified," and others will remain in special education but under a different classification category, or become "reclassified."

Declassification From Special Education

Leaving special education, or declassification, is a process that varies from one district to another across the United States and is intended to be individualized and based on child need. There are no Federal criteria that prescribe how or when a child should be declassified. Regulation 300.534 (c)(1) of *IDEA 2004* simply states that "a public agency must evaluate a child with a disability in accordance with Secs. 300.532 and 300.533 before determining that the child is no longer a child with a disability" (*IDEA* Amendments of 2004, P.L. 108-446). Sections 300.532 and 300.533 outline evaluation procedures and determination of needed evaluation data.

As noted above, the determination that a child will no longer be eligible for special education is a process that varies across districts. School districts may develop their own guidelines, but no consistent regulations or transition plans exist for children who are no longer in need of special education services, and some research suggests that as many as 50 percent of programs do not have written specifications for exit criteria (Thurlow, Lehr, and Ysseldyke 1987; Thurlow, Ysseldyke, and Weiss 1988). While declassification, as reported by respondents in this study, may be interpreted as a result of a child having improved in skills or performance, a child may also leave special education at the parent's request and against the recommendation of a school district. In this respect, the process of declassification is very similar to the process of determining eligibility for special education: both are subject to parental consent and vary across schools, districts, and states in the United States. A child who is eligible *or* ineligible in one school, district, or state may have a different eligibility status in another.

Among school-aged children, estimated rates of declassification range from 6 percent over 4 years for secondary-aged youth (Carlson 1997) to 17 percent over 2 years for children in elementary school (Walker et al. 1988). Past estimates of declassification among preschoolers range from 7 percent (VESID 2003) to 50 percent (Innocenti 2005). The high end of this range could reflect actual differences in declassification practices among younger children, such as children receiving services for only a short time during their preschool years. Different methodologies, including the time at which the estimates were taken, may also account for some variation. For example, if declassification is measured only at exit from preschool, rates may be higher, since this is a time when IEPs and eligibility are often reevaluated.

In PEELS, all children had an active IEP or IFSP at the time they were recruited into the study. Since then, some children have been declassified. In this report, declassification status is based primarily

on teacher report of whether the participant had either an IEP or IFSP, a question that appears in each form of the teacher questionnaire. For children with missing teacher data in 2003-04, teacher response to a question about whether the child received services during the previous year was used to determine declassification status. Parent response was used to fill in any remaining missing data. Seventeen percent of children were missing data for declassification status based on teacher report alone and had their declassification status imputed as a result.

The Wave 1 and Wave 2 PEELS data were used to examine several issues related to declassification for young children, including differences in demographic, school, and district characteristics; transition status; and academic abilities of children who were declassified. The augmented sample was used for all analyses in this chapter. Also, all declassification and reclassification variables used for analyses in this chapter were developed from longitudinal data. Table 12 presents data on IEP/IFSP status during both the 2003-04 and 2004-05 school years; the remainder of the section focuses on changes in IEP/IFSP status between the first and second waves of data collection. As such, in tables 13 to 16, we present data only for those children who had an IEP or IFSP during the 2003-04 data collection (later winter/early spring) and were no longer eligible by the 2004-05 data collection at the same time of year.

Declassification and Demographics

In some earlier studies, demographic characteristics, such as race, have been associated with changes in eligibility status. For example, for children with speech or language impairments, declassification was less likely for Black children than for those of other races (Walker et al. 1988). Across age groups, children with speech or language impairments and learning disabilities were the most likely to leave special education (Carlson 1997; Carlson and Parshall 1996; Halgren and Clarizio 1993; Innocenti 2005; Walker et al. 1988; Wong 1997). Family characteristics, including income, may also relate to eligibility decisions. For example, Carlson (1997) observed a positive association between income and declassification, and Markowitz and Cline (1991) documented a positive association between income and the frequency of checks on progress (Markowitz and Cline 1991). In this section, we examine the relationship between several demographic and family characteristics and declassification status.

As noted above, all children were eligible for special education services at the time of recruitment, but by the time of the 2003-04 data collection, some of these children no longer had an IEP or IFSP. Table 12 presents the percentage of children who did and did not have an IEP/IFSP at the Wave 1 and Wave 2 data collection points. Nearly 80 percent of children (79%, *S.E.* = 1.1) had an IEP/IFSP at both data collection points. Fourteen percent (*S.E.* = 1.0) of children had an IEP/IFSP in 2003-04 but were declassified by 2004-05; two percent (*S.E.* = 0.4) did not have an IEP/IFSP in 2003-04 but had one in 2004-05; and 5 percent (*S.E.* = 0.7) did not have an IEP/IFSP at either point in time because they were declassified before the Wave 1 data collection. The relationship between demographic and family characteristics and declassification status was examined, and there were no statistically significant differences in declassification by gender ($\chi = 0.639$, $p = .424$), race/ethnicity ($\chi = 3.881$, $p = .099$), or by family income ($\chi = 0.173$, $p = .871$). Table 13 provides the percentage of children by gender, race/ethnicity, and family income who received preschool special education services during the 2003-04 school year and were declassified by 2004-05.

Table 12. Percentage of young children who received preschool special education services and had or did not have an IEP/IFSP in 2003-04 and 2004-05

		2004-05		
		Total	IEP/IFSP	No IEP/IFSP
	Total	100	80.8	19.4
2003-04	IEP/IFSP	92.8	78.5	14.3
	No IEP/IFSP	7.4	2.3[1]	5.1

! Interpret data with caution.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," "Early Childhood Teacher Questionnaire," and "Parent Interview."

Table 13. Percentage of young children who received preschool special education services during the 2003-04 school year and were declassified in 2004-05

	Declassified
Expected percentage based on the total sample	15.4[18]
Gender	
Male	14.8
Female	16.7
Race/Ethnicity	
Black	9.2
Hispanic	13.9
White	15.9
Family income at Wave 1	
$20,000 or less	15.0
$20,001-$40,000	16.0
More than $40,000	14.7
Metropolitan status	
Urban	12.9
Suburban	14.7
Rural	20.9

[18] Analyses in tables 13, 14, and 15 excluded 7.4 percent of the original sample children who had no IEP at the time of initial data collection in 2003-04 thus making the overall sample declassification rate to be 15.4 percent.

Table 13. Percentage of young children who received preschool special education services during the 2003-04 school year and were declassified in 2004-05 (continued)

	Declassified
District size	
Very large	12.6
Large	14.6
Medium	11.9
Small	21.5
District wealth	
High	14.9
Medium	15.5
Low	16.1
Very Low	15.0

NOTE: Detail may not sum to totals because of rounding. The chi-square analysis result was significant at the $p < .05$ level for metropolitan status and district size.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," "Early Childhood Teacher Questionnaire," and "Parent Interview."

The relationship between primary disability category and declassification status was examined, and there was a statistically significant difference ($\chi = 29.286$, $p < .001$). As shown in table 14, a total of 21 percent of children in two groups were declassified between 2003-04 and 2004-05: those with a speech or language impairment (*S.E.* = 5.3) and those with an emotional disturbance (*S.E.* = 0.9). The percentage of children with developmental delays, learning disabilities, orthopedic impairments, and low-incidence disabilities who were declassified ranged from 2 percent for autism (*S.E.* = 0.4) to 13 percent for developmental delay and learning disability (*S.E.* = 3.7 and 0.6, respectively) (see table 14).

Table 14. Percentage of young children who received preschool special education services during the 2003-04 school year and were declassified in 2004-05: by disability

	Primary disability at Wave 1									
	Expected percentage based on the total sample	AU	DD	ED	LD	MR	OI	OHI	SLI	LI
Percentage of children in each disability group who were declassified	15.3	1.8!	13.1!	20.7	13.1	‡	11.3	‡	21.2!	8.0!

! Interpret data with caution.
‡ Reporting standards not met.
NOTE: The chi-square analysis result was significant at the $p < .05$ level.
AU = Autism; DD = Developmental delay; ED = Emotional disturbance; LD = Learning disability; MR = Mental retardation; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; LI = Low incidence.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," "Early Childhood Teacher Questionnaire," and "Parent Interview."

Declassification and School/District Characteristics

Previous research suggests that the characteristics of a child's district or school may play a role in determining whether a child is declassified from special education. A New York study found that suburban programs had the highest percentage of preschoolers declassified, and New York City had the lowest percentage (VESID 2003). A comparison in North Carolina found that preschoolers from an urban county were more likely to leave special education than those from a rural county (Wong 1997), and the reason for this remains unclear. Halgren and Clarizio (1993) found no association between declassification rates and district size.

In PEELS, researchers examined declassification status by three characteristics of the child's school district: metropolitan status, district size, and district wealth. Statistically significant differences were observed within two of these characteristics, metropolitan status and district size; there were no statistically significant differences in declassification by district wealth. Urban, suburban, and rural districts differed significantly in the frequency of declassification among young children ($\chi = 6.915$, $p = .029$). The findings were similar to those in the New York study (VESID 2003) in that both studies found significant results by metropolitan status, although with somewhat different patterns. As shown in table 13, the percentages of children declassified in PEELS were 13 percent (*S.E.* = 2.1), 15 percent (*S.E.* = 1.3), and 21 percent (*S.E.* = 2.1) for urban, suburban, and rural districts, respectively.

Declassification also differed significantly by district size ($\chi = 14.512$, $p = .002$). As shown in table 13, a total of 13 percent of children from very large districts (SE = 2.2), 15 percent from large districts (*S.E.* = 2.2), 12 percent from medium districts (*S.E.* = 2.2), and 22 percent (*S.E.* = 1.6) from small districts were declassified.

Declassification and Transitions

Past research suggests that changes in eligibility status are more likely during transition periods (Walker et al. 1988). To explore whether this was the case in PEELS, children were characterized as either remaining in preschool, transitioning from preschool to kindergarten, or transitioning from kindergarten to first grade between 2003-04 and 2004-05. Some children in the sample remained out of school throughout this time period A small group of children transitioned from being in school to out of school, were in an ungraded program, or were missing information on their grade level. This group of children, the "other status" group, was included in the analysis; it constitutes 3.6 percent of the analysis sample.

Statistically significant differences were observed for transition status by declassification status ($\chi = 37.333$, $p < .001$). Table 15 provides the percentages of children declassified, by transition status (n = 2,553). Thirty-one percent (*S.E.* = 8.7) of children who were out of school at both data collection points were declassified from special education compared with 24 percent (*S.E.* = 4.3) of those transitioning from kindergarten to first grade, and 20 percent (*S.E.* = 1.4) of those transitioning from preschool to kindergarten. Only 6 percent (*S.E.* = 1.1) of children who remained in preschool were declassified.

Another way to consider the interaction between declassification and transitions is to examine declassification status by whether the child underwent any type of transition between 2003-04 and 2004-05, e.g., from preschool to kindergarten, from kindergarten to first grade, or from being out of school to entering school. When viewed in this way, the relationship between transition status and declassification was significant ($p < .001$). In percentage terms, 21 percent of those making some type of transition were declassified (*S.E.* = 1.0), whereas only 7 percent of non-transitioning children were declassified (*S.E.* = 1.4).

Table 15. Percentage of young children who received preschool special education services during the 2003-04 school year and were declassified in 2004-05, by transition status

	Transition status					
	Expected percentage based on the total sample	Remained in preschool	Transitioned from preschool to kindergarten	Transitioned from kindergarten to first grade	Remained out of school	Other status
Percentage of children in each transition status who were declassified	15.3	6.4	19.7	23.8	30.9!	35.7!

! Interpret data with caution.
NOTE: The chi-square analysis result was significant at the $p < .05$ level.
Other status includes children who transitioned from being in school to out of school, children who were in an ungraded program, and children who were missing information on their grade level.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," "Early Childhood Teacher Questionnaire," and "Parent Interview."

Declassification, Standardized Assessments, and Academic Ability

PEELS data were used to compare children who were declassified between the 2003-04 and 2004-05 school years and those who remained eligible for services on three assessments administered at Wave 2. On the Woodcock-Johnson III: Letter-Word Identification subtest, children who were declassified (M = 104.5, $S.E.$ = 1.0) scored significantly higher than those who remained eligible for services (M = 99.2, $S.E.$ = 0.6) *(*t = -4.474, $p < .001$). The difference between these groups of children was greater on the Applied Problems subtest. Declassified children had a mean score of 101.3 ($S.E.$ = 1.2); children who retained eligibility had a mean score of 90.1 ($S.E.$ = 0.7) (t = -9.612, $p < .001$). A statistically significant difference between groups was also observed on the PPVT, with means of 96.1 ($S.E.$ = 1.3) and 87.6 ($S.E.$ = 0.6) for declassified and non-declassified children, respectively (t = -6.405, $p < .001$) (see table 16).

Table 16. Mean performance of young children who received preschool special education services during the 2003-04 school year on tests of emerging literacy and early math skills, by eligibility status

	Letter-Word Identification*	Applied Problems*	PPVT*
Total	100.1	92.0	89.0
Remained eligible	99.2	90.1	87.6
Declassified	104.5	101.3	96.1

*t, $p < .05$
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Letter-Word Identification," "Woodcock-Johnson III: Applied Problems," "Peabody Picture Vocabulary Test III–R."

Changes in Disability Categories

In addition to being declassified from special education, children may also be reclassified, that is, moved from one primary disability category to another. Previous research from two studies suggests that 5 to 6 percent of students with disabilities are reclassified each year, one estimate based on a sample of children in grades K-6 (Walker et al. 1988) and the other based on data for children in preschool through secondary school (Halgren and Clarizio 1993). In one of the studies, Halgren and Clarizio (1993) reported that reclassification occurred at the greatest rate in preschool (25%). Earlier studies also suggest that students with physical/multiple disabilities, speech or language impairments, or emotional disturbance were more likely to be reclassified (Halgren and Clarizio 1993; Walker et al. 1988). In their study, Halgren and Clarizio (1993) found movement within disability categories; for example, children with mild mental retardation were often reclassified as having severe mental retardation (39% of students in the sample). They found that students' cognitive abilities were also related to declassification and reclassification; students with mild cognitive disability were prone to declassification, whereas students with more significant cognitive disability were prone to reclassification.

In the year that passed between the first PEELS data collection and the second, 23 percent (N = 546) of children who remained in special education changed primary disability categories, and 77 percent stayed in the same category (*S.E.* = 1.7 and 2.0, respectively). Reclassification rates did not vary by gender ($\chi = 0.143$, $p = .706$) or race/ethnicity ($\chi = 0.845$, $p = .626$). Twenty-three percent of males (*S.E.* = 1.8) and 22 percent of females (*S.E.* = 2.0) were reclassified; 20 percent of Blacks (*S.E.* = 2.6), 22 percent of Hispanics (*S.E.* = 2.4) and 23 percent of Whites (*S.E.* = 1.7) were reclassified.

PEELS data were used to compare children who were reclassified between the 2003-04 and 2004-05 school years and those who were not reclassified on three assessments administered at Wave 2. Mean scores on the Woodcock-Johnson III: Letter-Word Identification subtest (M = 96.8, *S.E.* = 1.5; M = 99.0, *S.E.* = 0.8) (t = 1.429, p = .158) did not differ significantly by reclassification status. However, mean scores on the Woodcock-Johnson III: Applied Problems subtest (M = 82.9, *S.E.* = 1.7 for reclassified; M = 90.0, *S.E.* = 1.1 for not reclassified) (t = 3.881, $p < .001$) and on the PPVT (M = 82.2, *S.E.* = 1.4 for reclassified; M = 89.1, *S.E.* = 0.7 for not reclassified) (t = 4.69, $p < .001$), were significantly different between children who were and were not reclassified (see table 17). On both measures, reclassified children were lower performing than those whose disability classification was unchanged, in part supporting findings by Halgren and Clarizio (1993).

Table 17. Mean performance of young children who received preschool special education services during the 2003-04 and 2004-05 school years on tests of emerging literacy and early math skills, by reclassification status

	Letter-Word Identification	Applied Problems*	PPVT*
Total	98.4	88.2	87.3
Reclassified	96.8	82.9	82.2
Not reclassified	99.0	90.0	89.1

* *t, $p < .05$

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Letter-Word Identification," "Woodcock-Johnson III: Applied Problems," "Peabody Picture Vocabulary Test III–R."

Table 18 shows the percentage of children in each of nine disability groups who received preschool special education services and whose disability label was stable, that is, it remained the same from 2003-04 to 2004-05. The overall percentage of children retaining the same disability label from one year to the next was 77.4 (*S.E.* = 1.2) and ranged from 57 percent for children with other health impairments (*S.E.* = 10.2) to 89 percent for children with autism (*S.E.* = 2.9).

Table 18. Percentage of young children who received preschool special education services whose disability classification remained the same from 2003-04 to 2004-05

Disability classification	Percent
Total	77.4
Autism	89.4
Developmental delay	64.1
Emotional disturbance	60.4
Learning disability	69.7
Mental retardation	71.4
Orthopedic impairment	66.7
Other health impairment	57.2
Speech or language impairment	88.3
Low-incidence disability	61.6

NOTE: Percentages do not include children who were declassified between 2003-04 and 2004-05.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Early Childhood Teacher Questionnaire," "Kindergarten Teacher Questionnaire," "Elementary School Teacher Questionnaire," "Parent Interview."

