NATIONAL CENTER FOR
SPECIAL EDUCATION RESEARCH
Institute of Education Sciences

A Longitudinal View of the Receptive Vocabulary and Math Achievement of Young Children with Disabilities

NCSER 2011-3006

U.S. DEPARTMENT OF EDUCATION

A Longitudinal View of the Receptive Vocabulary and Math Achievement of Young Children with Disabilities

Authors:
Elaine Carlson
Frank Jenkins
Amy Bitterman
Brad Keller
Westat

Kristen Lauer
Project Officer
Institute of Education Sciences

NCSER 2011-3006 U.S. DEPARTMENT OF EDUCATION

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
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 Kristen Lauer in the National Center for Special Education Research.

August 2011

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., Jenkins, F., Bitterman, A., and Keller, B. (2011). A Longitudinal View of Receptive Vocabulary and Math Achievement of Young Children with Disabilities, (NCSER 2011-3006). U.S. Department of Education. Washington, DC: National Center for Special Education Research.

Content Contact

Kristen Lauer
(202) 219-0377
Kristen.Lauer@ed.gov

Contents

List of Tables

List of Tables (continued)

Page

List of Tables (continued)

List of Tables (continued)

List of Figures

Executive Summary

Reading and math foundational skills that are developed during a child's early years are important as predictors of later academic skills and promoting success at later stages. Vast differences exist in young children's precursor reading and math skills (e.g., Clements 2004; Denton and West 2002; Denton, West, and Walston 2003), and gaps seen in academic performance between groups of young students based on demographic characteristics or initial skill level can persist across grades (Chatterji 2005; LoGerfo, Nichols, and Reardon 2006; Morgan, Farkas, and Wu 2007; Princiotta, Flanagan, and Germino Hausken 2006).

While there are some data documenting the academic performance of older students with disabilities and their typically developing peers from efforts such as state data collected in accordance with the Elementary and Secondary Education Act of 1965 (ESEA), as amended by the No Child Left Behind Act of 2001 (NCLB) and the National Assessment of Educational Progress (NAEP), less is known about the academic skills and skills growth of young children with disabilities.

The Pre-Elementary Education Longitudinal Study (PEELS), which is 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. PEELS data were collected through several different instruments and activities, including direct one-on-one assessments of the children at five points in time.

While several comprehensive reports have been prepared using the PEELS data, this one is designed to address two specific research questions:

- How do children who received preschool special education services perform over time on assessments of receptive vocabulary and math skills?
- How does their receptive vocabulary and math performance vary over time by primary disability category?

Receptive Vocabulary Performance

Psychometrically Adapted and Shortened Version of the Peabody Picture Vocabulary Test III (PPVT-III adapted version)

- At age 3, children in PEELS had a mean score of 61,[1] and at age 10, children had a mean score of 113.

[1] Direct assessments are scored on different scales, so scores on PPVT-III cannot be compared to scores on Woodcock-Johnson III: Applied Problems. To develop the version of the PPVT-III used for PEELS, item response theory (IRT) proficiency scores were put on the publisher's *W*-ability scale through a linking process. As a result, the PPVT-III scores for the PEELS children can be compared to the national norming sample of the publisher (Dunn and Dunn 1997b). The linking procedure for PPVT was refined since the release of other PEELS reports, so comparisons of PPVT scores across PEELS reports should not be made.

- Children's growth on the PPVT-III (adapted version) decelerated, or slowed down, as the children got older, with scores for children at age 3 growing 12.9 points and scores for children at age 10 growing 1.4 points.

- At age 3, children with a speech or language impairment had a significantly higher mean on the PPVT-III (adapted version) than children with a developmental delay.[2] There were no statistically significant differences in growth at age 3 between disability groups, and the gap persisted at age 10 between children with a speech or language impairment and children with a developmental delay.

Math Performance

Woodcock-Johnson III: Applied Problems

- At age 3, children in PEELS had a mean score on Applied Problems of 362, and at age ten, children had a mean score of 488.

- Growth was decelerating, or slowing down, as the children got older, with scores for children at age 3 growing 32.1 points and scores for children at age 10 growing 4.3 points.

- Children with a speech or language impairment had significantly higher mean scores at age 3 than children with autism or a developmental delay. There were no statistically significant differences in growth at age 3 between disability groups. The gap between scores for children with speech or language impairments and children with a developmental delay persisted at age 10. Children with autism caught up to children with a speech or language impairment by age 10.

[2] The disability categories used for these analyses are based on the child's primary disability category in the first wave of data collection. For the purposes of these analyses, children remained in their initial primary disability category even if their classification status changed. Because of the small sample sizes for some disability categories, only the disability categories with sample sizes appropriate for the analyses (set at 40 children or more, which is justified by guidance from Muthén and Muthén 2002) were included: autism, developmental delay, and speech or language impairment.

Chapter 1: Introduction

Achievement in reading and mathematics is largely cumulative, with more advanced skills building on prerequisites (Ehri 1991, 1995; Mazzocco and Myers 2003). Skills developed during a child's early years are important as predictors of later academic skills and success at later stages of life. For example, the National Early Literacy Panel (2009) found that early literacy skills like phonological awareness and alphabet knowledge displayed by young children are predictive of later literacy development. Similarly, a longitudinal study by Storch and Whitehurst (2002) found that preschool children's oral language directly predicts reading comprehension outcomes in fourth grade. For math, children who recognized their basic numbers and shapes and understood the mathematical concept of relative size as they entered kindergarten were more than twice as likely as those who did not to be proficient in addition, subtraction, multiplication, and division by the spring of first grade (Denton and West 2002). Furthermore, entering kindergarteners who recognized their basic numbers and shapes and understood the mathematical concept of relative size were more likely than children who had not acquired these skills to understand ordinality or sequence by the spring of kindergarten and the spring of first grade. One major goal for early reading and mathematics education is developing students' proficiency with those skills needed to master more complicated content, such as reading in the content areas and algebra (Ehri 1995; National Mathematics Advisory Panel 2008).

Vast differences exist in young children's precursor reading and math skills (e.g., Clements 2004; Denton and West 2002; Denton, West, and Walston 2003). For example, variations in vocabulary knowledge (Hart and Risley 1995); ability to recognize letters, sounds, and words (Denton, West, and Walston 2003); count the number of elements in small sets; and carry out simple calculations have been noted in the research literature (Klibanoff et al. 2006). Children who demonstrate difficulties or lower performance in early schooling often have troubles that persist into later grades. For example, a four-year longitudinal study of early elementary school students conducted by Vukovic and Siegel (2010) found that children with persistent math difficulty had weak practical problem solving skills over all four school years, that these children also had low calculation and number fact skills in third and fourth grade, and that poor numbers facts in second grade was an especially strong predictor of persistent math difficulty.