The sample sizes for children with developmental delay and speech or language impairment were large enough to permit more in-depth analysis of reclassification. By age 10 (or the age at which states and districts stop using the developmental delay category), children identified as having a developmental delay must either be reclassified into another disability group or declassified (see *IDEA* 2004, Section 1401). In 2003-04, 28 percent of all preschoolers with disabilities were identified as having a developmental delay as their primary disability (Markowitz et al. 2006).[19] From 2003-04 to 2004-05, 64 percent of children initially identified as having a developmental delay retained that label (*S.E.* = 3.2). Fourteen percent (*S.E.* = 2.0) moved from the developmental delay to the speech or language impairment category; and 4 percent (*S.E.* = 0.7) moved to the learning disability category.[20]

Some children also moved into the developmental delay category from other disability groups. For example, 13 percent of children identified as having an emotional disturbance (*S.E.* = 6.3), 9 percent of children identified as having an orthopedic impairment or other health impairment (*S.E.* = 6.4 and 6.6, respectively), and 10 percent of children identified as having a low-incidence disability (deafness, deaf-blindness, vision impairment, or traumatic brain injury) (*S.E.* = 2.5) in 2003-04 were identified as having a developmental delay in 2004-05. These shifts involve relatively few children, however, since only 1 to 3 percent of children ages 3 through 5 with disabilities were identified as having emotional disturbance, orthopedic impairment, or other health impairment in 2003-04; 6 percent were identified as having a low-incidence disability.

[19] Developmental delay is an optional Federal disability category for children from birth through age 9 (or a subset of that age group) used by 44 states in 2003 (Danaher, Kraus, Armijo, and Hipps 2003).

[20] Because of small sample sizes, estimates for the remaining disability categories were imprecise.

In PEELS, of children identified as having a speech or language impairment in 2003-04, 21 percent were declassified by 2004-05, which was one of the higher declassification rates, and 79 percent continued to receive services (*S.E.* = 1.7). Of those who were still receiving special education services in 2004-05, 88 percent retained the speech or language impairment label, and 12 percent were reclassified (*S.E.* = 1.3). Six percent were reclassified from speech or language impairment to developmental delay (*S.E.* = 1.2) and 6 percent to various other disability categories.

Summary

In PEELS, between 2003-04 and 2004-05, 15.4 percent of children were declassified. Declassification was significantly associated with primary disability category, district metropolitan status, district size, and transition status. Children who were declassified scored significantly higher than those who remained eligible for special education services on several tests of early academic skills.

Between the first and second wave of data collection, 23 percent of children who continued to receive special education services had a change in their primary disability category. Children who were reclassified scored lower on measures of emerging literacy and math skills than children who were not reclassified.

Chapter 4: Changes in Services for Preschoolers With Disabilities

Special education services for preschool-aged children may include a wide array of instructional and related services. In PEELS, teachers were asked to identify all services provided to the participating children through the school system. A list of 23 services was included as well as a place to enter other services[21]. Teachers indicated that nearly 90 percent (89%, *S.E.* = 1.5) of the children received speech or language therapy in the 2003-04 school year, and 86 percent (*S.E.* = 1.6) did so in 2004-05, making it the most common service both years. Between one-fifth and one-third of children received occupational therapy (32%, *S.E.* = 1.6 in 2003-04 and 35%, *S.E.* = 1.5 in 2004-05) and learning strategies/study skills assistance by a special educator (30%, *S.E.* = 2.0 in 2003-04 and 20%, *S.E.* = 1.0 in 2004-05).

In this chapter, the augmented sample was used for all analyses. Also, all analyses in this chapter were longitudinal. T-tests for dependent samples were used to assess statistically significant changes in services over 2 years within subgroups. No analyses were conducted to compare differences across the subgroups. To control the family-wise error rate and avoid making false positive claims, the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995) was used for all multiple testing situations.

On average, preschoolers with disabilities in PEELS were nearly 3 years old when they started receiving special education or therapy services from a professional (Markowitz et al. 2006). As children get older, one might expect their services to change, either to reflect the emerging demands of an elementary-school setting or to reflect children's evolving educational and related service needs. From 2003-04 to 2004-05, teachers reported a decrease in the percentage of children receiving many services. It should be noted that although the outcome variable is categorical, with large sample sizes, changes in proportion can be assumed to have a normal *t* distribution, so a *t* test is appropriate.

There was a statistically significant reduction in the number of children receiving nine services (see table 19). For example, service coordination/case management decreased from 25 percent (*S.E.* = 2.4) in 2003-04 to 9 percent (*S.E.* = 1.0) in 2004-05; training, counseling, or other supports/services for the children's family decreased from 16 percent (*S.E.* = 1.5) in 2003-04 to 5 percent (*S.E.* = 0.6) in 2004-05; and learning strategies/study skills assistance decreased from 30 percent (*S.E.* = 2.0) in 2003-04 to 20 percent (*S.E.* = 1.0) in 2004-05. The only service showing a significant increase from 2003-04 to 2004-05 was help from a one-to-one paraeducator or assistant, which increased from 10 percent (*S.E.* = 0.8) to 13 percent (*S.E.* = 1.1).

In 2003-04, young children with disabilities, on average, received 3.5 different special education and related services (*S.E.* = 0.1); in 2004-05, that figure was 2.8 (*S.E.* = 0.1) (see table 20). From 2003-04 to 2004-05, the mean number of services received by children decreased significantly overall and for children in each cohort, for males and females, and for all disability categories except other health impairment.

[21] This list did not include indirect services, such as teacher consultation.

Table 19. The percentage of young children who received preschool special education services that received specific services through their school system, by school year: School years 2003-04 and 2004-05

	2003-04	2004-05	*t* value	*p* value
Adaptive physical education	9.6	11.0	1.261	0.212
Assistive technology services/devices	10.1	8.5	-0.91	0.366
Audiology*	9.7	4.2!	-2.76	0.008
Augmentative or alternative communication system*	10.0	6.5	-3.198	0.002
Behavior management program	14.4	11.9	-1.948	0.056
Learning strategies/study skills assistance*	29.5	20.4	-4.701	0
Occupational therapy*	31.9	35.4	2.215	0.03
One-to-one paraeducator/assistant*	9.8	13.0	2.548	0.013
Physical therapy	17.6	17.2	-0.377	0.707
Service coordination/case management*	25.4	8.6	-7.872	0
Social work services*	8.7	4.9	-3.377	0.001
Special transportation because of disability*	19.0	13.2	-4.384	0
Specialized computer software or hardware	6.4	5.0	-1.195	0.237
Speech or language therapy	88.6	86.4	-1.146	0.256
Training, counseling, or other supports/services for family*	16.4	4.5	-7.013	0
Tutoring/remediation by a special education teacher*	16.8	10.7	-4.013	0
Other services	17.0	13.6	-2.002	0.05

! Interpret data with caution.
* t, $p < .05$.
NOTES: Other services include health services; instruction in American Sign Language, Manual English, Cued Speech, or Braille; mental health services; reader or interpreter; vision services; and other services specified by the respondent. Denominators do not include children who were declassified from special education, so percentages include only children with an IEP.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," and "Early Childhood Teacher Questionnaire."

Time Spent in Different Educational Settings

Some young children with disabilities receive relatively low-intensity services, for example, an hour per week of speech therapy. Others receive services for an entire school day. They may include all-day public or private preschool, kindergarten, or elementary school programs.

Teachers in PEELS were asked to indicate the amount of time per week the children spent in a variety of settings, including a regular education classroom, special education setting (e.g., resource room, self-contained class, or separate school for children with disabilities), therapy setting, non-special education setting outside the classroom for remedial or special assistance (e.g., resource room for Title I or English-as-a-second-language instruction), and home instruction. From 2003-04 to 2004-05, the mean hours per week spent in a regular education classroom increased significantly, from 8.2 hours (*S.E.* = 0.5) to 15.0 hours (*S.E.* = 0.5) (t = 15.535, $p < .001$). Time in special education settings decreased significantly, from 8.0 hours a week in 2003-04 (*S.E.* = 0.5) to 6.2 hours a week in 2004-05 (*S.E.* = 0.4) (t = -4.315, $p < .001$). The amount of time children spent in a therapy setting, non-special education

Table 20. Mean number of services provided to young children who received preschool special education services, by age cohort, gender, and disability: School years 2003-04 and 2004-05

	2003-04	2004-05	*t* value	*p* value
Total	3.5	2.8	-6.03	0
Age cohort				
Cohort A*	3.6	2.7	-6.74	0
Cohort B*	3.6	2.7	-5.65	0
Cohort C*	3.3	2.8	-2.72	0.008
Gender				
Male*	3.5	2.8	-6.28	0
Female*	3.5	2.7	-3.77	0
Primary disability at Wave 1				
Autism*	5.5	4.1	-2.93	0.005
Developmental delay*	4.2	3.2	-5.48	0
Emotional disturbance*	4.4	3.1	-2.79	0.007
Learning disability*	3.6	2.4	-4.69	0
Mental retardation*	5.9	4.4	-3.62	0.001
Orthopedic impairment*	5.8	4.0	-2.06	0.044
Other health impairment	4.9	3.9	-1.43	0.158
Speech or language impairment*	2.2	1.8	-4.33	0
Low-incidence disability*	5.7	4.4	-2.15	0.036

*t, $p < .05$

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," and "Early Childhood Teacher Questionnaire."

setting outside the classroom, and home instruction remained fairly constant from 2003-04 to 2004-05 (see table 21).

Parent Satisfaction With Special Education Services

Parents were asked how satisfied they were with special education services their children received. The number of parents who reported being *very satisfied* decreased significantly ($t = -2.382$, $p = .02$). Forty-seven percent of parents reported being *very satisfied* in 2003-04 compared to 42 percent in 2004-05 (*S.E.* = 1.9 and 2.1, respectively) (see table 22). The number of parents who reported being *satisfied*, *dissatisfied*, and *very dissatisfied* did not change significantly. Forty-six percent of parents reported being *satisfied* in 2003-04 and 48 percent *satisfied* in 2004-05 (*S.E.* = 1.7 and 1.7, respectively) ($t = 1.314$, $p = .194$); 7 percent of parents reported being *dissatisfied* in both years (*S.E.* = 0.7 and 0.8, respectively) ($t = 1.013$, $p = .315$); and 2 percent of parents reported being *very dissatisfied* in both years (*S.E.* = 0.3 and 0.3, respectively) ($t = 1.391$, $p = .419$). So despite the other changes in educational services from 2003-04 to 2004-05, including time in various settings and types and number of services provided, parent satisfaction remained relatively stable, with a statistically significant decline only in the percentage of parents who were very satisfied.

Table 21. Mean hours per week that young children who received preschool special education services spent in various educational settings, by age cohort and school year: School years 2003-04 and 2004-05

	2003-04	2004-05
Total		
Regular education classroom*	8.2	15.0
Special education setting*	8.0	6.2
Therapy setting	0.8	0.8
Non-special education setting outside the classroom for remedial or special assistance	0.2!	0.2
Home instruction	0.2	0.2
Cohort A		
Regular education classroom*	4.3	7.1
Special education setting	8.7	9.0
Therapy setting*	0.6	0.8
Non-special education setting outside the classroom for remedial or special assistance	0.1	0.1
Home instruction	0.2!	0.1
Cohort B		
Regular education classroom*	6.7	13.5
Special education setting*	8.4	6.6
Therapy setting	0.7	0.8
Non-special education setting outside the classroom for remedial or special assistance	0.1	0.1
Home instruction	0.2!	0.1!
Cohort C		
Regular education classroom*	11.4	20.5
Special education setting*	7.2	4.4
Therapy setting	0.9!	0.8
Non-special education setting outside the classroom for remedial or special assistance	0.3!	0.4!
Home instruction	0.1!	0.2!

! Interpret data with caution.

*t, $p < .05$

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," and "Early Childhood Teacher Questionnaire."

Table 22. The percentage of young children who received preschool special education services whose parents were satisfied with special education services to various degrees: School years 2003-04 and 2004-05

	2003-04	2004-05
Very satisfied*	46.6	42.1
Satisfied	45.5	48.4
Dissatisfied	6.5	7.4
Very dissatisfied	1.5[!]	2.1

! Interpret data with caution.
* t, $p < .05$
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Parent Interview."

Summary

Teachers indicated that 89 percent of the children in the 2003-04 school year and 86 percent in 2004-05 received speech or language therapy. Occupational therapy (32% in 2003-04 and 35% in 2004-05) and learning strategies/study skills assistance by a special educator (30% in 2003-04 and 20% in 2004-05) were also commonly reported services both years. From 2003-04 to 2004-05, there was a statistically significant decrease in the percentage of children receiving nine types of services. The only service showing a significant increase was help from a one-to-one paraeducator or assistant. Furthermore, the mean number of special education and related services provided to young children with disabilities decreased from 3.5 in 2003-04 to 2.8 in 2004-05. Parent satisfaction remained relatively stable over that time period. There was a significant decline only in the number of parents who were *very satisfied*.

Chapter 5: A Year of Growth: The Knowledge and Skills of Preschoolers With Disabilities

Overall, preschoolers with disabilities who participated in the PEELS direct assessment in 2003-04 performed within one standard deviation of the national mean on tests of emerging literacy, early math proficiency, social-behavior skills, and motor skills. Mean scores on selected assessments varied by disability, age cohort, race/ethnicity, household income, and gender (Markowitz et al. 2006). This chapter takes those earlier analyses one step further by exploring change in scale scores from 2003-04 to 2004-05.

The augmented sample was used for the analyses presented in this chapter.[22] All analyses in this chapter were longitudinal, except those using the Adaptive Behavior Assessment System – II (ABAS-II), which were cross-sectional. To examine changes in scale scores from 2003-04 to 2004-05 for each subgroup, *t* tests for dependent samples were conducted. No analyses were conducted to compare differences in the amount of growth across the subgroups. The Benjamini-Hochberg procedure (Benjamini and Hochberg 1995) was used to control the family-wise error rate and avoid making false positive conclusions for multiple comparison analyses.

Emerging Literacy

This section describes the 1-year growth in emerging literacy skills for children who received preschool special education services and subgroups of that population defined by age cohort, gender, and disability category. Results for two assessments are included:

- the Woodcock-Johnson III: Letter-Word Identification and
- an adapted version of the Peabody Picture Vocabulary Test (PPVT).

Test scores for the emerging literacy measures reviewed in this section are standardized, so the mean for the population at each age is 100, for example. Even though children receive the same test items from year to year, more correct items are required for a 5-year-old to receive a standard score of 100 than for a 3-year-old to receive the same standard score. Population means remain constant, although subgroup means may go up or down. Performance on the various literacy measures that were included in the PEELS direct assessment suggest that children with disabilities maintained or improved their norm-referenced standing from school year 2003-04 to 2004-05. [23]

Woodcock-Johnson III: Letter-Word

The Letter-Word Identification test measures the child's word identification skills, without requiring the child to know the meaning of any word. It has a population mean of 100 and a standard deviation of 15. In school year 2003-04, young children who received preschool special education services performed as well, on average, as their peers without disabilities on Letter-Word Identification—with an overall mean performance of 98.2 (*S.E.* = 0.5). In school year 2004-05, the mean overall performance increased significantly to 100.2 (*S.E.* = 0.5) ($t = 7.427$, $p < .001$, see table 23).

---

[22] For children in the Supplemental Sample (6 percent of the augmented sample), we used imputed Wave 1 data.

[23] National norms for the assessments reported in this chapter are based on general population data, not data specifically for children with disabilities. National norms are used as benchmarks in the analyses to compare the PEELS children's growth from Wave 1 to Wave 2.

Table 23. Mean performance of young children who received preschool special education services on Woodcock-Johnson III: Letter-Word Identification, by age cohort: School years 2003-04 and 2004-05

	Total*	Cohort A*	Cohort B*	Cohort C*
2003-04	98.2	101.5	98.3	96.6
2004-05	100.2	102.7	100.9	98.2

* t, $p < .05$

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Letter-Word Identification."

Between 2003-04 and 2004-05, the mean performance on the Letter-Word Identification test increased significantly for each age cohort. Children in Cohort A had an increase in their performance, from 101.5 (*S.E.* = 1.1) in 2003-04 to 102.7 (*S.E.* = 0.8) in 2004-05 ($t = 2.334$, $p = .023$). The mean performance of children in Cohort B increased from 98.3 (*S.E.* = 0.8) to 100.9 (*S.E.* = 0.9) ($t = 6.239$, $p < .001$), and the mean performance of children in Cohort C increased from 96.6 (*S.E.* = 0.6) to 98.2 (*S.E.* = 0.7) (see table 23) ($t = 4.475$, $p < .001$).

Both males and females had a significant improvement in performance. The mean performance of males increased significantly, from 97.4 (*S.E.* = 0.6) in 2003-04 to 99.7 (*S.E.* = 0.5) in 2004-05 ($t = 8.119$, $p < .001$). The mean performance of females also increased significantly, from 99.9 (*S.E.* = 0.9) in 2003-04 to 101.2 (*S.E.* = 0.8) in 2004-05 ($t = 2.545$, $p = .013$).

Children identified as having a developmental delay and children identified as having a speech or language impairment had statistically significant improvements in performance on the Letter-Word Identification test from 2003-04 to 2004-05 (DD: $t = 4.008$, $p < .001$; SL: $t = 5.691$, $p < .001$) (see table 24).

Table 24. Mean performance of young children who received preschool special education services on Woodcock-Johnson III: Letter-Word Identification, by Wave 1 primary disability: School years 2003-04 and 2004-05

	AU	DD*	ED	LD	MR	OI	OHI	SLI*	LI
2003-04	109.0	94.1	95.7	95.2	87.3	96.3	96.4	99.8	94.8
2004-05	105.5	96.9	96.7	97.9	85.2	98.4	93.8	102.2	98.1

* t, $p < .05$

NOTE: AU = Autism; DD = Developmental delay; ED = Emotional disturbance; LD = Learning disability; MR = Mental retardation; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; LI = Low incidence.

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Letter-Word Identification."

PPVT-III

The direct assessment included a measure of receptive vocabulary using an adapted version of the PPVT-III. Receptive vocabulary also is referred to as listening vocabulary or oral vocabulary. It is considered a strong predictor of language acquisition and cognitive development and is a key component in emerging literacy. Overall, the PPVT-III performance of young children who received preschool

special education services did not change significantly between 2003-04 and 2004-05. In 2003-04, the mean performance was 89.9 (*S.E.* = 0.7), and in 2004-05, it was 89.6 (*S.E.* = 0.6; see table 25).

However, children in Cohort A had a statistically significant increase in performance on the PPVT-III (t = 3.406, p = .001). Their mean scale scores increased from 88.5 (*S.E.* = 0.7) in 2003-04 to 88.9 (*S.E.* = 0.6) in 2004-05 (see table 25).

Table 25. Mean performance of young children who received preschool special education services on Peabody Picture Vocabulary Test III, by age cohort: School years 2003-04 and 2004-05

	Total	Cohort A*	Cohort B	Cohort C
2003-04	89.9	88.5	89.3	91.0
2004-05	89.6	88.9	89.0	90.4

*t, $p < .05$

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Peabody Picture Vocabulary Test III."

The mean performance of males and females on the PPVT-III did not change significantly between 2003-04 and 2004-05. For males, scores averaged 89.6 (*S.E.* = 0.7) in 2003-04 and 89.2 (*S.E.* = 0.7) in 2004-05. For females, mean scores were 90.4 and 90.5 (*S.E.* = 1.2 and 1.1, respectively).

Table 26 reveals few statistically significant changes in mean performance on the PPVT-III by disability category. Overall, only the mean performance of children identified as having a low-incidence disability increased significantly, from 85.2 (*S.E.* = 1.8) in 2003-04 to 90.0 (*S.E.* = 1.7) in 2004-05 (t = 2.676, p = .01).

Table 26. Mean performance of young children who received preschool special education services on Peabody Picture Vocabulary Test III, by Wave 1 primary disability: School years 2003-04 and 2004-05

	AU	DD	ED	LD	MR	OI	OHI	SLI	LI*
2003-04	86.0	85.1	92.5	85.4	70.6	83.4	87.1	93.1	85.2
2004-05	81.1	85.0	94.1	85.0	66.5	86.7	82.0	93.4	90.0

* t, $p < .05$

NOTE: AU = Autism; DD = Developmental delay; ED = Emotional disturbance; LD = Learning disability; MR = Mental retardation; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; LI = Low incidence.

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Peabody Picture Vocabulary Test III."

Early Math Proficiency

Results for two norm-referenced math proficiency measures are presented in this section:

- Woodcock-Johnson III: Applied Problems, and
- Woodcock-Johnson III: Quantitative Concepts, including both concepts and number series.

Similar to the emerging literacy measures reviewed in the previous section, test scores for the early math measures reviewed in this section are standardized, so the mean for the population at each age is 100, for example.

Overall, performance on the early math measures significantly improved for PEELS children suggesting that children with disabilities maintained or improved their norm-referenced standing from school year 2003-04 to 2004-05.

Woodcock-Johnson III: Applied Problems

The Applied Problems test is a measure of children's ability to analyze and solve practical math problems using simple counting, addition, or subtraction operations. In school year 2003-04, the mean overall performance of young children who received preschool special education services was 90.8 (*S.E.* = 0.7). In school year 2004-05, the mean overall performance was 91.9 (*S.E.* = 0.7), which was a statistically significant increase ($t = 4.556$, $p < .001$).

The performance of children in Cohort A on the Applied Problems test increased from 88.9 (*S.E.* = 1.1) in 2003-04 to 91.7 (*S.E.* = 0.9) in 2004-05 (see table 27). Similarly, the mean performance of children in Cohort C increased from 90.7 (*S.E.* = 1.1) in 2003-04 to 93.1 (*S.E.* = 1.1) in 2004-05. These increases in performance were both statistically significant (Cohort A: $t = 5.608$, $p > .001$; Cohort C: $t = 5.046$, $p < .001$). The average performance of children in Cohort B did not change significantly.

Table 27. Mean performance of young children who received preschool special education services on Woodcock-Johnson III: Applied Problems, by age cohort: School years 2003-04 and 2004-05

	Total*	Cohort A*	Cohort B	Cohort C*
2003-04	90.8	88.9	91.9	90.7
2004-05	91.9	91.7	90.8	93.1

* t, $p < .05$

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Applied Problems."

The mean performance for males on the Applied Problems test was 90.2 (*S.E.* = 0.8) in 2003-04 and increased significantly to 91.8 (*S.E.* = 0.7) in 2004-05 ($t = 5.704$, $p < .001$). Females had the same mean performance of 92.2 in both 2003-04 (*S.E.* = 1.1) and 2004-05 (*S.E.* = 1.2, see table 28).

Table 28. Mean performance of young children who received preschool special education services on Woodcock-Johnson III: Applied Problems, by gender: School years 2003-04 and 2004-05

	Male*	Female
2003-04	90.2	92.2
2004-05	91.8	92.2

* t, $p < .05$

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Applied Problems."

Changes in mean performance on the Applied Problems test from 2003-04 to school year 2004-05 varied by disability category. In 2003-04, children identified as having a developmental delay had a mean performance of 83.7 (*S.E.* = 1.2), which increased to 86.0 (*S.E.* = 0.9) in 2004-05. This increase was statistically significant ($t = 3.488$, $p = .001$). Similarly, the mean performance of children identified as having a learning disability increased significantly ($t = 4.257$, $p < .001$). Children identified as having a speech or language impairment also had statistically significant increases in their mean performance ($t = 3.517$, $p = .001$) (see table 29).

Table 29. Mean performance of young children who received preschool special education services on Woodcock-Johnson III: Applied Problems, by Wave 1 primary disability: School years 2003-04 and 2004-05

	AU	DD*	ED	LD*	MR	OI	OHI	SLI*	LI
2003-04	81.8	83.7	91.6	85.3	62.7	91.0	84.0	96.1	84.2
2004-05	78.8	86.0	90.2	91.1	59.8	90.1	78.1	97.8	86.4

* $t, p < .05$

NOTE: AU = Autism; DD = Developmental delay; ED = Emotional disturbance; LD = Learning disability; MR = Mental retardation; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; LI = Low incidence.

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Applied Problems."

Woodcock-Johnson III: Quantitative Concepts

The Quantitative Concepts test measures basic mathematical concepts, symbols, and vocabulary. The score on this subtest is derived from questions that test important relational concepts and questions about numbers and number patterns. Children in Cohort C, the only cohort for which both years of data are available, had a statistically significant increase in performance on the Quantitative Concepts test, from 90.9 (*S.E.* = 0.8) in 2003-04 to 93.9 (*S.E.* = 0.9) in 2004-05 ($t = 5.945$, $p < .001$) (see table 30).

Table 30. Mean performance of young children who received preschool special education services on Woodcock-Johnson III: Quantitative Concepts, by age cohort: School years 2003-04 and 2004-05

	Total*	Cohort A	Cohort B	Cohort C*
2003-04	90.9	†	†	90.9
2004-05	93.9	†	94.0	93.9

† Not applicable.

* $t, p < .05$

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Quantitative Concepts."

Social Behavior

This section presents results on children's social behavior based on a series of teacher rating scales that include the PKBS Social Skills, Adaptive Behavior Assessment System II – Self-Care, and Adaptive Behavior Assessment System II – Self-Direction scales. Note that for most children, different

teachers rated the child's behavior in school years 2003-04 and 2004-05, which may represent a source of bias.

Preschool and Kindergarten Behavior Scales 2 – Social Skills

The Social Skills scale of the PKBS-2 assesses age-appropriate personal and interpersonal behaviors of preschool and early elementary-age children. Age-appropriate personal behaviors include such things as, "works or plays independently," "follows rules," and "accepts decisions made by adults." Age-appropriate interpersonal behaviors include such things as "is cooperative," "comforts other children who are upset," and "takes turns with toys and other objects." Teachers were asked to rate how frequently the identified child exhibited a series of skills or behaviors such as those noted above during the previous 3 months. The measurement scale consists of four points, labeled *never*, *rarely*, *sometimes*, and *often*. A higher rating on the Social Skills composite index indicates a higher level of social adjustment. Scores are standardized with a mean of 100 and a standard deviation of 15. Rating scales were often completed by different teachers in 2003-04 and 2004-05, introducing a potential source of bias.

On the Social Skills subscale, mean teacher ratings for young children who received special education services increased significantly from 92.9 (*S.E.* = 0.5) in school year 2003-04 to 96.0 (*S.E.* =0.6) in 2004-05 ($t = 6.287$, $p < .001$) (see table 31). Children in Cohort A had a significant increase in teacher ratings on the Social Skills subscale, from 84.7 (*S.E.* = 0.9) in 2003-04 to 93.7 (*S.E.* = 0.9) in 2004-05 ($t = 11.06$, $p < .001$). Children in Cohort B also had a statistically significant increase in their mean teacher rating, from 93.7 (*S.E.* = 1.0) in 2003-04 to 95.7 (*S.E.* = 0.7) in 2004-05 ($t = 2.068$, $p = .043$). Changes for children in Cohort C were not significant (2003-04, $M = 96.6$, *S.E.* = 1.2; 2004-05, $M = 97.6$, *S.E.* = 1.1).

Table 31. Mean teacher ratings of young children who received preschool special education services on the Social Skills subscale of the Preschool and Kindergarten Behavior Scale, by age cohort: School years 2003-04 and 2004-05

	Total*	Cohort A*	Cohort B*	Cohort C
2003-04	92.9	84.7	93.7	96.6
2004-05	96.0	93.7	95.7	97.6

* t, $p < .05$

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Social Skills subscale of the Preschool and Kindergarten Behavior Scale."

Both males and females had a statistically significant increase in performance on the Social Skills subscale from 2003-04 to 2004-05. The mean teacher rating for males increased from 91.4 (*S.E.* = 0.6) in 2003-04 to 94.4 (*S.E.* = 0.6) in 2004-05 ($t = 4.876$, $p < .001$) (see table 32). The mean for females increased from 96.6 (*S.E.* = 1.6) in 2003-04 to 99.8 (*S.E.* = 1.2) in 2004-05 ($t = 3.566$, $p = .001$).

Few statistically significant changes in teacher ratings were found on the Social Skills subscale when analyzed by disability category. However, mean teacher ratings for children identified as having a developmental delay ($t = 2.921$, $p = .005$) and for children identified as having a speech or language impairment ($t = 3.963$, $p < .001$) increased significantly (see table 33).

Table 32. Mean teacher ratings of young children who received preschool special education services on the Social Skills subscale of the Preschool and Kindergarten Behavior Scale, by gender: School years 2003-04 and 2004-05

	Male*	Female*
2003-04	91.4	96.6
2004-05	94.4	99.8

* t, $p < .05$

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Social Skills subscale of the Preschool and Kindergarten Behavior Scale."

Table 33. Mean teacher ratings of young children who received preschool special education services on the Social Skills subscale of the Preschool and Kindergarten Behavior Scale, by Wave 1 primary disability: School years 2003-04 and 2004-05

	AU	DD*	ED	LD	MR	OI	OHI	SLI*	LI
2003-04	73.7	89.3	89.9	90.3	69.2	100.2	90.4	99.7	85.5
2004-05	74.7	93.3	87.8	99.7	71.0	99.1	102.0	102.5	86.0

* t, $p < .05$

NOTE: AU = Autism; DD = Developmental delay; ED = Emotional disturbance; LD = Learning disability; MR = Mental retardation; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; LI = Low incidence.

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Social Skills subscale of the Preschool and Kindergarten Behavior Scale."

Adaptive Behavior Assessment System II (ABAS-II) – Self-Care and Self-Direction

As mentioned previously, the Teacher/Daycare Provider (TDP) Form of the ABAS-II was used for children in preschool or not in school and the Teacher Form (TF) for children in kindergarten or elementary school. The following analyses are based upon cross-sectional data of different samples of children for 2003-04 and 2004-05, reported by different teachers, using different forms. Therefore, the average change in performance between the two waves was not calculated for the ABAS-II subscales. Scaled scores are based on a mean of 10 and a standard deviation of 3.

The Self-Care scale of the ABAS-II includes items that measure the child's basic personal care skills, including eating, dressing, bathing, toileting, grooming, and hygiene. The TDP Form includes items such as "swallows liquids with no difficulty" "nurses, drinks, or eats willingly, with little encouragement" and "tells teacher/daycare provider or other adult when he/she needs to use the bathroom." The TF includes some of the same items, but also includes items that are more appropriate for an older child. These items include such things as "uses the school restroom alone," "uses a fork to eat solid food" and "opens a tab-top can, milk carton, or screw-top bottle." For each of the skill areas, teachers rated the frequency with which the identified child exhibited specific behaviors using a 4-point scale, with ratings of the child's abilities as *is not able, never when needed, sometimes when needed,* and *always when needed.*

In 2003-04, the mean teacher/daycare provider rating on the Self-Care subscale for young children who received preschool special education services was 8.9 (*S.E.* = 0.1), and in 2004-05, it was 8.3 (*S.E.* = 0.2). The mean teacher rating for kindergarteners was 8.0 (*S.E.* = 0.4) in 2003-04 and 8.5 (*S.E.*

= 0.2) in 2004-05. The Self-Direction subscale measures the child's skills in independence, self-control, and personal responsibility. The TDP Form includes such items as child "shows interest in a toy or other object by looking at it for a few seconds," "follows an adult's request to quiet down or behave," and "looks for misplaced toys or games until he/she finds them." The TF includes such items as the child "routinely arrives at school or class on time," "works independently and asks for help only when necessary," and "stops a fun activity, without complaints, when told that time is up." As with the Self-Care subtest, scaled scores are based on a mean of 10 and a standard deviation of 3.

In school year 2003-04, the mean teacher/daycare provider rating on the Self-Direction subscale was 9.6 (*S.E.* = 0.1). In school year 2004-05, the mean rating was 9.5 (*S.E.* = 0.1). The mean for kindergarteners was 8.3 (*S.E.* = 0.3) in 2003-04 and 8.5 (*S.E.* = 0.2) in 2004-05.

Motor Skills

To provide a measure of each child's motor skills, the Fine Motor and Gross Motor subscales from the Vineland Adaptive Behavior Scales Classroom Edition were included in the teacher questionnaires. The results presented in this section are based on longitudinal analyses conducted on the data for the augmented sample. In school year 2003-04, the mean Motor Skills scale score for children who received preschool special education services was 94.4 (*S.E.* = 0.9). In school year 2004-05, the mean was 96.2 (*S.E.* = 1.0), representing a statistically significant increase ($t = 3.134$, $p = .003$).

The ratings of children in Cohort A significantly increased from 92.8 (*S.E.* = 0.9) in 2003-04 to 95.2 (*S.E.* = 1.0) in 2004-05 ($t = 3.207$, $p = .002$) (see table 34). Similarly, the mean ratings of children in Cohort C significantly increased from 95.4 (*S.E.* = 1.6) in 2003-04 to 98.4 (*S.E.* = 1.6) in 2004-05 ($t = 2.696$, $p = .009$). On average, the ratings of children in Cohort B did not change.

Table 34. Mean teacher ratings of young children who received preschool special education services on the Vineland Adaptive Behavior Scale: Motor Skills Domain, by age cohort: School years 2003-04 and 2004-05

	Total*	Cohort A*	Cohort B	Cohort C*
2003-04	94.4	92.8	94.3	95.4
2004-05	96.2	95.2	94.4	98.4

* $t, p < .05$

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Vineland Adaptive Behavior Scale: Motor Skills Domain."

The overall mean ratings of males on the Motor Skills test increased significantly from 93.9 (*S.E.* = 0.8) in 2003-04 to 95.8 (*S.E.* = 1.0) in 2004-05 ($t = 2.804$, $p = .007$). Ratings on the Motor Skills test also increased significantly for male children in Cohort A and Cohort C (Cohort A: $t = 2.25$, $p = .028$; Cohort C: $t = 2.347$, $p = .022$). Mean ratings for children identified as having a learning disability increased significantly on the Motor Skills scale, from 90.5 (*S.E.* = 3.1) in 2003-04 to 101.9 (*S.E.* = 5.1) in 2004-05 ($t = 2.223$, $p = .03$). Children identified as having a developmental delay also had statistically significant increases in their ratings from 89.8 (*S.E.* = 1.2) in 2003-04 to 91.4 (*S.E.* = 1.4) in 2004-05 ($t = 2.722$, $p = .008$).

Alternate Assessment Results

Alternate assessments were completed for PEELS children who were not capable of participating in the direct assessment. For the alternate assessment, the child's teacher or service provider completed the Adaptive Behavior Assessment System (ABAS-II). An alternate assessment was completed for 331 children (12 percent of the total sample) in Wave 1 and 228 children (7 percent of the total sample) in Wave 2. Children who had an alternate assessment completed for them in Wave 1 did not necessarily have one completed for them in Wave 2, and vice versa. Thus, the following analyses are based on cross-sectional data of different samples of children for 2003-04 and 2004-05.