In addition, gaps seen in academic performance between groups of young students based on demographic characteristics or initial skill level can persist across grades. Several studies using data from the Early Childhood Longitudinal Study-Kindergarten cohort (ECLS-K) reported that reading and math scores for kindergarteners from low income households were lower than scores for kindergartners from higher income households, and the gap between the income groups persisted through third grade (Chatterji 2005; LoGerfo et al. 2006; Princiotta, Flanagan, and Germino Hausken 2006). Similarly, tracking the growth trajectories of children with and without math difficulties from kindergarten through fifth grade, Morgan and his colleagues (2007) found that while the mean scores of all groups increased, growth was basically parallel, meaning that those with math difficulties in kindergarten neither caught up to nor fell further behind their peers. Morgan and his colleagues also found that children from families with higher incomes began higher and grew faster than those from families with lower incomes.

While some research documents the academic performance of younger students without disabilities or older students with disabilities using state data collected in accordance with the Elementary and Secondary Education Act of 1965 (ESEA), as amended by the No Child Left Behind Act of 2001 (NCLB), and the National Assessment of Educational Progress (NAEP), less is known about the academic skills and growth in skills of young children with disabilities nationwide. This report tracks over time a single group of young children who received preschool special education services and describes their variation in performance and growth in receptive vocabulary and math skills. The report uses data from the Pre-Elementary Education Longitudinal Study (PEELS). It addresses initial status, final status, and growth for children as they mature from age 3 through 5 to age 8 through 10, a range that includes performance prior to most state data collections and NAEP assessments. It also addresses variation in skills across subgroups of students with disabilities. Specifically, it investigates whether receptive vocabulary and early mathematics skills and growth vary by disability category, one factor previously identified as being associated with academic performance (Blackorby et al. 2005; Center on Education Policy 2009;).

The analyses presented in this report are designed to address two specific research questions:

- How do children who received preschool special education services perform over time on assessments of receptive vocabulary and math skills?
- How does their receptive vocabulary and math performance vary over time by primary disability category?

This is one of several PEELS reports that have been prepared under contract with the National Center for Special Education Research (NCSER) in the U.S. Department of Education's Institute of Education Sciences (IES). Other PEELS reports include the following:

Technical Reports

- *Preschoolers with Disabilities: Characteristics, Services, and Results*;
- *Changes in the Characteristics, Services, and Performance of Preschoolers with Disabilities from 2003-04 to 2004-05*;
- *Early School Transitions and the Social Behavior of Children with Disabilities*; and
- *Access to Educational and Community Activities for Young Children with Disabilities*.

PEELS Progress Notes (2-page briefs)

- *Preschoolers with Disabilities: A Look at School Readiness Skills;*
- *Preschoolers with Disabilities: A Look at Transitions from Preschool to Kindergarten;*
- *Preschoolers with Disabilities: A Look at Parent Involvement;*
- *Preschoolers with Disabilities: A Look at Social Behavior;*
- *Preschoolers with Disabilities: Early Math Performance;*
- *Preschoolers with Disabilities: Reclassification Across Disability Categories;*
- *Young Children with Disabilities: Access to Community Activities; and*
- *Young Children with Disabilities: Access to Educational Activities in Kindergarten.*

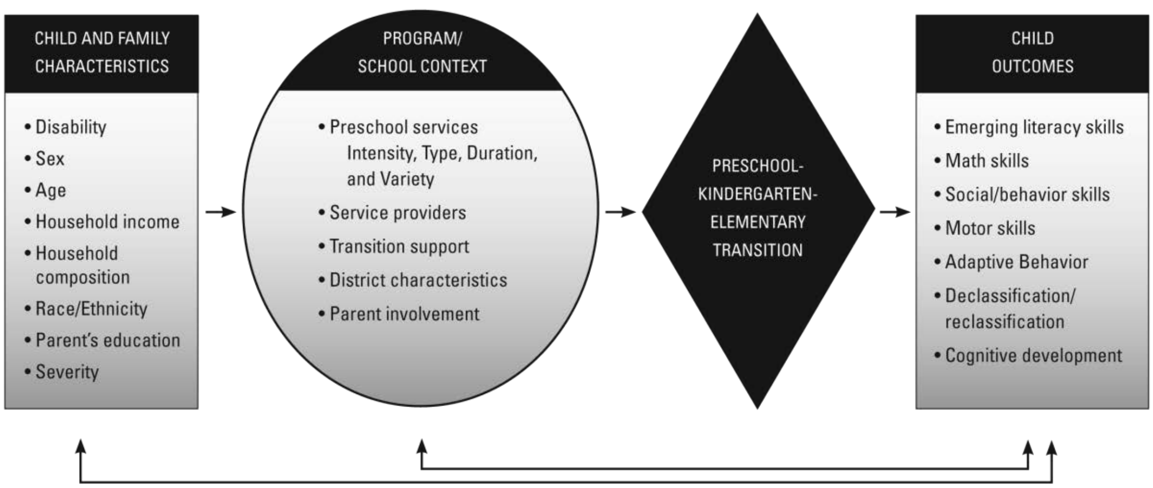
All IES-released reports are available through the project website: ies.ed.gov.

Figure 1 provides an overall model that has guided the PEELS analyses. Five broad descriptive research questions drive 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?

While PEELS is a broad, descriptive, longitudinal study, the analyses presented in this report are more narrowly focused, looking at child characteristics associated with children's emerging receptive vocabulary and math skills over time.

FIGURE 1: OVERALL CONCEPTUAL MODEL FOR PEELS ANALYSIS



This report is organized as follows. Chapter 2 describes the overall PEELS study design and methods relevant to this report. Chapter 3 presents results for growth on the receptive vocabulary and mathematics measures. Appendix A contains a diagram of local education agency (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 describes the number of children who received various test accommodations. Appendix E documents characteristics of the final augmented LEA sample. Appendix F includes likelihood ratio results for the growth models included in chapter 3. Appendix G includes a discussion of cohort effects and the way in which they were addressed.

Appendix H provides a description of the hierarchical linear modeling procedure. For access to PEELS data collection instruments, data tables, and publications, please go to www.peels.org.

Chapter 2: Methods

PEELS is designed to describe children 3 through 5 years of age 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 methods relevant to the results presented in this report.

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]

Different samples are referred to 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 LEAs that initially did not agree to participate in PEELS to examine potential differences between the respondents and nonrespondents.[4] A random sample of 32 initially nonparticipating LEAs in Wave 1 were sampled. While 25 of those LEAs agreed to participate, only 23 actually followed through with their participation, meaning they successfully recruited one or more families. 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,[5] 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.