The children were identified for an alternate assessment in several ways. The first was during a screening interview prior to the assessment. If the respondent, who was familiar with the child, indicated that the child (1) could not understand and follow simple instructions that were spoken aloud or given in sign language; (2) had a visual impairment that would prohibit participation in an assessment that primarily involved pictures, text, and numbers; or (3) did not speak English or Spanish, then an alternate assessment was given. Alternate assessments were also completed for children who scored four or less on the two English-language subtests that were included in the direct assessment.

The ABAS-II is designed to evaluate whether an individual displays the adaptive skills necessary for daily living without the assistance of others. The adaptive skills measured by the ABAS-II are defined as "those practical, everyday skills required to function and meet environmental demands, including effectively and independently taking care of oneself and interacting with other people" (Harrison and Oakland 2003, p.3). Examples of adaptive skills that preschoolers use on a daily basis include those related to eating, dressing, expressing needs, interacting with peers, controlling one's behavior in a structured setting, and communicating with other people.

As noted in chapter 2, two forms of the ABAS-II were used for the PEELS alternate assessment. The TDP Form is designed for children ages 2 to 5, while the TF is designed for children in grades kindergarten through 12 (or ages 5 to 21). For each of the skill areas, teachers rated the frequency with which the identified child exhibited specific behaviors using a 4-point scale, with ratings of the child's abilities as *is not able, never when needed, sometimes when needed,* and *always when needed.* The specific skill areas measured by the ABAS-II used for the alternate assessment are Functional Academics, Self-Care, and Self-Direction, which were completed for all PEELS children as part of the teacher questionnaire, plus Communication, Community Use, School Living, Health and Safety, Leisure, Social, and Motor skills.[24] Data for children whose teachers completed the alternate assessment were not included in the Self-Care and Self-Direction analyses reported earlier in this chapter. The norm-referenced standardized scores for the ABAS-II have a mean of 10 and a standard deviation of 3.

Overall, the mean performance of preschoolers for whom an alternate assessment was completed was more than one standard deviation lower than the population mean for each of the skill areas measured by the ABAS-II. In most cases, the performance of preschoolers for whom an alternate assessment was completed was more than two standard deviations below the normed mean of 10. Table 35 presents the average teacher ratings (using the TDP Form) for PEELS children not yet in kindergarten on each of the ABAS-II skill areas. Table 36 presents the average teacher ratings (using the TF) for PEELS children in kindergarten or elementary school.

[24] The Community Use subscale is not included on the TDP Form. The Motor subscale is not included on the TF.

In 2003-04, mean teacher/daycare provider ratings for non-kindergarteners was 4.0 for the Motor subscale (*S.E.* = 0.2) and 2.2 for the Communication (*S.E.* = 0.2) and Social (*S.E.* = 0.1) subscales (see table 35). In school year 2004-05, non-kindergarteners' rating was 4.1 on the Self-Direction subscale (*S.E.* = 0.3) and 1.8 on the Communication subscale (*S.E.* = 0.1) (see table 35).

Table 35. Mean teacher ratings of young children who received preschool special education services and participated in an alternate assessment: ABAS-II—Skill area scores (Teacher/Daycare Provider Form): School years 2003-04 and 2004-05

	Total
Communication	
2003-04	2.2
2004-05	1.8
Functional (Pre) academics	
2003-04	3.2
2004-05	3.1
Health and safety	
2003-04	2.5
2004-05	2.6
Leisure	
2003-04	2.9
2004-05	3.0
Motor	
2003-04	4.0
2004-05	3.6
School living	
2003-04	3.5
2004-05	3.9
Self-care	
2003-04	3.2
2004-05	3.4
Self-direction	
2003-04	3.6
2004-05	4.1
Social	
2003-04	2.2
2004-05	2.3

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "ABAS-II—skill area scores."

In 2003-04, mean teacher ratings using the TF were 1.3 for the Communication (*S.E.* = 0.2) and Self-Direction (*S.E.* = 0.1) subscales and 3.8 for the Functional Academics (*S.E.* = 0.6) subscale. In school year 2004-05, mean teacher ratings using the TF were 1.4 for the Self-Direction subscale (*S.E.* = 0.1) and 3.8 for the Functional Academics (*S.E.* = 0.3) subscale (see table 36).

Table 36. Mean teacher ratings of young children who received preschool special education services and participated in an alternate assessment: ABAS-II—Skill area scores (Teacher Form): School years 2003-04 and 2004-05

	Total
Communication	
2003-04	1.3
2004-05	1.5
Community use	
2003-04	1.8
2004-05	1.8
Functional academics	
2003-04	3.8
2004-05	3.8
Health and safety	
2003-04	2.3[!]
2004-05	1.8
Leisure	
2003-04	2.1[!]
2004-05	2.6
School living	
2003-04	3.3
2004-05	2.9
Self-care	
2003-04	2.6[!]
2004-05	2.0
Self-direction	
2003-04	1.3
2004-05	1.4
Social	
2003-04	1.7
2004-05	1.8

! Interpret data with caution.

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "ABAS-II—skill area scores."

Summary

Data from 2004-05 indicate that children who received preschool special education services and took the direct PEELS assessment performed close to the mean for the norm-population on letter-word identification, social skills, and motor skills, and these scores were significantly higher than in 2003-04. The children scored within one standard deviation of the norm-referenced mean on a test of vocabulary and on both early math tests. Math scores were significantly higher in 2004-05 than in 2003-04. Children

who participated in the alternate assessment had mean ratings more than two standard deviations below the population mean in most skill areas.

References

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of Royal Statistical Society*, *Series B*, *57*: 289-300.

Carlson, E. (2004). *Report of PEELS Sample Recruitment*. Rockville, MD: Westat.

Carlson, E. (1997). Outcomes for Students Declassified From Special Education. *Dissertation Abstracts International, 58*(11): 4231 (UMI No. 9815240).

Carlson, E., and Parshall, L. (1996). Academic, Social, and Behavioral Adjustment for Students Declassified From Special Education. *Exceptional Children, 63*: 89-100.

Carrow-Woolfolk, E. (1995). Oral and Written Language Scales: Listening Comprehension and Oral Expression. Circle Pines, MN: American Guidance Service.

Connors, C.K. (1990). *Connors' Rating Scales Manual*. North Tonawanda, NY: Multi-Health Systems.

Council for Exceptional Children. (2000). *Making Assessment Accommodations: A Toolkit for Educators.* Reston, VA: Author.

Danaher, J., Kraus, R., Armijo, C., and Hipps, C. (2003). *Section 619 Profile*. Chapel Hill, NC: National Early Childhood Technical Assistance Center.

De Avila, E., and Duncan, S. (1998). *preLAS 2000.* Monterey, CA: CTB McGraw-Hill.

Dunn, L.M., and Dunn, L.M. (1997). *Peabody Picture Vocabulary Test-Third Edition*. Circle Pines, MN: American Guidance Services.

ECRI-MGD (Early Childhood Research Institute on Measuring Growth and Development). (2001). *Individual Growth and Development Indicators*. Minneapolis, MN: Center for Early Education and Development, University of Minnesota.

Gresham, F.M., and Elliott, S.N. (1990). *Social Skills Rating System*. Circle Pines, MN: American Guidance Service.

Halgren, D.P., and Clarizio, H.F. (1993). Categorical and Programming Changes in Special Education Services. *Exceptional Children, 59*: 547-555.

Harrison, J., and Oakland, T. (2003). *Adaptive Behavior Assessment System* (2nd edition). San Antonio, TX: Harcourt Assessment Company, The Psychological Corporation.

Individuals with Disabilities Education Act, P. L. 05-17, 105 Stat. 587 (codified as amended at 20 U.S.C. §1400 et seq.

Innocenti, M.S. (2005). *Classification Status of Children Once Enrolled in Early Intervention Programs*. Report submitted to the Utah State Office of Education.

Kaufman, A.S., and Kaufman, N.L. (1993). *Kaufman Adolescent and Adult Intelligence Test*. Circle Pines, MN: American Guidance Service.

Kaufman, A.S., and Kaufman, N.L. (1990). *Kaufman Brief Intelligence Test*. Circle Pines, MN: American Guidance Service.

Kaufman, A.S., and Kaufman, N.L. (1985). *Kaufman Test of Educational Achievement*. Circle Pines, MN: American Guidance Service.

Lambert, N., and Windmiller, M. (1981). *AAMD Adaptive Behavior Scale – School Edition*. Monterey, CA: Publishers Test Service.

Markowitz, J., and Bridges-Cline, F. (1991, November). *A Six-Year Study of Children Enrolled in Special Education Under the Age of Five Years*. Rockville, MD: Montgomery County Public Schools.

Markowitz, J., Carlson, E., Frey, W., Riley, J., Shimshak, A., Heinzen, H., Strohl, J., Lee, H., and Klein, S. (2006). *Preschoolers' Characteristics, Services, and Results: Wave 1 Overview Report from the Pre-Elementary Education Longitudinal Study (PEELS)*. Rockville, MD: Westat.

Markwardt, F. (1989). *Peabody Individual Achievement Test-Revised*. Circle Pines. MN: American Guidance Service.

Matson, J.L., Esvelt-Dawson, K., and Kazdin, A.E. (1983). Validation Methods for Assessing Social Skills in Children. *Journal of Clinical Psychology, 12*: 175-180.

McGrew, K.S., and Woodcock, R.W. (2001). Technical Manual. *Woodcock-Johnson III.* Itasca, IL: Riverside Publishing.

Merrell, K.W. (2002). *Preschool and Kindergarten Behavior Scales, Second Edition.* Austin, TX: PRO-ED.

Merrell, K.W. (1995). Relationships Among Early Childhood Behavior Rating Scales: Convergent and Discriminant Construct Validity of the Preschool and Kindergarten Behavior Scales. *Early Education and Development, 6*: 253-264.

Merrell, K.W. (1993). *School Social Behavior Scales*. Eugene, OR: Assessment-Intervention Resources.

Roid, G.H., and Miller, L.J. (1995, 1997). *Leiter International Performance Scale-Revised*. Wood Dale, IL: Stoelting Co.

Sparrow, S.S., Balla, D.A., and Cicchetti, D.V. (1985). *Vineland Adaptive Behavior Scales* (Classroom Edition), Circle Pines, MN: American Guidance Service.

Stocking, M.L., and Lord, F.M. (1983). Developing a Common Metric in Item Response Theory. *Applied Measurement, 7*: 201-210.

Thurlow, M.L., Lehr, C.A., and Ysseldyke, J.E. (1987). Exit Criteria in Early Childhood Programs for Handicapped Children. *Journal of the Division for Early Childhood, 11*(2): 118-123.

Thurlow, M.L., Ysseldyke, J.E., and Weiss, J.A. (1988). Early Childhood Special Education Exit Decisions: How Are They Made? How Are They Evaluated? *Journal of the Division for Early Childhood, 12*(3): 253-262.

U.S. Department of Health and Human Services (US HHS), Administration for Children and Families, Head Start Bureau. (2005, February). *Report to Congress: Assessment of Children in Head Start Programs.* Washington, DC: U.S. Government Printing Office.

VESID Program Quality and Its Effect on the Placement of Preschool Education Children in School-age Programs. Special Education Quality Indicator Study. August 2003.

Walker, D., Singer, J., Palfrey, J., Orza, M., Wenger, M., and Butler, J. (1988). Who Leaves and Who Stays in Special Education: A 2-Year Follow-Up Study. *Exceptional Children, 54*: 393-402.

Wechsler, D. (1992). *Wechsler Individual Achievement Test.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children–Third Edition.* San Antonio, TX: Psychological Corporation.

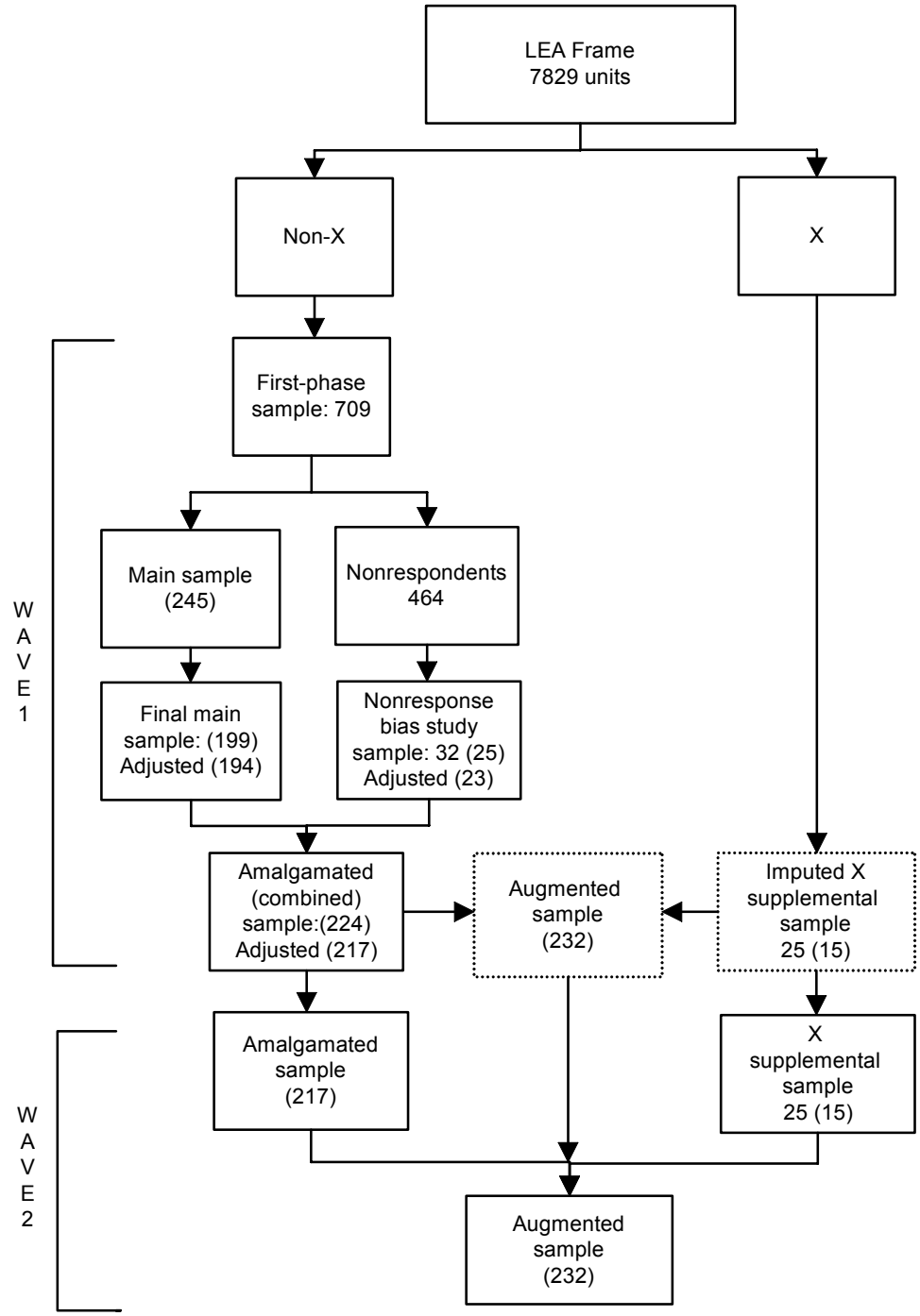
Westat. (2002). *WesVar Version 4.2*. Rockville, MD: Author.

Wong, M.M. (1997). Patterns of Special Education Placement for Preschool and School-age Children. *Dissertation Abstracts International, 58*(12): 4616 (UMI No. 9819362).

Woodcock, R.W., McGrew, K.S., and Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside Publishing.

Woodcock, R. and Muñoz-Sandoval, A. (1996). *Pruedas de Aprovechamiento Revisada*. Itasca, IL: Riverside Publishing.

Appendix A: Diagram of Selection of LEA Sample



Note: X stands for the state that originally did not participate. LEA counts for X and non-X were suppressed for confidentiality reasons. The figures in parentheses are the number of participating LEAs. They were adjusted as the LEAs, which did not contribute any data, were dropped. The dotted boxes represent a mirror image created by imputation of the X supplemental sample selected in Wave 2.

Appendix B: Weighting Procedures

This appendix describes weighting procedures used in Waves 1 and 2 of PEELS. The PEELS study was designed to use a nationally representative sample of local education agencies (LEAs) and children ages 3 through 5 with disabilities to generate weighted estimates that reflect the characteristics of the population, not the sample.

District Weighting

The LEA weighting procedure includes developing base weights and replicate weights. Replicate weights were generated for each set of full-sample weights to allow the creation of estimated standard errors on all statistics.

District Base Weights

Calculation of the base weights started with the first-stage sample of 709 LEAs for the amalgamated sample and 25 LEAs for the supplemental sample. Analysis of nonresponse patterns revealed that nonresponse adjustment to the base sampling weights for the main sample could be carried out within the design stratum cells. Therefore, district base weights were recomputed within each sampling stratum cell as the number of districts on the sampling frame divided by the number of districts that participated in the study. The sum of the base weights represents 7,829 districts.[25] These weights will be denoted as w_h, which is the same for all LEAs within a stratum cell (defined by district size, region, and wealth category for non-supplemental LEAs and by district size alone for supplemental sample LEAs).

Replicate Weights

Replicate weights were developed to facilitate variance estimation using Westat's proprietary software, WesVar. Due to restrictions in the DAS software that will be used for data dissemination, the jackknife method JK2 with 62 replicates was used instead of the JKn method used previously for Wave 1 weighting.

The JK2 method requires defining the variance strata and two variance units per variance stratum. The variance strata were defined by the sampling strata by size, region, and wealth at the beginning. However, sampling strata with no or a small number of responding LEAs were collapsed with a neighboring stratum cell with similar sampling rates. Sampling strata with a large number of LEAs were split into two variance strata. Altogether, 62 variance strata were created. Variance units were formed by randomly grouping districts within each variance stratum up to three variance units. The number of groups was determined by the number of replicates.

The replicate weights were then created for the JK2 method. If there are two variance units, this is done by assigning a zero weight to records in one variance unit chosen randomly and doubling the weights for records in the other variance units from the same variance stratum but leaving the weights for records in other variance strata unchanged. If the randomly chosen variance unit from the *i*-th variance

[25] This number is different from the total number of LEAs in the country because the smallest LEAs were not covered by the sample design.

stratum is denoted as U_{i1} and the other variance unit as U_{i2}, algebraically the *i*-th replicate weight for the *j*-th LEA record, w_{ij}^{*}, is given by

$$w_{ij}^{*} = \begin{cases} 0 & \text{if the } j \text{ - th record is in } U_{i1} \\ 2w_h & \text{if the } j \text{ - th record is in } U_{i2} \\ w_h & \text{if the } j \text{ - th record is not in the } i \text{ - th variance stratum} \end{cases}$$

where w_h is the full sample base weight for the stratum cell *h* to which the *j*-th LEA belongs, *i* = 1, 2, …, 62; *j* = 1, 2, …, 232.

If there are three variance units, replicate weight calculation is more complex. In this case, another variance stratum number is needed; usually, an existing number is arbitrarily assigned. Let this be *k* and the three variance units be randomly ordered as U_{i1}, U_{i2}, and U_{i3}. The replicate weight that corresponds to this situation is defined as:

$$w_{ij}^{*} = \begin{cases} 0 & \text{if } j \text{ - th record is in } U_{i1} \\ 1.5w_h & \text{if } j \text{ - th record is in } U_{i2} \\ 1.5w_h & \text{if } j \text{ - th record is in } U_{\mathrm{i3}} \end{cases}$$

and

$$w_{kj}^{*} = \begin{cases} 1.5w_h & \text{if } j \text{ - th record is in } U_{i1} \\ 0 & \text{if } j \text{ - th record is in } U_{i2} \\ 1.5w_h & \text{if } j \text{ - th record is in } U_{\mathrm{i3}} \end{cases}$$

Consequently, each LEA has a base weight w_h and 62 replicate weights, w_{1j}^{*}, w_{2j}^{*}, …, w_{62j}^{*}.

Child Weighting: Within LEA Child Base Weight

After the child sampling was finished, the sampling status was defined by child status ID, which has 15 categories shown in table B-1.

The status codes 1, 2, and 4 are interim codes, and no child should have this code at the end of data collection in each wave. A large number of children have a status code of 3 since they were passed through the sampling system but not selected into the sample (those who were selected had a code value of 4 but subsequently moved to one of the remaining categories). Only children in category 6 are enrolled for the study. Children in categories 9 and 11 were selected first but then deselected due to the maximum 80-children limit for each district or district-wide non-participation. These and 1, 2, 8, and 12 are treated as not passed in the sampling system. Status codes 60, 61, and 62 are relevant only to the children in Wave 2.

Table B-1. Child status codes

Code	Definition	Description
1	Entering	The child record is entered into the computer system.
2	Ready sample	The child record is ready for sampling.
3	Sampled	The child record has gone through the sampling system.
4	Selected	The child record is selected into the sample.
5	Ineligible	The child is ineligible.
6	Enrolled	The child is enrolled for the study.
7	Declined	The child has declined.
8	Max reached/not sampled	The record is not sampled because the district has reached the cap of 80.
9	Max reached/deselected	The record is selected but subsequently deselected because the district has reached the cap of 80.
10	Nonresponse	The child was selected but did not respond.
11	Deselected-No LEA/child participation	The child was selected but subsequently deselected because neither LEA questionnaire was filled out nor any child participated in the study.
12	Desampled/district nonparticipation	The child was sampled but subsequently desampled because the whole district dropped out of the study.
60	Deceased	The child died after Wave 1.
61	Ineligible	The child turned out to be ineligible after Wave 1.
62	Study withdrawal	The child withdrew from the study after Wave 1.

Child sampling was done using the sampling system within sampling strata (called LEA-cohort) defined by District ID and the five cohort IDs [3-years-old ongoing (A_O), 4-years-old ongoing (B_O), 4-years-old historical (B_H), 5-years-old ongoing (C_O), 5-years-old historical (C_H)].

During reweighting it was found that nine children had incorrect birthdates. The correction of their birthdates altered their sampling LEA-cohort strata. We recomputed sampling rates of those affected LEA-cohort strata assuming the realized strata are the real strata from which they were selected. Four children from two LEAs swapped their LEA-cohort strata within their LEAs, and thus no change in the sampling rate was necessary for them. This approach may be termed as conditional on the realized LEA-cohort strata. This may introduce some bias but will reduce the variance. We believe that the bias introduced by this approach is negligible because the number of problem cases is small, and the sampling rate changes are not great.

A within-LEA base sampling weight for children by child sampling stratum was created for all sampled and selected children (categories 5, 6, 7, 10, 60, 61, 62) based on the sampling rate. The weight for a selected child i in an LEA-cohort within LEA stratum h is defined as the inverse of the sampling rate that was applied:

$$w_{hi}^{c} = \frac{1}{r_{hi}} .$$

Note that the subscript i now identifies sample children, so it has a different meaning from the one used in the previous section. The sampling rate r_{hi} depends on the LEA stratum h, where the child's LEA is contained, and the child's particular LEA-cohort.

The sampling rate changed during the sampling process for many LEA-cohort strata, so children in those LEA-cohort strata were selected with a different sampling rate from that of other children in the same LEA-cohort stratum, depending on the time of sampling. Therefore, the children from the same LEA may have different base weights.

The sum of unconditional base weights in a cohort is close but not equal to the child list total of the cohort. We first considered using a conditional approach that defines the within-LEA child weight based on the realized sample size instead of using the sampling rate. This approach cuts down the variance due to random sample sizes that resulted from the Bernoulli sampling procedure used for child sampling from the ongoing lists. However, this approach became problematic because 48 LEA-cohort strata did not have any children selected due to small sampling rates and inaccurate list size estimates used to calculate the sampling rates and also by chance. Therefore, if we used the conditional approach, children from the 48 LEA-cohort strata would not be represented. To avoid this problem, we used the unconditional approach and the corresponding formula given above.

There are two exceptions to using unconditional weights:

- First, for LEA-cohort strata that have some children in categories 1, 2, 8, and 9, we used the conditional weighting method because not all the children were covered by the unconditional weighting; that is, some children were unsampled or deselected, which makes the sampling rate used for sample selection wrong. For these cases, the conditional weight was calculated by dividing the child list total of the LEA-cohort by the actual number of children selected for the LEA-cohort:

$$w_{hi}^{c} = \frac{N_{hi}}{n_{hi}}.$$

  The conditional weight was the same for every child and summed exactly to the list total of the LEA-cohort stratum.

- Second, after we performed the weighting using the methods above, we checked the sum of weights against the list counts, by cohort, and found some large differences, which were mainly due to large discrepancies for the following LEA-cohorts: 1457B_O, 1457C_O, 3319C_H, 3495C_O, 1060C_O, 2044B_H, 2596B_H, 1917C_H, 1519B_H, 3256B_H, 9002A_O, 9002_B_O, 2549C_H, 1519A_O, 2864B_H, and 1472B_H. We recalculated the sampling weights using the conditional approach for them.

With this correction, the sum of weights was almost the same as the overall list total. The weights also agree quite well at various levels of aggregation.

Child Base Weight

The overall weight for the selected children was created by multiplying the child base weight and the LEA full sample weights, w_h, defined earlier:

$$w_{hi} = w_h w_{hi}^{c}.$$

The overall child replicate weights are then obtained by multiplying the child base weight and the LEA replicate weights.

Noncoverage Adjustment for Smallest LEAs

In the PEELS sample design, size 5 (very small) LEAs were not sampled. This is because size 5 LEAs accounted for only a small percentage of the whole target population but required more resources to sample because they are numerous. We decided to adjust for the noncoverage of size 5 children by increasing the size 4 children's base weights by a ratio factor calculated from the original frame stratified by region and wealth. Note that only size 4 children's weights are adjusted. The adjusted weights are given by

$$w_{hi}^{*} = \begin{cases} w_{hi}, & \text{if size less than 4,} \\ w_{hi} f_{hi}^{\text{cov}}, & \text{if size = 4,} \end{cases}$$

where f_{hi}^{cov} is the coverage adjustment factor for size 4 LEAs. Table B-2 shows the factors by region and wealth class.

Table B-2. Non-coverage adjustment factors

Region	Wealth	Non-coverage factor
1	1	1.0798
1	2	1.1203
1	3	1.2089
1	4	1.4796
2	1	1.0530
2	2	1.0391
2	3	1.0517
2	4	1.0699
3	1	1.1428
3	2	1.2300
3	3	1.4222
3	4	1.5694
4	1	1.2022
4	2	1.3007
4	3	1.3887
4	4	1.4203

Nonresponse Adjustment of Child Base Weight

The child base weights were adjusted to compensate for the nonresponding sample children. Each of the four input datasets contain all the children who have child status ID equal to 5, 6, 7, or 10, where 5 = ineligible, 6 = enrolled, 7 = declined, and 10 = nonresponse. Only children with child status ID = 6 are enrolled in the study. The eligibility of children with status 10 was unknown for most records; however, for 182 records this could be determined by a subcoded value of child status ID (see table B-3). The weights of the enrolled children were adjusted to account for the unknown eligibility and nonresponse.

Table B-3. Subcodes for child eligibility

Code	Description	Eligibility
1	Received, eligibility status not reported/not known	Unknown
2	Received, eligible case, district could not reach family	Known
3	Received, eligible case, problem not resolved	Known
4	Enrollment form not received	Unknown
5	Enrollment form received late	Unknown

We first tried to use CHAID analysis to define the adjustment cells for the main sample based on the size, region, wealth, age, and placement on the ongoing or historical lists. We found that the stratification variables size, region, and wealth were the most significant predictors of nonresponse. We decided to use the stratification cell as the initial nonresponse adjustment cell.

Since the eligibility of some children was not known, adjustment was done in two stages. First, the nonresponse status was redefined as

Status	Meaning
1	Enrolled
2	Eligible but declined
3	Ineligible
4	Nonresponse, eligibility unknown

In the first stage adjustment, the adjusted weight was $w_{hi}^{**} = w_{hi}^{*} f_{hi}^{NR1}$, where f_{hi}^{NR1} is the factor defined in the table below. S_j is defined as the sum of weights of all cases within each of the nonresponse cells. The nonresponse adjustment factor f_{hi}^{NR1} is then determined depending on the child sample status by:

Status	Adjustment factor
1	$\frac{S_1 + S_2 + S_3 + S_4}{S_1 + S_2 + S_3}$
2	$\frac{S_1 + S_2 + S_3 + S_4}{S_1 + S_2 + S_3}$
3	$\frac{S_1 + S_2 + S_3 + S_4}{S_1 + S_2 + S_3}$
4	0

In the second stage adjustment, the adjusted weight is $w_{hi}^{***} = w_{hi}^{**} f_{hi}^{NR2}$, where the nonresponse adjustment factor f_{hi}^{NR2} is determined as follows:

Status	Adjustment factor

1	$\frac{S_1 + S_2}{S_1}$
2	0
3	1

Truncation of Weight Outliers for Child Base Weights

After nonresponse adjustment, we truncated the weight outliers within five cohorts (A_O, B_O, B_H, C_O, and C_H). This was deemed necessary because the weights vary too much to contain the variance at a reasonable level. Sometimes a simple rule, such as the three-median rule, was used to set truncation of boundary. This rule truncates weights that are larger than three times the median weight to three times the median weight:

$$w_{hi}^{***} = \begin{cases} w_{hi}^{***}, & \text{if } w_{hi}^{***} \le 3\text{Median}, \\ 3\text{Median}, & \text{if } w_{hi}^{***} > 3\text{Median}. \end{cases}$$

However, for some child sampling strata, the three-median rule caused too many weights to be truncated. We tried to keep the percentage of truncated weights to less than 3 percent, so, for some child sampling strata, we used a three-and-a-half-median or four-median rule. For the children who had their full sample weight truncated, all the replicate weights were reduced by the same percentage.

Post-stratification of Enrolled Child Weight

The nonresponse adjusted children's weight was further adjusted by a post-stratification procedure. The control totals for post-stratification contained the number of special education children enrolled by December 2003, by age, for each of the 50 states and the District of Columbia.

Post-stratification was necessary because several states did not have any children sampled, either because, by chance, no LEAs in those states were selected, or none of the selected LEAs in a state responded. It should be noted that the control totals are snapshot figures, while the PEELS population includes children enrolled during a certain time period. The control totals also include children from the very small (size 5) school districts, which were not covered (but were adjusted for) by the PEELS sample.

The post-strata were formed by crossing the three age groups and nine subregions formed by combining states within the same region by their geographical proximity. The size of states in terms of number of children was also taken into consideration in order to obtain similar-sized post-strata.

After the post-stratification was applied, we created the final enrolled children's base weight. This weight is called the children's base weight, although it resulted from various adjustments, because it will be the base for further nonresponse adjustments for different data collection instruments. These are discussed in the following section.

Parent Interview Weights

The parent interview was attempted for all enrolled children, but some parents did not respond. The weights for the parent interview data were created by adjusting the enrolled children's base weights for parent nonresponse. The nonresponse adjustment cells were the same as the ones formed for the nonresponse adjustment to obtain the enrolled children's base weight. This worked well because the response rate for the parent interview was very high. In Waves 1 and 2 at the completion of imputation, parent interview data and corresponding weights were available for 96 percent and 93 percent, respectively, of the children in the augmented sample. Parent interview data and corresponding weights were available for 91 percent of the children in the augmented sample in both waves.

Child Assessment Weights

The child assessment was done in two ways. Most of the children were assessed directly, but for children who could not complete the direct assessment, an alternate assessment was conducted. Together, they represent the whole population of either directly assessable children or unassessable children. The child assessment weight was created by using the enrolled children's weights as base weights and adjusting for child nonresponse in the assessment data. The nonresponse adjustment cells were the same as the ones formed for the nonresponse adjustment to create the enrolled children's base weight. The response rate for child assessment was very high. In Waves 1 and 2 at the completion of imputation, assessment data and corresponding weights were available for 96 percent and 95 percent, respectively, of the children in the augmented sample. Assessment data and corresponding weights were available for 92 percent of the children in the augmented sample in both waves.

Teacher Weights

The teacher interview was attempted for the teachers of all enrolled children, but some teachers did not respond. The weights for the teacher interview data were created by adjusting the enrolled children's base weights for teacher nonresponse. The nonresponse adjustment cells were the same as the ones formed for the nonresponse adjustment to create the enrolled children's base weight. The response rate for teachers was lower than for parents and child assessment. In Waves 1 and 2 at the completion of imputation, teacher interview data and corresponding weights were available for 79 percent and 84 percent, respectively, of the children in the augmented sample. Teacher interview data and corresponding weights were available for 65 percent of the children in the augmented sample in both waves.

Parent-Child Weights

In many analyses, both parent interview and child assessment information are needed; the parent-child weight was for children with both child assessment data and parent interview data. The enrolled children's weights were used as base weights and adjusted for the nonresponse of children in the parent-child data. The nonresponse cells were the same as the ones formed in the nonresponse adjustment for children's base weight. In Waves 1 and 2 at the completion of imputation, both parent interview and assessment data and corresponding weights were available for 92 percent and 89 percent, respectively, of the children in the augmented sample. Child assessment and parent interview data and corresponding weights were available for 85 percent of the children in the augmented sample in both waves.

Parent-Child-Teacher Weights

In some analyses, information from all three instruments is needed. The parent-child-teacher weight is for children with completed interviews for parent interview, child assessment, and the teacher interview. The enrolled children's weights were used as base weights and adjusted for the nonresponse of children in the parent-child data. The nonresponse cells were the same as the ones formed in the nonresponse adjustment for children's base weight. Because of the lower response rate in the teacher interview, the response rate for the parent-child-teacher data is relatively low. In Waves 1 and 2 at the completion of imputation, child assessment, parent interview, and teacher interview data and corresponding weights were available for 70 percent and 76 percent, respectively, of the children in the augmented sample. Child assessment, parent interview, and teacher interview data and corresponding weights were available for 57 percent of the children in the augmented sample in both waves.

Use of Weights in Analysis

Table B-4 provides a description of each weight available after Wave 2 and the analyses for which it is used. For this report, cross-tabulations with covariates from the PEELS demographics file, such as age cohort, sex, and race/ethnicity, use Wave 1 and Wave 2 cross-sectional weights because the demographics file has no missing data and no specific weights. Cross-tabulations with covariates from the Wave 1 files, such as household income, use Wave 1 cross-sectional weights for the Wave 1 cross-tabulations and longitudinal weights for the Wave 2 cross-tabulations because the Wave 2 cross-tabulations use data from Wave 1 and Wave 2 sources. Wave 2 cross-sectional weights were used in table columns with Wave 2 covariates or demographics analyzed with Wave 2 dependent variables.