---

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

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

[5] Data were collected in school years 2003-04, 2004-05, 2005-06, 2006-07, and 2008-09, which are referred to as Wave 1, Wave 2, Wave 3, Wave 4, and Wave 5, respectively.

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.

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 (i.e., more than 25,000 students) 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[6] contacted 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[7] for the region in which the state is located. (This nonparticipating state is referred to 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.[8] By spring 2003, the state education agency (SEA) in state X lifted the ban and allowed its districts to participate in the study. 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.[9] Not all of

---

[6] The authors of this report are among those who conducted the activities described in this chapter (e.g., data collection, imputation, weighting).

[7] 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.

[8] This dropout worsened the response rate among the selected LEAs in the region but did not aggravate the undercoverage problem.

[9] 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.

these LEAs agreed to participate in PEELS, and recruitment of children was very low (approximately 30 percent); therefore, the undercoverage was largely unresolved.

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 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.

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),[10] so only size was used to stratify the 464 nonparticipating LEAs to select the random sample of 32.[11] Twenty-five of those LEAs (78 percent) 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 children in this augmented sample. The Wave 1 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 of children ages 3 through 5 with disabilities because the supplemental sample eliminated the undercoverage issue, and weighting of this sample was done to produce nationally representative estimates for that age group.

A diagram 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

---

[10] 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.

[11] This sample (10% 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.

which were stratified by size only.[12] 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.

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: District wealth was not used as a stratification factor for either the nonresponse bias sample or the supplemental sample, but the counts include the sample LEAs that happened to fall in the respective classes.

NOTE: District size was obtained through the LEA Policies and Practices Questionnaire and was based on report of total district enrollment. Using cutoffs from the National Center for Education Statistics (NCES) Common Core of Data, the districts were categorized as *small* if they had 300-2,500 students, *medium* if they had 2,501-10,000 students, *large* if they had 10,001-25,000 students, and *very large* if they had more than 25,000 students. District wealth was defined as a percentage of the district's children falling below the federal government poverty guidelines, where *high wealth* was 0-12 percent, *medium wealth* was 13-34 percent, *low wealth* was 35-40 percent, and *very low wealth* was more than 40 percent.

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 for which districts identified 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 3 through 5 years of age)—(see table 2 for age eligibility). The second set of lists, called ongoing lists, were submitted monthly for 1 year for which districts identified 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

[12] 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.

ongoing lists late in the 2002-03 school year.[13] 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.

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

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.

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.

To determine the sampling rates for the five child sampling classes in the main sample, district-level sampling weights and district-level child counts by cohort were used. 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. District child counts were obtained 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

[13] Sampling rates were based on district-level enrollment counts for children 3 through 5 years of age with disabilities.

was reselected so that the overall sample size did not exceed 80, and no further ongoing sample selection was performed for the district.[14]

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 ongoing 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).

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. The following five questions on the form asked site coordinators for non-identifying information for each child sampled.

1. Is the child of Hispanic origin?
2. What is the child's race?
3. Is the child in foster care?
4. Does the family receive any kind of public assistance?
5. What is the primary reason for the child's eligibility in preschool special education?

---

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

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 the family 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 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, this set of final recruited families was properly weighted to produce national estimates. Details of the weighting procedure are given in appendix B.

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, so the final tally of the participating districts in the child-based surveys is 223.[15] See appendix E for tables that show participating LEA sample sizes by size of the LEA, region, and wealth. 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.

---

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

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

	Total number of children	Very large	Large	Medium	Small
Total	3,104	736	851	729	788
Cohort A	985	225	256	238	266
Cohort B	1,124	300	323	253	248
Cohort C	995	211	272	238	274

NOTE: District size was obtained through the LEA Policies and Practices Questionnaire and was based on report of total district enrollment. Using cutoffs from the National Center for Education Statistics (NCES) Common Core of Data, the districts were categorized as *small* if they had 300-2,500 students, *medium* if they had 2,501-10,000 students, *large* if they had 10,001-25,000 students, and *very large* if they had more than 25,000 students.

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

	Total number of children	Northeast	Southeast	Central	West/ Southwest
Total	3,104	756	727	658	963
Cohort A	985	287	177	209	312
Cohort B	1,124	261	287	225	351
Cohort C	995	208	263	224	300

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

	Total number of children	High	Medium	Low	Very low
Total	3,104	848	856	796	604
Cohort A	985	292	295	222	176
Cohort B	1,124	301	305	273	245
Cohort C	995	255	256	301	183

NOTE: District wealth was defined as a percentage of the district's children falling below the federal government poverty guidelines, where *high wealth* was 0-12 percent, *medium wealth* was 13-34 percent, *low wealth* was 35-40 percent, and *very low wealth* was more than 40 percent.

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

	Total number of children	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	985	72	328	13	9	23	15	20	443	49	13
Cohort B	1,124	75	280	12	22	30	18	16	590	52	29
Cohort C	995	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 (including deaf/blindness, deafness, hearing impairment, traumatic brain injury, visual impairment, and other disabilities identified by parents but not specified in IDEA (e.g., comprehension problems, hand-eye coordination)).

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

	Total number of children	Male	Female
Total	3,104	2,189	915
Cohort A	985	692	293
Cohort B	1,124	802	322
Cohort C	995	695	300

Data Collection Instruments and Activities

The PEELS design called 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 included 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 was conducted in Wave 5. Additionally, questionnaires were mailed to SEA, LEA, and program/school administrators to obtain contextual information. Table 9 provides response rates for each of the data collection instruments in each wave. Because this report focuses on results from the direct child assessment, it does not include a description of the other data collections. For more information on them, see Markowitz et al. 2006.

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.

Table 9. Total number of respondents for each PEELS instrument

	Wave 1		Wave 2		Wave 3		Wave 4		Wave 5	
Instrument type	Frequency	Response rate	Frequency	Response rate	Frequency	Response rate	Frequency	Response rate	Frequency	Response rate
Parent interview	2,802	96%	2,893	93%	2,719	88%	2,488	80%	--	--
LEA questionnaire	207	84%	--	--	--	--	--	--	--	--
SEA questionnaire	51	100%	--	--	--	--	--	--	--	--
Principal/program director questionnaire[a]	852	72%	665	77%	406	56%	--	--	--	--
Teacher mail questionnaire	2,287	79%	2,591	84%	2,514	81%	2,502	81%	--	--
Early childhood teacher questionnaire	2,018	79%	1,320	86%	346	82%	--	--	--	--
Kindergarten teacher questionnaire	269	73%	957	79%	992	81%	419	79%	--	--
Elementary teacher questionnaire	--	--	314	86%	1176	81%	2083	81%	--	--
Child assessment	2,794	96%	2,932	94%	2,891	93%	2,632	85%	2,520	81%
English/Spanish direct assessment	2,463	97%	2,704	96%	2,726	93%	2,507	85%	2,404	81%
Alternate assessment only	331	93%	228	79%	165	93%	125	84%	116	82%

-- Not available

[a]Quality Education Data (QED) data were used to impute missing items for the principal/program director questionnaires, bringing the percentage of children with some school context information in Waves 1-3 to 94, 95, and 94 percent, respectively.