Table B-4. Description and uses of Wave 1 and Wave 2 cross-source and longitudinal weight variables used in this report

Description	Use of weight
Cross-sectional Wave 1 assessment weight	Analyses using only data from the Wave 1 assessment file
Cross-sectional Wave 2 assessment weight	Analyses using only data from the Wave 2 assessment file
Longitudinal assessment weight	Analyses using only data from the Wave 1 and Wave 2 assessment files
Cross-sectional Wave 1 parent interview weight	Analyses using only data from the Wave 1 parent interview file
Cross-sectional Wave 2 parent interview weight	Analyses using only data from the Wave 2 parent interview file
Longitudinal parent interview weight	Analyses using only data from the Wave 1 and Wave 2 parent interview files
Cross-sectional Wave 1 teacher weight	Analyses using only data from the Wave 1 teacher files
Cross-sectional Wave 2 teacher weight	Analyses using only data from the Wave 2 teacher files
Longitudinal teacher weight	Analyses using only data from the Wave 1 and Wave 2 teacher files
Cross-sectional Wave 1 parent/assessment weight	Analyses using data from the Wave 1 parent interview and Wave 1 assessment files
Cross-sectional Wave 2 parent/assessment weight	Analyses using data from the Wave 2 parent interview and Wave 2 assessment files
Cross-sectional Wave 2 parent/assessment/ teacher weight	Analyses using data from the Wave 2 parent interview, Wave 2 assessment, and Wave 2 teacher files
Longitudinal parent/assessment/teacher weight	Analyses using Wave 1 and Wave 2 data from (1) the parent interview, (2) assessment, or (3) teacher files and Wave 1, Wave 2, or Wave 1 and Wave 2 data from the other types of files (parent interview, assessment, or teacher)
Longitudinal parent/assessment weight	Analyses using Wave 1 and Wave 2 data from the parent interview or assessment files and (1) Wave 1, (2) Wave 2, or (3) Wave 1 and Wave 2 data from the other type of file (assessment or parent interview)
Cross-sectional Wave 1 parent/assessment/ teacher weight	Analyses using data from the Wave 1 parent interview, Wave 1 assessment, and Wave 1 teacher files

Appendix C: Results from PEELS Nonresponse Bias Study

This report presents results of a nonresponse bias analysis of PEELS Wave 1 data. The study was conducted in response to concerns about potential bias from low stage 1 response rates. As a result, terms of clearance for the Pre-Elementary Education Longitudinal Study (PEELS) (OMB #1820-0656) required the U.S. Department of Education's Office of Special Education (OSEP) to submit to the Office of Management and Budget (OMB) a nonresponse analysis report.

To provide the needed confidence to data users, data producers, and study sponsors, OSEP funded a small-scale sample survey of LEAs that initially did not agree to participate in PEELS (464 LEAs or 65 percent of the original LEA sample). Westat selected a random sample of 32 nonparticipating LEAs in Wave 1, allocating the sample to the existing size strata. While 25 of those LEAs agreed to participate, only 23 (72%) actually followed through with their participation, meaning they successfully recruited one or more families[26]. This nonresponse study sample is roughly 10 percent of the size of the main LEA sample. Table C-1 shows the size distribution of the LEAs participating in the nonresponse study.

Table C-1. Frequency of LEAs in PEELS by size stratum and sample type

Size stratum	U.S.	Main sample	Nonresponse sample
Total	7,818	194	23
Very Large	117	33	2
Large	629	32	5
Medium	1,897	43	6
Small	5,175	86	10

The instruments and data collection procedures were exactly the same for the main and nonresponse study participants, so any differences between the two samples can be attributed to the differences in the characteristics of the subpopulations that the samples represent (main study sample and nonresponse study sample).

This nonresponse bias study has three primary research questions. They include the following:

1. Can we produce weighted data from the main sample that provides unbiased national estimates of student performance on key outcome variables?

2. Do statistical differences exist between the performances of students in participating districts and students in nonresponse study districts on key outcome variables?

3. Is student performance on key outcome variables a factor in the decision to participate or not in PEELS?

[26] Nonresponse may cause some bias in estimates obtained from a sample of only respondents if nonrespondents are different from respondents in terms of their characteristics of interest. Nonresponse adjustment weighting was performed so that the bias due to nonresponse is minimized. Even if the nonresponse adjustment weighting was not perfect, bias would not be serious because the response rate of 72 percent is reasonably high.

Methods Used to Analyze Nonresponse Bias

Our general strategy for assessing bias due to nonresponse includes three types of analyses. The first set of analyses involves comparisons between weighted data of the *main* sample versus weighted data of the *amalgamated* sample (which includes the main and nonresponse samples). The second set of analyses compares unweighted data in the main sample with the nonresponse sample. A final set of analyses involves logistic regressions using participation status as the dependent variable and child performance among the independent variables. Each of these analyses is discussed in more detail below.

The amalgamated sample, which includes the main plus nonresponse study samples, with proper weighting, will provide unbiased estimates because the amalgamated sample will represent the entire population. Statistical tests that compare these unbiased estimates and estimates obtained solely from the (weighted) main sample will reveal whether the main sample estimates are significantly different from the unbiased estimates. We will refer to this method as the *amalgamated-main comparison.*

Nonresponse is of less concern if nonrespondents are not systematically different from the respondents in terms of the study variables. The second analysis focuses on this aspect using the super-population framework in which the two samples are assumed to be selected from hypothetical infinite populations of respondents and nonrespondents. This framework enables us to ignore the weights, simplifying the comparison. We performed *t* tests to determine whether the differences between estimates obtained from the unweighted data are significant. This method of comparison is termed the *unweighted comparison.*

The final set of analyses involved a series of logistic regressions in which participation status (main or initial respondents v. initial nonrespondents) was predicted using child age, disability category, and assessment scores. Significant coefficients for the assessment scores would provide evidence for potential bias due to nonresponse for those variables.

It should be noted that a significant difference in the unweighted analysis does not imply that the weighted main sample would be biased for the variable in question. It simply means that bias potential is greater. It is possible to eliminate the bias potential through effective nonresponse adjustment weighting. Therefore, greater emphasis should be given to the results of the amalgamated-main comparison.

Outcome Variables

Wave 1 demographic and direct assessment data were used to analyze nonresponse bias. Among the PEELS data, the direct assessment data are very key, as they will characterize the performance of preschoolers with disabilities and be used to model factors affecting that performance. Further, one might expect children's assessment performances to differ for districts that initially refused to participate in PEELS relative to those that initially accepted the PEELS invitation. Participating children completed a one-on-one assessment of school readiness with a trained assessor. The assessment included the following subtests:

- preLAS 2000 Simon Says, a measure of English/Spanish language ability;
- preLAS 2000 Art Show, a measure of English/Spanish language ability;
- Peabody Picture Vocabulary Test (PPVT), a measure of receptive language ability;
- Woodcock-Johnson III: Letter-Word Identification, a measure of pre-reading skill;

- Woodcock-Johnson III: Applied Problems, a measure of practical math skills;
- Woodcock-Johnson III: Quantitative Concepts-Concepts, a measure of conceptual math skills;
- Woodcock-Johnson III: Quantitative Concepts-Number Series;
- Leiter-R Attention Sustained Scale, a measure of attention;
- Individual Growth and Development Indicators (IGDI): Picture Naming, a measure of pre-reading skills;
- IGDI: Rhyming, a measure of pre-reading skills;
- IGDI: Alliteration, a measure of pre-reading skills;
- IGDI: Segment Blending, a measure of pre-reading skills; and
- Test of Early Math Skills, a measure of general math skills.

The above measures include a combination of performance (achievement) outcomes that we expect to be sensitive to the effects of programs and services that are provided to pre-elementary children and other variables (factors) that may help to explain performance. The PreLAS (Simon Says and Art Show) was used primarily to identify children needing a Spanish-language assessment rather than the Direct Assessment (in English). As such, these two measures were excluded from the nonresponse bias analysis. The PPVT, a measure of receptive language, is not considered to be an achievement measure. It was also excluded from the nonresponse bias analysis. Finally, the Test of Early Math Skills was thought to be largely duplicative of the several Woodcock-Johnson math measures already included in the analysis. Therefore, in order to reduce the complexity of the study, we elected to use only the Woodcock-Johnson measures. Thus, the remaining nine measures were used in the analysis.

Results

In the comparison of main and amalgamated sample estimates of child assessment scores, we assumed that the estimates obtained from the amalgamated sample were unbiased because they were based on the combination of main and nonresponse samples. To address the question of whether the main sample alone, which suffers a high rate of nonresponse, can produce unbiased estimates of the child assessment variables after weighting adjustment for nonresponses, we performed *t* tests on the differences of the estimates obtained from the amalgamated sample and the main sample. If a test result was significant for a variable, we interpreted the result as a piece of evidence to indicate a potential for bias in the main sample estimates for the variable. A non-significant result indicated a lack of such evidence. Tables C-2 through C-4 present the test results for nine outcome performance score variables[27] and eight additional demographic variables, including age, sex, and disability category.

In the following discussion, we will use a 5-percent significance level for all tests. The test results are given in terms of the *p*-value. If a *p*-value is greater than 5 percent, the test result (i.e., the comparison being examined), to which that *p*-value applies, is not statistically significant. Thus, for a comparison

[27] An Attention variable (Leiter-R) was constructed for each age group (3-, 4-, and 5-year-olds). The other eight variables were analyzed using age group as an independent variable.

yielding a *p*-value above 5 percent, the assumption is that there is no statistical difference between those means.

Comparisons Between the Weighted Main and Amalgamated Samples

First, we looked at the sex, age, and disability category distributions as presented in table C-2. The percentage of males in the amalgamated sample is 71.5 percent, which is slightly higher than the main sample estimate of 69.8 percent. The difference is not significant, with a 31.2 percent *p*-value. The percentage of each age group is also not significantly different between the two samples. The *p*-values range from 12.7 to 84.6 percent. No significant differences in individual disability categories were detected either.

Comparison of the two estimates of each score across the age groups is shown in table C-3. Among the 11 variables, only one variable, the WJLWSCORE (Letter-Word), had a significant difference, with a *p*-value of 3.2 percent. All other *p*-values were non-significant. In fact, most results were quite distant from the significance level of 5 percent, with the exception of the WJQCNSCORE (Quantitative Concepts: Number Series) variable, whose *p*-value (6.7 percent) was just over 5 percent.

When the data were analyzed by age group, no differences were significant. The ATTEN variables cannot be analyzed by age because they are already specific to a particular age. Results for these three variables are presented in table C-3. Results for the other assessment-by-age variables are presented in table C-4.

The *t* test results presented here, based on the amalgamated-main comparison, do not indicate any systematic bias in the main sample estimates. Even for the case of the WJLWSCORE (Letter-Word) variable where the overall age comparison yielded a statistically significant result, no significant difference was detected for the comparisons performed within age groups. This provides strong evidence that the main sample is unbiased for the great majority of the assessment variables considered in this study.

Comparisons Between the Unweighted Main and Nonresponse Samples

In the comparison of unweighted means from the main and nonresponse samples, one—WJAPSCORE—of the eight across-age comparisons revealed a significant difference. Among the 8 across-age comparisons and the 18 by-age comparisons, 3 of the by-age results yielded a significant difference—ATTEN4, WJLWSCORE age 4, and WJAPSCORE age 4. These results are provided in detail in tables C-5 and C-6.

While these results in isolation might raise some concerns about possible bias, particularly in cohort B (age 4), it is important to remember that the analyses were unweighted, and weighting is designed to reflect the sampling probability as well as reduce bias due to nonresponse.

Grouped Overall Comparisons

If we look at the results from the view point of overall comparisons, we can make even stronger statements about such comparisons than about individual comparisons. We performed chi-square tests to compare the overall distributions of age and disability. For the age distribution, the difference between the

amalgamated and main samples is strongly insignificant at a *p*-value of 79 percent. Similarly, the difference in the disability distribution in the two samples is insignificant with a *p*-value of 69 percent.

The Bonferroni inequality is often used to perform multiple comparisons. If we perform a family of *t* tests to compare *k* pairs of means with a significance level *a* for each of the *k* individual *t* tests, then the overall significance level (type I error) of the family of *t* tests is at most *ka*. For example, if $k = 10$ and the *ka* is set at 5 percent, then $a = 0.5$ percent.

If we apply this procedure to the result given in table C-3 with an overall significance level of 5 percent, we can say that the differences in the 11 pairs of means are collectively insignificant. We can say the same for the result presented in table C-4 even more forcefully. Furthermore, the Bonferroni procedure enables us to claim that unweighted comparisons shown in tables C-5 and C-6 are not significantly different either in terms of overall comparison.

Logistic Regression Results

Logistic regression analysis was used to examine whether participation status depends on the assessment scores. Dependency indicates possible bias in the score variables. Since the participation status variable is dichotomous, we can examine such dependency using logistic regression, where we use participation status as the dependent variable and assessment scores, disability category, and age as independent variables. By adding age and disability category in the regression models, the dependency is studied by subgroups of age and disability category.

Researchers tried to put as many score variables as possible together in a single model. However, since many score variables are age dependent, we had to limit the age groups permissible in each model. Furthermore, for some scores (e.g., IGDI Alliteration and Rhyming scores), although the tests shared a common age group, we could not estimate the regression coefficients when the tests were placed in a single model. This occurred because the score variables are defined not only based on age but also based on other differing restrictions and this, in turn, created many cases with missing values on one of the score variables. Separate models were developed for those variables. In every model, assessment scores were insignificant predictors of participation status (see tables C-7-A through 7-H).

Conclusions

Based on the three sets of analyses presented here, we conclude that there is little evidence of response bias in the PEELS main sample data. While a few individual comparisons of unweighted data were significantly different, the comparisons of the weighted data were not, in particular when run by age. Furthermore, even those significantly different individual comparisons were not significant as a collective group. This suggests that the weights have eliminated bias in the unweighted main sample. In addition, none of the regressions indicated that assessment scores were significant predictors of participation status. Based on this evidence, we believe no systematic differences exist between the main and nonresponse bias study samples.

Table C-2. Main and amalgamated sample comparison of sex, age, and disability categories

Variable Name	Main		Amalgamated		Difference on main and amalgamated sample est					
	N	est	*N*	est	est	*s.e.*	Lower C.L.	Upper C.L.	*t* test *p*-value	Significant?
SEX_1	2,242	0.698	2,426	0.715	-0.018	0.017	-0.052	0.017	0.312	No
SEX_2	2,242	0.302	2,426	0.285	0.018	0.017	-0.017	0.052	0.312	No
AGE_3	2,242	0.182	2,426	0.194	-0.012	0.008	-0.027	0.003	0.127	No
AGE_4	2,242	0.368	2,426	0.358	0.010	0.013	-0.017	0.036	0.471	No
AGE_5	2,242	0.418	2,426	0.421	-0.003	0.013	-0.028	0.023	0.846	No
DDCAT_1	2,242	0.345	2,426	0.331	0.014	0.032	-0.050	0.077	0.666	No
DDCAT_2	2,242	0.505	2,426	0.491	0.014	0.028	-0.042	0.070	0.622	No
DDCAT_3	2,242	0.030	2,426	0.026	0.004	0.009	-0.014	0.021	0.690	No
DDCAT_4	2,242	0.035	2,426	0.051	-0.016	0.013	-0.042	0.010	0.229	No
DDCAT_5	2,242	0.046	2,426	0.059	-0.012	0.015	-0.043	0.018	0.426	No
DDCAT_6	2,242	0.006	2,426	0.006	0.001	0.003	-0.005	0.006	0.873	No
DDCAT_7	2,242	0.033	2,426	0.037	-0.004	0.010	-0.023	0.016	0.704	No

Table C-3. Main and amalgamated sample comparison of the means of child assessment scores

	Main		Amalgamated		Difference					
Variable Name	***N***	**est**	***N***	**est**	**est**	***s.e.***	**Lower C.L.**	**Upper C.L.**	***t* test *p*-value**	**Significant?**
WJQCCScore	807	7.37	863	7.30	0.06	0.28	-0.49	0.62	0.822	No
WJQCNSScore	807	3.55	863	3.16	0.40	0.22	-0.03	0.82	0.067	No
WJAPScore	2,242	10.38	2,426	10.10	0.29	0.24	-0.18	0.76	0.225	No
WJLWScore	2,239	7.93	2,423	7.50	0.43	0.20	0.04	0.82	0.032	No
IGDIPNScore	2,014	14.70	2,178	15.04	-0.34	0.32	-0.98	0.30	0.296	No
IGDIAScore	720	4.96	775	5.07	-0.11	0.34	-0.77	0.56	0.751	No
IGDIRScore	774	6.55	823	6.67	-0.12	0.49	-1.08	0.84	0.812	No
IGDISBScore	1,562	10.17	1,681	10.69	-0.52	0.52	-1.56	0.51	0.317	No
ATTEN3	533	9.15	586	8.96	0.18	0.31	-0.44	0.81	0.557	No
ATTEN4	859	9.07	930	8.70	0.37	0.25	-0.12	0.86	0.139	No
ATTEN5	776	9.30	826	9.59	-0.29	0.38	-1.05	0.47	0.445	No

Table C-4. Main and amalgamated sample comparison of the means of child assessment scores, by age group

		Main		Amalgamated		Difference					
Variable Name	Age group	*N*	est	*N*	est	est	*s.e.*	Lower C.L.	Upper C.L.	*t* test *p*-value	Significant?
WJAPScore	Age 3	587	5.19	641	5.17	0.01	0.43	-0.83	0.86	0.973	No
	Age 4	848	9.11	922	8.68	0.43	0.41	-0.39	1.24	0.302	No
	Age 5	749	13.28	801	13.19	0.09	0.43	-0.75	0.94	0.825	No
WJLWScore	Age 3	586	4.10	640	4.24	-0.14	0.45	-1.03	0.75	0.756	No
	Age 4	846	5.98	920	5.56	0.42	0.27	-0.12	0.97	0.124	No
	Age 5	749	10.84	801	10.22	0.62	0.42	-0.21	1.45	0.142	No
IGDIPNScore	Age 3	477	10.95	519	11.56	-0.61	0.46	-1.51	0.29	0.183	No
	Age 4	773	13.81	842	13.41	0.40	0.51	-0.60	1.41	0.429	No
	Age 5	711	16.50	760	17.45	-0.94	0.59	-2.10	0.22	0.110	No
IGDIAScore	Age 4	254	3.48	279	3.26	0.22	0.32	-0.40	0.85	0.486	No
	Age 5	426	5.48	454	5.93	-0.45	0.62	-1.66	0.77	0.470	No
IGDIRScore	Age 4	302	5.11	320	4.97	0.14	0.27	-0.38	0.67	0.596	No
	Age 5	431	7.02	459	7.31	-0.30	0.73	-1.73	1.14	0.683	No
IGDISBScore	Age 4	785	7.30	852	7.60	-0.30	0.54	-1.37	0.77	0.579	No
	Age 5	719	12.06	768	12.61	-0.55	0.90	-2.32	1.23	0.545	No

Table C-5. Main and nonresponse sample comparison of the unweighted means of child assessment scores

	Main		Nonresponse		Difference					
Variable Name	***N***	**est**	***N***	**est**	**est**	***s.e.***	**Lower C.L.**	**Upper C.L.**	***t* test *p*-value**	**Significant?**
M_WJQCCScore	807	7.24	56	7.16	0.08	0.450	-0.80	0.96	0.843	No
M_WJQCNSScore	807	3.34	56	2.91	0.43	0.413	-0.38	1.24	0.293	No
M_WJAPScore	2,242	9.68	184	8.50	1.18	0.457	0.29	2.08	0.010	No
M_WJLWScore	2,239	7.10	184	6.29	0.81	0.441	-0.06	1.67	0.064	No
M_IGDIPNScore	2,014	14.50	164	14.61	-0.11	0.509	-1.11	0.89	0.836	No
M_IGDIAScore	720	4.89	55	4.60	0.29	0.559	-0.81	1.39	0.556	No
M_IGDIRScore	774	6.42	49	6.35	0.07	0.680	-1.26	1.40	0.919	No
M_IGDISBScore	1,562	9.91	119	9.90	0.01	0.830	-1.62	1.64	0.989	No
M_ATTEN3	533	9.18	53	8.58	0.59	0.463	-0.32	1.50	0.283	No
M_ATTEN4	859	9.26	71	8.21	1.05	0.439	0.19	1.91	0.009	No
M_ATTEN5	776	9.50	53	9.40	0.10	0.561	-1.00	1.20	0.868	No

Table C-6. Main and nonresponse sample comparison of the unweighted means of child assessment scores, by age

		Main		Nonresponse		Difference					
Variable Name	Age group	*N*	est	*N*	est	est	*s.e.*	Lower C.L.	Upper C.L.	*t* test *p*-value	Significant?
M_WJAPScore	Age 3	587	5.16	54	5.17	-0.01	0.615	-1.21	1.20	0.992	No
	Age 4	848	9.31	74	7.65	1.66	0.610	0.47	2.86	0.009	No
	Age 5	749	13.14	52	12.83	0.31	0.780	-1.22	1.84	0.698	No
M-WJLWScore	Age 3	586	4.03	54	4.04	-0.01	0.539	-1.06	1.05	0.994	No
	Age 4	846	5.99	74	4.96	1.03	0.542	-0.04	2.09	0.035	No
	Age 5	749	10.20	52	10.12	0.08	0.900	-1.68	1.86	0.928	No
M_IGDIPNScore	Age 3	477	10.93	42	11.71	-0.78	0.869	-2.49	0.92	0.324	No
	Age 4	773	14.24	69	13.42	0.82	0.733	-0.62	2.26	0.282	No
	Age 5	711	16.82	49	18.43	-1.61	0.888	-3.35	0.14	0.069	No
M_IGDIAScore	Age 4	254	3.70	25	3.20	0.50	0.621	-0.72	1.72	0.289	No
	Age 5	426	5.41	28	5.75	-0.34	0.847	-2.00	1.32	0.676	No
M_IGDIRScore	Age 4	302	5.13	18	4.67	0.46	0.963	-1.43	2.36	0.587	No
	Age 5	431	7.05	28	7.43	-0.38	0.924	-2.19	1.44	0.706	No
M_IGDISBScore	Age 4	785	7.43	67	7.28	0.15	0.887	-1.59	1.89	0.850	No
	Age 5	719	12.06	49	12.78	-0.72	1.388	-3.44	2.01	0.617	No

Table C-7-A. Logistic regression results for model of Woodcock-Johnson III: Quantitative Concepts scores

HYPOTHESIS TESTING RESULTS: 863 (UNWEIGHTED)

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F	NOTE
OVERALL FIT	0.413	8	114	0.911	
WJQCCScore	1.914	1	121	0.169	
WJQCNSScore	2.436	1	121	0.121	
ddiscat2[7]	0.186	6	116	0.98	

ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB>|T|	COMMENT
INTERCEPT	0.3	1.279	0.237	0.813	
WJQCCScore	-0.11	0.078	-1.384	0.169	
WJQCNSScore	0.13	0.082	1.561	0.121	
ddiscat2.1	-0.13	0.804	-0.158	0.874	
ddiscat2.2	0.06	0.922	0.06	0.952	
ddiscat2.3	0.55	34.731	0.016	0.987	Unstable Standard Error
ddiscat2.4	-0.5	1.351	-0.372	0.711	
ddiscat2.5	0.32	2.068	0.156	0.877	
ddiscat2.6	0.32	32.915	0.01	0.992	Unstable Standard Error

Table C-7-B. Logistic regression results for model of Woodcock Johnson III Letter-Word and Applied Problems scores

HYPOTHESIS TESTING RESULTS: 2178 (UNWEIGHTED)

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F
OVERALL FIT	2.1327	11	111	0.0234
ddiscat2[7]	0.5529	6	116	0.7669
WJLWScore	2.6736	1	121	0.1046
WJAPScore	0.5406	1	121	0.4636
IGDIPNScore	1.4604	1	121	0.2292
CHLDAGE2[3]	0.5636	2	120	0.5707

ESTIMATES FULL REGRESSION COEFFICIENTS

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB>|T|
INTERCEPT	-0.18	1.1105	-0.1638	0.8702
ddiscat2.1	0.16	0.6333	0.2587	0.7963
ddiscat2.2	0.29	0.6419	0.4593	0.6469
ddiscat2.3	-0.13	1.2519	-0.1015	0.9193
ddiscat2.4	-0.73	1.1091	-0.6582	0.5117
ddiscat2.5	-0.27	1	-0.2701	0.7875
ddiscat2.6	0.81	32.9739	0.0245	0.9805
WJLWScore	0.03	0.0208	1.6351	0.1046
WJAPScore	0.03	0.0361	0.7353	0.4636
IGDIPNScore	-0.05	0.0384	-1.2085	0.2292
CHLDAGE2.1	0.14	0.7784	0.1809	0.8568
CHLDAGE2.2	0.35	0.5473	0.635	0.5266

Table C-7-C. Logistic regression results for model of IGDI Alliteration scores

HYPOTHESIS TESTING RESULTS: 775 (UNWEIGHTED)

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F
OVERALL FIT	0.043	5	117	0.999
ddiscat3[4]	0.013	3	119	0.998
CHLDAGE2[2]	0.045	1	121	0.832
IGDIAScore	0.216	1	121	0.643

ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB>|T|
INTERCEPT	0.25	1.955	0.126	0.9
ddiscat3.1	-0.17	1.831	-0.095	0.924
ddiscat3.2	-0.1	1.901	-0.054	0.957
ddiscat3.3	-0.14	2.352	-0.058	0.954
CHLDAGE2.1	-0.14	0.64	-0.213	0.832
IGDIAScore	-0.03	0.07	-0.465	0.643

Table C-7-D. Logistic regression results for model of IGDI Rhyming scores

HYPOTHESIS TESTING RESULTS: 823 (UNWEIGHTED)

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F	NOTE
OVERALL FIT	0.304	5	117	0.91	
ddiscat3[4]	0.201	3	119	0.896	
CHLDAGE2[2]	0.157	1	121	0.693	
IGDIRScore	0.195	1	121	0.66	

ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB>|T|	COMMENT
INTERCEPT	0.59	1.47	0.399	0.691	
ddiscat3.1	-0.11	1.728	-0.066	0.948	
ddiscat3.2	-0.5	1.538	-0.325	0.746	
ddiscat3.3	-0.55	34.21	-0.016	0.987	Unstable Standard Error
CHLDAGE2.1	0.28	0.697	0.396	0.693	
IGDIRScore	-0.03	0.067	-0.442	0.66	

Table C-7-E. Logistic regression results for model of IGDI Segment Blending scores

HYPOTHESIS TESTING RESULTS: 1681 (UNWEIGHTED)

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F
OVERALL FIT	0.639	5	117	0.67
CHLDAGE2[2]	0.076	1	121	0.783
ddiscat3[4]	0.229	3	119	0.876
IGDISBScore	0.441	1	121	0.508

ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB>|T|
INTERCEPT	-0.25	0.794	-0.315	0.753
CHLDAGE2.1	0.15	0.555	0.276	0.783
ddiscat3.1	0.28	0.873	0.32	0.749
ddiscat3.2	0.41	0.771	0.538	0.591
ddiscat3.3	1.28	1.716	0.746	0.457
IGDISBScore	-0.01	0.022	-0.664	0.508

Table C-7-F. Logistic regression results for model of Leiter-R Attention Sustained scores, age 3

HYPOTHESIS TESTING RESULTS: 586 (UNWEIGHTED)

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F
OVERALL				
FIT	0.631	4	118	0.641
ddiscat3[4]	0.515	3	119	0.672
ATTEN3	0.618	1	121	0.433

ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB>|T|
INTERCEPT	-1.58	1.727	-0.915	0.362
ddiscat3.1	0.66	1.35	0.486	0.628
ddiscat3.2	1.19	1.513	0.785	0.434
ddiscat3.3	-0.37	2.354	-0.156	0.876
ATTEN3	0.06	0.073	0.786	0.433

Table C-7-G. Logistic regression results for model of Leiter-R Attention Sustained scores, age 4

HYPOTHESIS TESTING RESULTS: 929 (UNWEIGHTED)

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F
OVERALL				
FIT	1.005	4	118	0.408
ddiscat3[4]	0.426	3	119	0.734
ATTEN4	3.082	1	121	0.082

ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB>|T|
INTERCEPT	-1.59	1.6	-0.991	0.324
ddiscat3.1	0.67	1.476	0.452	0.652
ddiscat3.2	1.1	1.477	0.746	0.457
ddiscat3.3	1.64	1.828	0.898	0.371
ATTEN4	0.1	0.059	1.756	0.082

Table C-7-H. Logistic regression results for model of Leiter-R Attention Sustained scores, age 5

HYPOTHESIS TESTING RESULTS: 829 (UNWEIGHTED)

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F	NOTE
OVERALL					
FIT	0.139	4	118	0.967	
ddiscat3[4]	0.032	3	119	0.992	
ATTEN5	0.459	1	121	0.5	

ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB>|T|	COMMENT
INTERCEPT	0.19	1.104	0.176	0.861	
ddiscat3.1	0.16	0.971	0.169	0.866	
ddiscat3.2	0.27	1.022	0.261	0.795	
ddiscat3.3	0.57	34.718	0.016	0.987	Unstable Standard Error
ATTEN5	-0.04	0.065	-0.677	0.5	

Appendix D: Standard Error Tables

The tables in Appendix D contain standard errors for the corresponding tables in the main body of the report. For example, table D-12 contains the standard errors for table 12.

Table D-12. Standard errors for the percentage of young children who received preschool special education services and had or did not have IEPs in 2003-04 and 2004-05

		2004-05		
		Total	IEP/IFSP	No IEP/IFSP
2003-04	Total		1.1	1.1
	IEP/IFSP	0.7	1.1	1.0
	No IEP/IFSP	0.7	0.4	0.7

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," "Early Childhood Teacher Questionnaire," and "Parent Interview."

Table D-13. Standard errors for the percentage of young children who received preschool special education services during the 2003-04 school year and were declassified in 2004-05

	Declassified
Expected percentage based on the total sample	1.0
Gender	1.2
Male	2.1
Female	1.0
Ethnicity	
Black	1.8
Hispanic	2.3
White	1.3
Family income	
$20,000 or less	2.8
$20,001-$40,000	1.9
More than $40,000	1.6
Metropolitan status	
Urban	2.1
Suburban	1.3
Rural	2.1
District size	
Very large	2.2
Large	2.2
Medium	2.2
Small	1.6
District wealth	
High	1.7
Medium	1.7
Low	2.7
Very low	2.2

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," "Early Childhood Teacher Questionnaire," and "Parent Interview."

Table D-14. Standard errors for the percentage of young children who received preschool special education services during the 2003-04 school year and were declassified in 2004-05, by disability

	Disability									
	Expected percentage based on the total sample	AU	DD	ED	LD	MR	OI	OHI	SLI	LI
Percentage of children in each disability group who were declassified	1.0	0.4	3.7	0.9	0.6	‡	0.9	‡	5.3	2.1

‡ Reporting standards not met.
NOTE: AU = Autism; DD = Developmental delay; ED = Emotional disturbance; LD = Learning disability; MR = Mental retardation; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; LI = Low incidence.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," "Early Childhood Teacher Questionnaire," and "Parent Interview."

Table D-15. Standard errors for the percentage of young children who received preschool special education services during the 2003-04 school year and were declassified in 2004-05, by transition status

	Transition status					
	Expected percentage based on the total sample	Remained in preschool	Transitioned from preschool to kindergarten	Transitioned from kindergarten to first grade	Remained out of school	Other status
Percentage of children in each transition status who were declassified	1.0	1.1	1.4	4.3	8.7	7.2

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," "Early Childhood Teacher Questionnaire," and "Parent Interview."

Table D-16 Standard errors for the mean performance of young children who received preschool special education services during the 2003-04 school year on tests of emerging literacy and early math skills, by eligibility status

	Letter-Word Identification	Applied Problems	PPVT
Total	0.5	0.7	0.6
Remained eligible	0.6	0.7	0.6
Declassified	1.0	1.2	1.3

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Letter-Word Identification," "Woodcock-Johnson III: Applied Problems," "Peabody Picture Vocabulary Test III–R."

Table D-17. Standard errors for the mean performance of young children who received preschool special education services during the 2003-04 and 2004-05 school years on tests of emerging literacy and early math skills, by reclassification status

	Letter-Word Identification	Applied Problems	PPVT
Total	0.7	1.0	0.7
Reclassified	1.5	1.7	1.4
Not reclassified	0.8	1.1	0.7

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Letter-Word Identification," "Woodcock-Johnson III: Applied Problems," "Peabody Picture Vocabulary Test III–R."

Table D-18. Standard errors for the percentage of young children who received preschool special education services whose disability classification remained the same from 2003-04 to 2004-05

Disability classification	
Total	1.2
Autism	2.9
Developmental delay	3.2
Emotional disturbance	9.2
Learning disability	6.8
Mental retardation	5.8
Orthopedic impairment	11.9
Other health impairment	10.2
Speech or language impairment	1.3
Low-incidence disability	5.5

NOTE: Percentages do not include children who were declassified between 2003-04 and 2004-05.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Early Childhood Teacher Questionnaire," "Kindergarten Teacher Questionnaire," "Elementary School Teacher Questionnaire," "Parent Interview" previously unpublished tabulation (August 2006).

Table D-19. Standard errors for the percentage of young children who received preschool special education services who received specific services through their school system, by school year: School years 2003-04 and 2004-05

	2003-04	2004-05
Adaptive physical education	1.1	1.2
Assistive technology services/devices	1.1	1.3
Audiology	1.4	1.2
Augmentative or alternative communication system	1.0	0.8
Behavior management program	1.5	0.9
Learning strategies/study skills assistance	2.0	1.0
Occupational therapy	1.6	1.5
One-to-one paraeducators/assistant	0.8	1.1
Physical therapy	1.2	1.3
Service coordination/case management	2.4	1.0
Social work services	1.0	0.9
Special transportation because of disability	1.4	1.2
Specialized computer software or hardware	1.1	0.6
Speech or language therapy	1.5	1.6
Training, counseling, or other supports/services for family	1.5	0.6
Tutoring/remediation by a special education teacher	1.6	1.1
Other services	1.3	1.0

NOTES: Other services include health services; instruction in American Sign Language, Manual English, Cued Speech, or Braille; mental health services; reader or interpreter; vision services; and other services specified by the respondent. Denominators do not include children who were declassified from special education, so percentages include only children with an IEP.

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," and "Early Childhood Teacher Questionnaire," previously unpublished tabulation (April 2006).