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 the following 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 that were being used in the Head Start National Reporting System and Head Start Impact Study (HSIS) when the PEELS study was initially designed (www.acf.hhs.gov/programs/opre/ hs/impact_study/index.html). The direct assessment in each wave averaged 40 minutes. Assessments in Waves 1 through 3 included one or more of the following subtests:

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

In Wave 4, the oldest of the PEELS children were 8 years old. Some of the subtests used in Waves 1, 2, and 3 were no longer appropriate for them, and new tests were required to capture their emerging academic skills. PPVT-III (adapted version), Letter-Word Identification, and Applied Problems were retained. The other previous assessments were discontinued and replaced by:

- Woodcock-Johnson III: Calculation (Woodcock, McGrew, and Mather 2001);
- Woodcock-Johnson III: Passage Comprehension (Woodcock, McGrew, and Mather 2001); and
- DIBELS Oral Reading Fluency (Good and Kaminski 2002).

The same subtests used in Wave 4 were used in Wave 5.

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, 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 video-based training only, 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 bi-weekly. 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 percent direct, 12 percent alternate). In Wave 2, a direct or alternate assessment was completed for 94 percent of participating children (87 percent direct, 7 percent alternate). In Wave 3, 93 percent of children completed an assessment (88 percent direct, 5 percent alternate). In Wave 4, 85 percent of children were assessed (81 percent direct, 4 percent alternate). In Wave 5, 81 percent of the children were assessed (77 percent direct, 4 percent alternate).

Description of Assessments. The following is a detailed description of the assessments included in this report, PPVT-III (adapted version) and Applied Problems. To support the growth curve analysis described in this report, we considered only those direct assessments used in all 5 waves: PPVT-III (adapted version), Letter-Word Identification, and Applied Problems. The Letter-Word Identification subtest was excluded because of cohort effects (see Appendix G for more information about cohort effects). For a description of the other assessments, see Markowitz et al. 2006.

Peabody Picture Vocabulary Test III (PPVT-III adapted version). 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.

The standard administration of the PPVT-III involves an assessor showing the child four pictures on a single page then asking the child to point to the picture that matches a word the assessor speaks aloud. For example, the child is shown a page with a picture of a lamp, a wagon, a hoe, and a mop. The child is asked to point to lamp. The child points to one of the pictures; actual articles are not used during administration. If the child points to the correct picture, he or she is given 1 point. Prior to beginning the actual test, the child is given two sets of practice items. If the child correctly completes two consecutive practice items on each set, he or she is administered the actual test. If the child fails to meet the performance criteria, then the test is not administered.

PEELS used a psychometrically adapted and shortened version of the PPVT-III[16]. Due to time constraints associated with the direct assessment, the same test-shortening strategy adopted by the HSIS was used to create a 5-minute version of the PPVT-III for PEELS. This strategy saved approximately 10 minutes of testing time. With the shortened version, all children were presented a core set of items. If their performance on the core set of items was extremely low (responding incorrectly on 8 to 14 of the 14 items in Wave 1, 10 to 16 of the 16 items in Waves 2 and 3, and 14 to 18 of the 18 items in Waves 4 and 5), they were administered an easier basal set of items. If their performance on the core set of items was high (responding incorrectly on 0 to 2 of the 14 items in Wave 1, 0 to 2 of the 16 items in Wave 2, 0 to 3 of the 16 items in Wave 3, 0 to 2 of the 18 items in Wave 4, and 0 to 4 of the 18 items in Wave 5), they were administered a harder ceiling set of items to determine their basic or extended level of performance. PEELS IRT proficiency scores were put on the publisher's *W*-ability scale through a linking process. As a result, the PPVT-III (adapted version) scores for the PEELS children can be compared to the national norming sample of the publisher (Dunn and Dunn 1997b).

The linking procedure for the PPVT-III (adapted version) has been refined since the release of other PEELS reports, so comparisons of PPVT-III (adapted version) scores across reports should not be made. To link PEELS PPVT-III (adapted version) scores to publisher's norms required information on the difficulty and discriminating power of various items. That information was originally taken from an item bank developed through the Head Start Family and Child Experiences Survey (FACES) because no similar item bank was available from the test publisher. After several waves of PEELS analysis, the PPVT-III publisher released an item bank that could be used for the same purpose, so the PEELS PPVT-III (adapted version) data were revised. All PPVT-III (adapted version) data in this report were generated using the same, new linking procedure based on the publisher's item bank.

Woodcock-Johnson III: Applied Problems. The Applied Problems subtest is a measure of children's ability to analyze and solve practical math problems using simple counting, addition, or subtraction operations. 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 (McGrew and Woodcock 2001).

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

[16] The reliability of the adapted version of the PPVT-III used in 2007, for example, was .80. The reported reliability for the full PPVT-III administered to the national norming sample as reported by the test publisher for a comparable age group was .95 (Williams 1997).

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. The assessors were paid $100 for each assessment they completed in Waves 1-3 and $110 for assessments completed in Waves 4 and 5.

If an alternate assessment was required, the assessor gave the Adaptive Behavior Assessment System-II (ABAS-II)[17] 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. In Waves 1-3, the assessor received $50, and the respondent completing the alternate assessment received $50; in Waves 4 and 5, each received $55.

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 Individuals with Disabilities Education Act (IDEA) of 1997, 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,

[17] See Markowitz et al. 2006 for more information on the alternate assessment.

- 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 Waves 1-5 is presented in appendix D. Children who completed English direct assessments with accommodations were included in direct assessment analyses. Their scores were analyzed in the same way as scores for children who did not require accommodations.

Data Preparation and Analysis

This section describes methods used to impute for item and unit nonresponse, develop sampling weights, estimate variance, create independent variables, test for statistical significance, and suppress scarcely populated cells.

Imputation

In data preparation, imputation was conducted for selected items on the child assessment as well as other data collections. For the Wave 1 assessment data, 80 percent of the variables had missing rates of 16 percent or less. Twenty percent of the variables' missing rates were between 24 and 26 percent. In Wave 2, a total of 95 percent of the variables had missing rates less than 2 percent, and 5 percent of the variables had missing rates of 2 to 3 percent. In Wave 3, 90 percent of the variables had missing rates of less than 2 percent. The other 10 percent had missing rates below 3 percent. In Wave 4, all variables had missing rates less than 0.5 percent. In Wave 5, all variables had missing rates less than 0.7 percent. 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 (e.g., 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 data and available information for imputation. The methods included hot-deck imputation, regression, external data source, and deterministic or derivation method[18], 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.

Weighting

The data presented in the report were weighted to generate national estimates. Different weights were 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[19]. Appendix B includes complete information on the weights.

Variance Estimation

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.

Independent Variable

The disability categories used in this data collection were those specified in IDEA. Children's primary disability categories were obtained from their teachers or service providers; however, if service provider data were missing, disability information was obtained from the child's parents. The disability categories used for these analyses are based on the child's primary disability category in the first wave of data collection. For the purposes of these analyses, children remained in their initial primary disability category even if their classification status changed. The disability categories with sufficient sample sizes to stand alone in the analyses[20] were autism, developmental delay, and speech or language impairment.

Trend Analyses

Children in PEELS were tested in five waves, using several different assessments of academic progress. This analysis used two of those assessments: the PPVT-III (adapted version), a measure of

---

[18] The deterministic imputation method imputes a missing value by using the internal relationship of multiple variables within the dataset. For example, if two variables provide complimentary percentages for segments of the population (e.g., percentages of male and female children in a district), and one variable is missing while the other is present, the missing value can be deterministically derived from the other variable (e.g., the percentage in each gender must sum to 100 percent).

[19] There are two types of nonresponses: unit level nonresponse, where the whole questionnaire is missing, and item level nonresponse, where the unit responded but some items were missing. For the latter type of missing values, imputation was used but for the former type, weight adjustment was used.

[20] The threshold set at 40 or more is justified by guidelines from Muthén and Muthén (2002).

receptive vocabulary, and the Woodcock-Johnson III: Applied Problems subtest, a measure of practical math skills. The analysis was designed to describe average growth for subgroups defined by primary disability. As discussed earlier, the PEELS sample includes three overlapping cohorts that entered the study at different ages: Cohort A, starting at age 3; Cohort B, starting at age 4; and Cohort C, starting at age 5. These overlapping cohorts were combined in the analysis using an overlapping cohort design (Raudenbush and Chan 1993). With this design, the initial time point is defined by age, so children of a similar age are compared at each repeated measure. Although each cohort has five waves of data, the combined sample spans eight time points, providing a longer timeframe for modeling average growth curves. Observations are collated by age across cohorts.

By merging the data for the three cohorts, growth over a wider range of ages can be modeled than could be modeled with any one cohort. However, for the merging of cohorts to be valid, an assumption is made that there are no cohort-by-age interactions, implying that the growth profile over ages is the same for all cohorts. To test this assumption, a likelihood ratio test for hierarchical linear growth models was used to determine whether the merged-cohort model (which assumes that there is one growth curve for all cohorts) fits the data as well as the separate cohort model (which assumes that there are cohort-by-growth interactions). The likelihood ratio test of the merged-cohort model used in this study is based on an approach described by Miyazaki and Raudenbush (2000).This test revealed that the merged-cohort model fit the data for PPVT-III (adapted version) and Applied Problems when we used household income as a covariate. Introducing this income indicator variable was adequate for the merged cohort model to fit as well as the separate cohort model.

In selecting covariates, we considered two points: the concomitant reduction in degrees of freedom and possible measurement error in the covariates. To limit the loss in degrees of freedom, we wanted to enter as few covariates as possible while enhancing the model fit. To minimize measurement error, we considered use of demographic variables, which typically demonstrate high reliability. Age was already included as part of the model, and gender was not highly correlated with outcomes in previous PEELS reports (Markowitz et al. 2006, Carlson et al. 2009). Researchers had previously documented the correlation between household income and both PPVT and Applied Problems scores in PEELS (Markowitz et al. 2006) as well as in other studies of children and youth with disabilities (Wagner, Newman, Cameto, and Levine 2006). Household income is a common covariate in studies of educational performance (see, for example, Guarino, Hamilton, Lockwood, and Rathbun 2006; Walston and West 2004). In this analysis, household income enhanced model fit adequately and was used as the sole covariate (see appendix G for a more detailed description of the likelihood ratio tests of the merged and partial cohort models).

Table 10 shows the ages at each of the five waves of observation for each cohort. The dashed lines indicate a year in which no data collection occurred (2008-09). Ages 9 and 10 do not have any consecutive data collections. Also note that for ages 3, 9, and 10, observations come from a single cohort.

Table 10. Children's ages at each of the five waves of PEELS data collection, by cohort

	Age							
Cohort	3	4	5	6	7	8	9	10
A	√	√	√	√	----------	√		
B		√	√	√	√	----------	√	
C			√	√	√	√	----------	√

The longitudinal data were analyzed using three-level hierarchical linear modeling (HLM) (Raudenbush and Bryk 2002; Steele 2008) which models outcomes at several levels of aggregation, with repeated observations over time nested within individuals, and individuals nested within LEAs. HLM has the advantage of explicitly modeling outcomes as they are nested within organizational groupings so the correlations of observations within groups can be accounted for (e.g., the level 3 model, LEA, enables us to account for the clustering effect of children's scores within LEA). In addition, by simultaneously modeling outcomes at different levels of aggregation (i.e., repeated measures within children, children within LEA and LEA), covariates can be added at any level to explain variation in outcomes due to time-specific variables, child characteristics, and group context.

Using the HLM software package and the full information maximum likelihood (FIML) estimator, a growth trajectory was estimated for each individual[21]. In the analyses presented in this paper, linear and quadratic growth, centered at age 3, were estimated and tested for significance using *t*-tests that reflect the clustering of observations in the sample due to the complex sampling design. Linear growth is characterized by the rate of change at age 3. It indicates, on average, how much scores changed as the children got older[22]. However, children's scores do not necessarily grow at a constant rate from year-to-year; so quadratic growth was also estimated. Quadratic growth is curvilinear, with a concave-shaped curve indicating that growth got faster or accelerated as children got older and a convex-shaped curve indicating that growth slowed or decelerated as children got older. The analyses did not investigate whether a linear model or quadratic model fit the data better. Rather, both models are presented to show differences in growth across groups based on disability category. Presenting both types of growth models provides separate perspectives about differences across groups in terms of their average amount of growth and the rate at which children's academic achievement is increasing or decreasing over time.