Table D-20. Standard errors for the mean number of services provided to young children who received preschool special education services, by age cohort, gender, and disability: School years 2003-04 and 2004-05

	2003-04	2004-05
Total	0.1	0.1
Age cohort		
Cohort A	0.1	0.
Cohort B	0.1	0.1
Cohort C	0.1	0.1
Gender		
Male	0.1	0.1
Female	0.2	0.2
Disability		
Autism	0.4	0.3
Developmental delay	0.1	0.1
Emotional disturbance	0.5	0.4
Learning disability	0.3	0.3
Mental retardation	0.4	0.3
Orthopedic impairment	0.8	0.5
Other health impairment	0.7	0.5
Speech or language impairment	0.1	0.1
Low-incidence disability	0.6	0.4

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," and "Early Childhood Teacher Questionnaire," previously unpublished tabulation (April 2006).

Table D-21. Standard error for the mean hours per week that young children who received preschool special education services spent in various educational settings, by age cohort and school year: School years 2003-04 and 2004-05

	2003-04	2004-05
Total		
Regular education classroom	0.5	0.5
Special education setting	0.5	0.4
Therapy setting	0.1	#
Non-special education setting outside the classroom for remedial or special assistance	0.2	#
Home instruction	#	#
Cohort A		
Regular education classroom	0.4	0.4
Special education setting	0.5	0.6
Therapy setting	0.1	#
Non-special education setting outside the classroom for remedial or special assistance	#	#
Home instruction	0.1	#
Cohort B		
Regular education classroom	0.6	0.5
Special education setting	0.6	0.5
Therapy setting	0.1	0.1
Non-special education setting outside the classroom for remedial or special assistance	#	#
Home instruction	0.1	0.1
Cohort C		
Regular education classroom	1.0	0.8
Special education setting	0.8	0.4
Therapy setting	0.2	0.1
Non-special education setting outside the classroom for remedial or special assistance	0.4	0.1
Home instruction	0.1	0.1

Rounds to zero

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Elementary School Teacher Questionnaire," "Kindergarten Teacher Questionnaire," and "Early Childhood Teacher Questionnaire," previously unpublished tabulation (April 2006).

Table D-22. Standard error for the percentage of young children who received preschool special education services whose parents were satisfied with special education services to various degrees: School years 2003-2004 and 2004-05

	2003-04	2004-05
Very satisfied	1.9	2.1
Satisfied	1.7	1.7
Dissatisfied	0.7	0.8
Very dissatisfied	0.3	0.3

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Parent Interview," previously unpublished tabulation (April 2006).

Table D-23. Standard errors for the mean performance of young children who received preschool special education services on Woodcock-Johnson III: Letter-Word Identification, by age cohort: School years 2003–04 and 2004-05

	Total	Cohort A	Cohort B	Cohort C
2003-04	0.5	1.1	0.8	0.6
2004-05	0.5	0.8	0.9	0.7

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Letter-Word Identification."

Table D-24. Standard errors for the mean performance of young children who received preschool special education services on Woodcock-Johnson III: Letter-Word Identification, by Wave 1 primary disability: School years 2003-04 and 2004-05

	AU	DD	ED	LD	MR	OI	OHI	SLI	LI
Total									
2003-04	5.1	0.9	3.1	3.1	5.0	2.3	3.8	0.7	1.7
2004-05	5.4	1.0	3.0	2.8	5.6	2.1	4.9	0.6	2.7

NOTE: AU = Autism; DD = Developmental delay; ED = Emotional disturbance; LD = Learning disability; MR = Mental retardation; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; LI = Low incidence.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Letter-Word Identification."

Table D-25. Standard errors for the mean performance of young children who received preschool special education services on Peabody Picture Vocabulary Test III, by age cohort: School years 2003-04 and 2004-05

	Total	Cohort A	Cohort B	Cohort C
2003-04	0.7	0.7	0.7	1.3
2004-05	0.6	0.6	0.8	1.0

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Peabody Picture Vocabulary Test III."

Table D-26. Standard errors for the mean performance of young children who received preschool special education services on Peabody Picture Vocabulary Test III, by Wave 1 primary disability: School years 2003-04 and 2004-05

	AU	DD	ED	LD	MR	OI	OHI	SLI	LI
Total									
2003-04	3.0	1.1	2.6	1.4	4.0	2.6	3.7	0.7	1.8
2004-05	3.8	0.7	1.5	1.9	5.2	3.6	3.0	0.7	1.7

NOTE: AU = Autism; DD = Developmental delay; ED = Emotional disturbance; LD = Learning disability; MR = Mental retardation; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; LI = Low incidence.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Peabody Picture Vocabulary Test III."

Table D-27. Standard errors for the mean performance of young children who received preschool special education services on Woodcock-Johnson III: Applied Problems, by age cohort: School years 2003-04 and 2004-05

	Total	Cohort A	Cohort B	Cohort C
2003-04	0.7	1.1	1.1	1.1
2004-05	0.7	0.9	1.2	1.1

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Applied Problems."

Table D-28. Standard errors for the mean performance of young children who received preschool special education services on Woodcock-Johnson III: Applied Problems, by gender: School years 2003-04 and 2004-05

	Male	Female
Total		
2003-04	0.8	1.1
2004-05	0.7	1.2

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Applied Problems."

Table D-29. Standard errors for the mean performance of young children who received preschool special education services on Woodcock-Johnson III: Applied Problems, by Wave 1 primary disability: School years 2003-04 and 2004-05

	AU	DD	ED	LD	MR	OI	OHI	SLI	LI
Total									
2003-04	4.2	1.2	3.3	2.0	3.6	2.8	5.2	0.8	3.0
2004-05	5.3	0.9	3.5	2.5	4.5	2.7	5.7	0.6	4.1

NOTE: AU = Autism; DD = Developmental delay; ED = Emotional disturbance; LD = Learning disability; MR = Mental retardation; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; LI = Low incidence.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Applied Problems."

Table D-30. Standard errors for the mean performance of young children who received preschool special education services on Woodcock-Johnson III: Quantitative Concepts, by age cohort: School years 2003-04 and 2004-05

	Total	Cohort A	Cohort B	Cohort C
2003-04	0.8	†	†	0.8
2004-05	0.7	†	1.0	0.9

† Not applicable.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III: Quantitative Concepts."

Table D-31. Standard errors for mean teacher ratings of young children who received preschool special education services on the Social Skills subscale of the Preschool and Kindergarten Behavior Scale, by age cohort: School years 2003-04 and 2004-05

	Total	Cohort A	Cohort B	Cohort C
2003-04	0.5	0.9	1.0	1.2
2004-05	0.6	0.9	0.7	1.1

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Social Skills subscale of the Preschool and Kindergarten Behavior Scale."

Table D-32. Standard errors for mean teacher ratings of young children who received preschool special education services on the Social Skills subscale of the Preschool and Kindergarten Behavior Scale, by gender: School years 2003-04 and 2004-05

	Male	Female
Total		
2003-04	0.6	1.6
2004-05	0.6	1.2

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Social Skills subscale of the Preschool and Kindergarten Behavior Scale."

Table D-33. Standard errors for mean teacher ratings of young children who received preschool special education services on the Social Skills subscale of the Preschool and Kindergarten Behavior Scale, by Wave 1 primary disability: School years 2003-04 and 2004-05

	AU	DD	ED	LD	MR	OI	OHI	SLI	LI
Total									
2003-04	3.2	1.2	5.1	4.2	4.4	3.1	5.6	0.7	4.5
2004-05	2.4	1.1	4.5	4.7	4.3	3.4	5.0	0.7	4.7

NOTE: AU = Autism; DD = Developmental delay; ED = Emotional disturbance; LD = Learning disability; MR = Mental retardation; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; LI = Low incidence.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Social Skills subscale of the Preschool and Kindergarten Behavior Scale."

Table D-34. Standard errors for mean teacher ratings of young children who received preschool special education services on the Vineland Adaptive Behavior Scales, Motor Skills Domain, by age cohort: School years 2003-04 and 2004-05

	Total	Cohort A	Cohort B	Cohort C
2003-04	0.9	0.9	1.1	1.6
2004-05	1.0	1.0	1.0	1.6

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Vineland Adaptive Behavior Scales, Motor Skills Domain."

Table D-35. Standard errors for mean teacher ratings of young children who received preschool special education services and participated in an alternate assessment: ABAS II—Skill area scores (Teacher/Daycare Provider Form): School years 2003-04 and 2004-05

	Total
Communication	
2003-04	0.2
2004-05	0.1
Functional (pre) academics	
2003-04	0.1
2004-05	0.3
Health and safety	
2003-04	0.1
2004-05	0.2
Leisure	
2003-04	0.1
2004-05	0.2
Motor	
2003-04	0.2
2004-05	0.4
School living	
2003-04	0.2
2004-05	0.3
Self-care	
2003-04	0.2
2004-05	0.3
Self-direction	
2003-04	0.2
2004-05	0.3
Social	
2003-04	0.1
2004-05	0.2

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "ABAS-II—Skill area scores."

Table D-36. Standard errors for mean teacher ratings of young children who received preschool special education services and participated in an alternate assessment: ABAS II—Skill area scores (Teacher Form): School years 2003-04 and 2004-05

	Total
Communication	
2003-04	0.2
2004-05	0.1
Community use	
2003-04	0.3
2004-05	0.3
Functional academics	
2003-04	0.6
2004-05	0.3
Health and safety	
2003-04	0.5
2004-05	0.2
Leisure	
2003-04	0.3
2004-05	0.3
School living	
2003-04	0.6
2004-05	0.4
Self-care	
2003-04	0.7
2004-05	0.3
Self-direction	
2003-04	0.1
2004-05	0.1
Social	
2003-04	0.3
2004-05	0.2

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "ABAS-II—Skill area scores."

Appendix E: Number of Children Who Had Test Accommodations

Table E-1. Unweighted number of children who had various test accommodations in the PEELS Wave 2 direct assessment, by gender: School year 2004-05

	Male	Female
Abacus	‡	‡
Adaptive furniture	8	4
Communication device	‡	‡
Enlarged print	‡	‡
Familiar person administered test	‡	‡
Familiar person present	62	20
Multiple test sessions	64	21
Person to help child respond	‡	‡
Sign language interpreter	‡	‡
Other	15	3

‡ Reporting standards not met.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Parent Interview."

Table E-2. Unweighted number of children who had various test accommodations in the PEELS Wave 2 direct assessment, by race/ethnicity: School year 2004-05

	Black	Hispanic	White
Abacus	‡	‡	‡
Adaptive furniture	‡	3	7
Communication device	‡	‡	‡
Enlarged print	‡	‡	‡
Familiar person administered test	‡	‡	‡
Familiar person present	6	22	42
Multiple test sessions	9	14	56
Person to help child respond	‡	‡	‡5
Sign language interpreter	‡	‡	‡
Other	3	4	9

‡ Reporting standards not met.
NOTE: Some children who had accommodations are not included in this table because their race/ethnicity is not Black, Hispanic or White.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Parent Interview."

Table E-3. Unweighted number of children who had various test accommodations in the PEELS Wave 2 direct assessment, by primary disability: School year 2004-05

	AU	DD	ED	LD	MR	OI	OHI	SLI	LI
Abacus	‡	‡	‡	‡	‡	‡	‡	‡	‡
Adaptive furniture	‡	3	‡	‡	‡	3	‡	‡	‡
Communication device	‡	‡	‡	‡	‡	‡	‡	‡	‡
Enlarged print	‡	‡	‡	‡	‡	‡	‡	‡	‡
Familiar person present	14	21	‡	‡	5	‡	4	32	3
Multiple test sessions	9	28	‡	‡	3	‡	3	34	6
Person to help child respond	‡	‡	‡	‡	‡	‡	‡	‡	‡
Sign language interpreter	‡	‡	‡	‡	‡	‡	‡	‡	‡
Other	‡	4	‡	‡	3	‡	‡	3	4

‡ Reporting standards not met.
NOTE: AU = Autism; DD = Developmental delay; ED = Emotional disturbance; LD = Learning disability; MR = Mental retardation; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; LI = Low incidence. Some children who had accommodations are not included in this table because they did not have a disability at the time the teacher questionnaire was administered; the teacher questionnaire was the source of the disability variable.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Parent Interview," "Early Childhood Teacher Questionnaire," and "Kindergarten Teacher Questionnaire."

Table E-4. Unweighted number of children who had various test accommodations in the PEELS Wave 2 direct assessment, by age cohort: School year 2004-05

	Cohort A (age 3)	Cohort B (age 4)	Cohort C (age 5)
Abacus	‡	‡	‡
Adaptive furniture	7	3	‡
Communication device	‡	‡	‡
Enlarged print	‡	‡	‡
Familiar person present	40	25	17
Multiple test sessions	26	36	23
Person to help child respond	3	3	‡
Sign language interpreter	‡	‡	‡
Other	4	7	7

‡ Reporting standards not met.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Parent Interview."

Appendix F: Analysis Variables Used Throughout Report

Variable	Source	Response codes
CHILD BACKGROUND AND FAMILY CHARACTERISTICS		
Age cohort	LEA sampling frame and parent interview	1=Cohort A 2=Cohort B 3=Cohort C
Child's sex	Parent interview	1=Male 2=Female
Race/ethnicity	Parent interview	1=Hispanic and of any race 2=Black or African American only, not Hispanic 3=White only and not Hispanic
Household income	Parent interview	1=$20,000 or less 2=$20,001-$40,000 3=More than $40,000
Disability category	Teacher questionnaire	1=Autism 2=Developmental Delay 3=Emotional Disturbance 4=Learning Disability 5=Mental Retardation 6=Orthopedic Impairment 7=Other Health Impairment 8=Speech or Language Impairment 9=Low incidence

Variable	Source	Response codes
SCHOOL /PROGRAM CONTEXT		
District wealth (Percent of district's children living in poverty)	QED sampling frame	1=High Wealth (0-12%) 2=Medium Wealth (13-34%) 3=Low Wealth (35-40%) 4=Very Low Wealth (>40%)
District size (Number of schools within the district)	QED sampling frame	1=Very large (391 or more) 2=Large (118-390) 3=Medium (42-117) 4=Small (41 or less)
Metropolitan status	QED sampling frame	1=Urban (large or mid-sized central city) 2=Suburban (Urban fringe of a large or mid-sized city, large or small town) 3=Rural (population of less than 2500)
Did child receive: adaptive physical education	Teacher questionnaire	1=yes 2=no
Assistive technology services/ devices	Teacher questionnaire	1=yes 2=no
Audiology	Teacher questionnaire	1=yes 2=no
Augmentative communication	Teacher questionnaire	1=yes 2=no

Variable	Source	Response codes
Behavior management program	Teacher questionnaire	1=yes 2=no
Health services	Teacher questionnaire	1=yes 2=no
Instruction in ASL	Teacher questionnaire	1=yes 2=no
Instruction in manual English or cued speech	Teacher questionnaire	1=yes 2=no
Instruction in Braille	Teacher questionnaire	1=yes 2=no
Learning strategies/study skills	Teacher questionnaire	1=yes 2=no
Mental health services	Teacher questionnaire	1=yes 2=no
Occupational therapy	Teacher questionnaire	1=yes 2=no
One-to-one paraeducator assistance	Teacher questionnaire	1=yes 2=no
Physical therapy	Teacher questionnaire	1=yes 2=no
Reader or interpreter	Teacher questionnaire	1=yes 2=no
Service coordination/case management	Teacher questionnaire	1=yes 2=no
Social work services	Teacher questionnaire	1=yes 2=no
Special transportation	Teacher questionnaire	1=yes 2=no
Specialized computer software	Teacher questionnaire	1=yes 2=no
Speech or language therapy	Teacher questionnaire	1=yes 2=no
Training, counseling, and other services to family	Teacher questionnaire	1=yes 2=no
Tutoring/remediation	Teacher questionnaire	1=yes 2=no
Vision services	Teacher questionnaire	1=yes 2=no
Other services	Teacher questionnaire	1=yes 2=no
Number of disability-related services	Teacher questionnaire	continuous, count of services received
Hours per week in regular education classroom	Teacher questionnaire	continuous
Hours per week in special education setting	Teacher questionnaire	continuous
Hours per week in therapy setting	Teacher questionnaire	continuous

Variable	Source	Response codes
Hours per week in nonspecial education setting outside regular class	Teacher questionnaire	continuous
Hours per week in home instruction	Teacher questionnaire	continuous

Variable	Source	Response codes
TRANSITIONS		
Location of enrollment the year before kindergarten	Teacher questionnaire	1=Exact same school and class as now 2=Same school but different kindergarten classroom 3=Not sure 4=Preschool class in same school 5=Some other program or at home
Transition status	Teacher questionnaire	1=No transition between Wave 1 and Wave 2 2=Transition between Wave 1 and Wave 2
OUTCOMES		
IEP/IFSP during previous year	Teacher questionnaire	1=yes 2=no
Declassified between Wave 1 and Wave 2 [must have been eligible at Wave 1]	Teacher questionnaire, missing data filled in using parent report	1=Has IEP/IFSP at both time points 2=IEP/IFSP at Wave 1 and declassified at Wave 2
Does child have an IEP?	Teacher questionnaire	1 = yes, has IEP/IFSP 2 = no IEP/IFSP
PKBS Problem Behaviors Scale	Teacher questionnaire	continuous variable
PKBS Social Skills Scale	Teacher questionnaire	continuous variable
PPVT	Child assessment	continuous variable
WJ Letter-Word Identification	Child assessment	continuous variable
WJ Applied Problems	Child assessment	continuous variable
WJ Quantitative Concepts	Child assessment	continuous variable
Child's academic skills compared to typical children of same grade level	Teacher questionnaire	1=Below or far below average 2= Average 3=Above or far above average