Likelihood ratio tests were conducted to determine the ability of the models to predict the growth parameters of initial status, linear growth, and quadratic growth based on the inclusion of disability (results from these tests are included in appendix F). If disability was found to be a statistically significant predictor of initial achievement, linear growth, or quadratic growth, *t*-tests were performed to judge whether growth parameters between each pair of disability categories were significantly different from one another (see appendix G for a description of the coding of disability as a predictor).

Study Limitations

The analyses included in this report have several limitations. First, only three of the assessments were administered in all five waves of data collection, and several other assessments not included in this report were administered only once or twice. Therefore, opportunities for growth modeling for a variety of child outcomes were limited. While vertical scaling of different assessments within a content area could have expanded the availability of data across waves, the assessments measured different constructs, even within a general content area (e.g., math); therefore, vertical scaling was not conducted. One of the three assessments given in all five waves was the Woodcock-Johnson III Letter-Word Identification subtest. That assessment was excluded from the analyses because of cohort effects, leaving only two

[21] As seen in table 10, there is systematic missing data for children in a cohort by age. Also, assessment data are missing for some children at a particular wave for various reasons (e.g., child could not be located). HLM allows for missing outcome data and children were not excluded from the analysis unless they had only one wave of assessment data. However, children were excluded if data for the predictor (disability) was missing.

[22] The assessments were not necessarily conducted exactly one year apart. There was a four month window for each wave for data collection, and the timing of the assessments varied across schools and LEAs. The assessment schedules were not incorporated into the analyses.

outcome variables, one for receptive vocabulary and one for math (see appendix G for more information about cohort effects).

A second limitation involves sample size for each of the disability groups. Because PEELS did not oversample by disability category, 90 percent of the analytic sample fell into 3 of the 13 categories: speech or language impairment, developmental delay, and autism. The smaller sample sizes (i.e., below 40 children (Muthén and Muthén 2002)) for the other primary disability categories limited their use in the analyses and are not included in this report (see table 11 for sample sizes by disability). However, the overall analytic sample is large enough that small differences may be significant, even if they are not practically important. Therefore effect sizes are also presented to offset the concern.

Table 11. Number of children in each disability category in analytic sample

Disability													
AU	Deaf	DD	ED	HI	LD	Mild MR	Mod-Sev MR	MD	OI	OHI	SLI	TBI	VI
42	‡	297	17	7	37	10	5	4	17	20	807	‡	‡

‡ Reporting standards not met.
NOTE: AU = Autism; Deaf = Deafness; DD = Developmental delay; ED = Emotional disturbance; HI = Hearing impairment; LD = Learning disability; Mild MR = Mild mental retardation; Mod-Sev MR = Moderate to severe mental retardation; MD = Multiple disabilities; OI = Orthopedic impairment; OHI = Other health impairment; SLI = Speech or language impairment; TBI = Traumatic brain injury; VI = Visual impairment

Chapter 3: The Receptive Vocabulary and Mathematics Performance Over Time of Young Children with Disabilities

This chapter describes the receptive vocabulary and mathematics achievement and growth of young children who received preschool special education services. Skills that are developed during a child's early years are important as predictors of later academic skills and promoting success at later stages. One major goal for early reading and mathematics education is to develop students' proficiency with those skills needed to master more complicated content (Ehri 1995; National Mathematics Advisory Panel 2008).

Previous research has documented sizable differences in young children's early reading and math skills (e.g., Clements 2004; Denton and West 2002; Denton, West, and Walston 2003), and children who demonstrate reading and math difficulties or lower performance in early schooling often have troubles that persist into later grades (Vukovic and Siegel 2010). In addition, gaps in academic performance between subgroups defined by demographic characteristics or initial skill can persist over time (Chatterji 2005; LoGerfo et al. 2006; Morgan et al. 2007; Princiotta, Flanagan, and Germino Hausken 2006).

Although state achievement data and data from NAEP have documented academic performance for older students with disabilities, few data from a nationally representative sample are available on the academic skills and growth of young children with disabilities. This chapter uses longitudinal data from two assessments, one of receptive vocabulary, PPVT-III (adapted version), and one of practical math problem-solving, Woodcock-Johnson III Applied Problems, to address the following specific questions:

- How do children who received preschool special education services perform over time on assessments of receptive vocabulary and math skills?
- How does their receptive vocabulary and math performance vary over time by primary disability category?

Peabody Picture Vocabulary Test-III (PPVT-III Adapted Version)

PEELS used a single measure of receptive vocabulary, the PPVT-III (adapted version). The PPVT-III publisher reported mean W-ability scores for children ages 3 through 10 in the norming sample that averaged 9 percent annual rate of change from age 3 to 10 (see table 12). The rate of change was not consistent over time, however. For example, from age 3 to 4 it increased 11 percent; from age 8 to 9, it increased 3 percent; and from age 9 to 10, it increased 5 percent.[23]

[23] The average annual change was calculated by subtracting the mean at age 3 from the mean at age 10, dividing that result by the mean at age 3 then dividing by the number of comparisons, of which there were seven (e.g., change from age 3 to 4, age 4 to 5, etc.). Also, note that the age-specific means from the publisher's tables are cross-sectional means of different populations of children. These cross-sectional means do not represent growth within age groups. In contrast, the means of the PEELS sample are based on the longitudinal growth of three cohorts of children merged together. The PEELS growth-over-age represents longitudinal growth.

Table 12. Mean W-ability scores and standard deviations for the norming sample on the PPVT-III

Age	*M*	*SD*
3	71	11.3
4	79	10.7
5	87	10.7
6	93	11.4
7	101	12.7
8	107	13.1
9	110	14.3
10	115	21.3

NOTE: *M* = mean; *SD* = standard deviation.
SOURCE: Dunn, L. M., and Dunn, L. M. (1997). *Examiner's Manual for the Peabody Picture Vocabulary Test-Third Edition*. Circle Pines, MN: American Guidance Service.

Table 13 shows the model-based growth components for the PEELS sample. The first three growth parameters in the table (initial status, linear growth and quadratic growth) determine the average growth profile over time for the whole group, as displayed in figure 2[24]. The t-test for each parameter indicates whether the parameter is different from zero. A significant test result indicates that the parameter improves the model fit and therefore should be included as a description of whole group growth. All three growth parameters were significant and were used to define the group means graphed in figure 2. The final status parameter was added to show the model-based mean for children at age 10[25]. The average initial status score at age 3 was 61 (*SE* = 1.3), and the average final status score at age 10 was 113 (*SE* = 0.7). As shown in table 14, children's growth decelerated, or slowed down, as the children got older, with scores for children at age 3 growing 12.9 points and scores for children at age 10 growing 1.4 points. The average for the age-specific growth rates across the 8 year period was 7.1 points. This means children's receptive vocabulary skills, as measured by the PPVT-III (adapted version), increased at a faster rate when they were younger and slowed as children got older. The convex shape of the trend line, as show in figure 2, illustrates that the growth was slowing over time.

[24] Initial status, linear growth, and quadratic growth, are from a model in which age contrasts are centered at age 3.

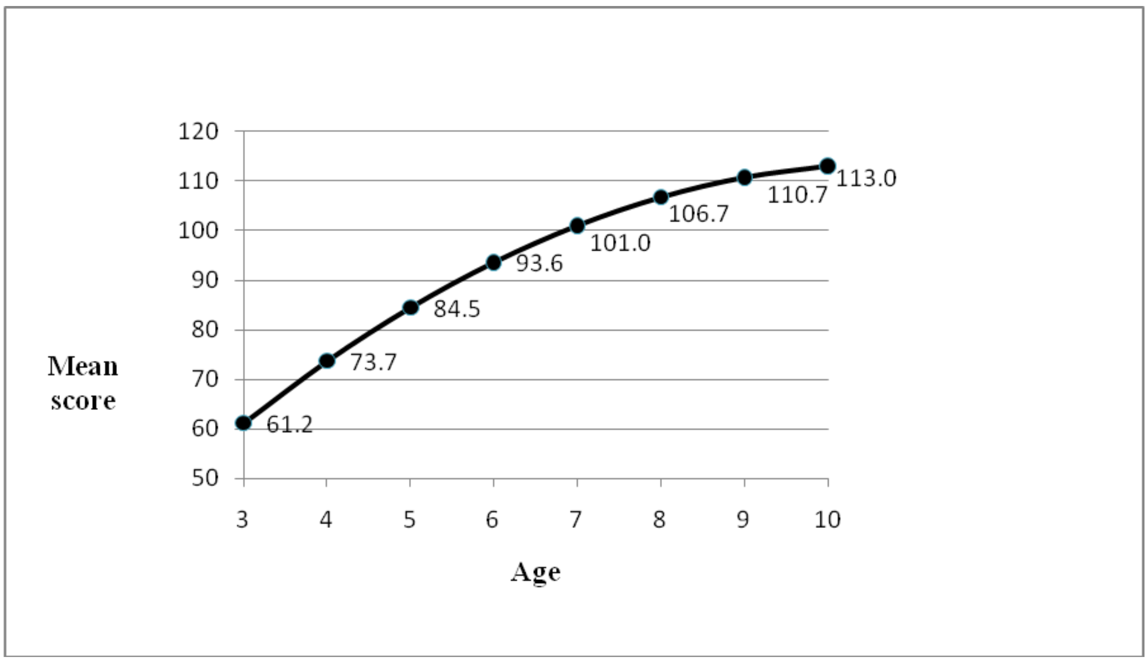
[25] The age 10 final status parameter was from a model in which age contrasts were centered at age 10.

Table 13. Model-based mean initial status at age 3, linear and quadratic growth, and final status at age 10 for young children with disabilities on the PPVT-III (adapted version)

Whole group growth components	Adjusted growth	*SE*	*t*	df	Prob
Initial status	61.17	1.55	39.39	205	**<0.001**
Linear growth	13.38	0.52	25.51	1409	**<0.001**
Quadratic growth	-0.85	0.06	-14.50	1409	**<0.001**
Final status	113.04	0.79	143.73	205	**<0.001**

NOTE: *SE* = standard error, df = degrees of freedom. Items in bold were statistically significant at $p < .05$. The growth effects were adjusted for the covariate, household income. The models were estimated twice, once for initial status with age centered at 3 and once for final status with age centered at 10. The linear growth included in the table is from the model run for initial status.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Peabody Picture Vocabulary Test-III" (December 2010).

Figure 2. Growth curve for children with disabilities on the PPVT-III (adapted version)



SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Peabody Picture Vocabulary Test-III" (December 2010).

Table 14. Amount of growth at each age on the PPVT-III (adapted version)

Age	Age-specific growth	Mean	*SE*	Effect size
3	12.85	61.17	1.55	5.6
4	11.21	73.69	1.18	6.8
5	9.58	84.51	0.94	7.8
6	7.95	93.63	0.80	8.6
7	6.32	101.04	0.72	9.3
8	4.68	106.74	0.69	9.9
9	3.05	110.74	0.72	10.2
10	1.42	113.04	0.79	10.4

NOTE: Effect sizes are the means divided by the standard deviation of the intercept.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Peabody Picture Vocabulary Test-III" (December 2010).

PPVT-III (adapted version) by Disability

Table 15 shows the model-based mean PPVT-III (adapted version) scores for each age, by disability category. At age 3, model-based mean scores were 59 for children with autism, 58 for children with a developmental delay, and 63 for children with a speech or language impairment. At age 10, model-based mean scores were 115 for children with autism, 111 for children with a developmental delay, and 114 for children with a speech or language impairment.

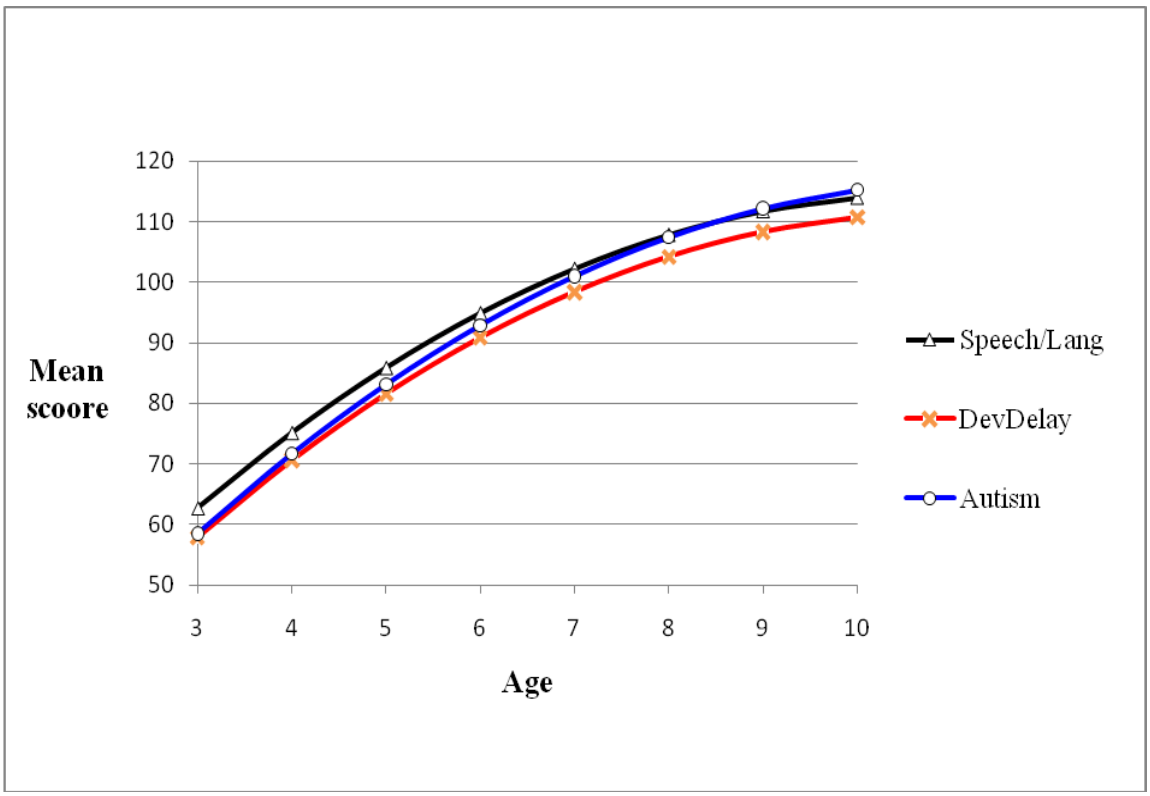
Table 15. Model-based means on the PPVT-III (adapted version), by age and disability category

	Autism			Developmental Delay			Speech or Language Impairment		
Age	Mean	*SE*	Effect size	Mean	*SE*	Effect size	Mean	*SE*	Effect size
3	58.52	3.52	5.4	57.93	1.56	5.3	62.70	1.62	5.8
4	71.67	2.86	6.6	70.62	1.21	6.5	75.13	1.24	6.9
5	83.13	2.40	7.7	81.60	0.98	7.5	85.85	1.00	7.9
6	92.92	2.16	8.6	90.86	0.83	8.4	94.87	0.85	8.8
7	101.02	2.15	9.3	98.40	0.75	9.1	102.18	0.76	9.4
8	107.43	2.33	9.9	104.22	0.72	9.6	107.79	0.72	9.9
9	112.17	2.67	10.4	108.31	0.78	10.0	111.70	0.74	10.3
10	115.23	3.17	10.6	110.69	0.97	10.2	113.90	0.89	10.5

NOTE: Effect sizes are the means divided by the standard deviation of the intercept.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Peabody Picture Vocabulary Test-III" (December 2010).

Likelihood ratio tests were conducted to determine the ability of the models to predict the growth parameters of initial status, linear growth, and quadratic growth based on the inclusion of disability (see Appendix F). The tests indicated that disability was a significant predictor of initial status and linear growth, but not quadratic growth. Figure 3 presents growth trajectories from the modeled PPVT-III (adapted version) means in table 15, and table 16 presents pairwise contrasts between each set of disability categories[26] on initial status, linear growth, quadratic growth, and final status at age 10 (see Appendix F for the likelihood ratio tests conducted to determine the ability of the models to predict the growth parameters of initial status, linear growth, and quadratic growth based on the inclusion of disability). At age 3, children with a speech or language impairment had a significantly higher model-based mean score on the PPVT-III (adapted version) than children with a developmental delay ($t = 4.48$, $p < .05$). While disability overall was a significant predictor of linear growth (see Appendix G), there were no statistically significant differences in the linear growth rates between disability groups at age 3. The analysis of pairwise contrasts between each set of disability categories for quadratic growth was not conducted because disability was not a significant predictor of quadratic growth (see Appendix F). The gap between scores for children with a speech or language impairment and scores for children with a developmental delay ($t = 4.58$, $p < .05$) persisted at age 10. None of the other comparisons by disability category on the PPVT-III (adapted version) were statistically significant.

Figure 3. Growth curve for children with disabilities on the PPVT-III (adapted version), by disability group



SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), “Peabody Picture Vocabulary Test-III” (December 2010).

---

[26] For the pairwise contrasts each model was run twice, once with autism as the reference group and once with speech or language impairment as the reference group.

Table 16. Pairwise contrasts for initial status, growth, and final status on the PPVT-III (adapted version), by disability category

Contrast	Difference in means	*SE*	*t*	Prob	Effect size
Initial status					
Autism – Speech or language	-4.18	2.66	-1.57	0.12	-0.39
Developmental delay – Speech or language	-4.78	1.07	-4.48	**<0.001**	-0.44
Developmental delay – Autism	-0.60	2.77	-0.22	0.83	-0.06
Linear growth					
Autism – Speech or language	0.71	0.70	1.02	0.31	0.07
Developmental delay – Speech or language	0.28	0.20	1.38	0.17	0.03
Developmental delay – Autism	-0.43	0.65	-0.66	0.51	-0.04
Quadratic growth	†	†	†	†	†
Final status					
Autism – Speech or language	1.33	2.70	0.49	0.62	0.27
Developmental delay – Speech or language	-3.21	0.70	-4.58	**<0.001**	-0.65
Developmental delay – Autism	-4.54	2.95	-1.54	0.12	-0.92

† Not applicable; pairwise contrasts were not conducted because disability was not a significant predictor of quadratic growth.
NOTE: Differences in means were adjusted for the covariate, household income. Items in bold were statistically significant at $p <$ .05. The models were estimated twice, once for initial status with age centered at 3 and once for final status with age centered at 10. Pairwise contrasts on linear growth are the pairwise differences of the growth at age 3 between groups. Effect size was calculated by dividing the differences in parameter by the standard deviation of the intercept random effect.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Peabody Picture Vocabulary Test-III" (December 2010).

Woodcock-Johnson III: Applied Problems

PEELS used the Applied Problems subtest as a measure of children's ability to analyze and solve practical math problems using counting, addition, or subtraction operations. On Applied Problems, McGrew and Woodcock (2001) reported mean W-ability scores for the norming sample of children ages 3 through 10 that averaged 5 percent annual growth[27] (see table 17). The rate of change slowed down as children got older. Namely, the rate was 6 percent for ages 3 to 4 and decreased to 2 percent for ages 8 to 9 and ages 9 to 10.

[27] The average annual growth was calculated by subtracting the mean at age 3 from the mean at age 10, dividing that result by the mean at age 3 then dividing the result by the number of comparisons, of which there were seven (e.g., change from age 3 to 4, age 4 to 5, etc.). Also, note that the age-specific means from the publisher's tables are cross-sectional means of different populations of children. These cross-sectional means do not represent growth within age groups. In contrast, the means of the PEELS sample are based on the longitudinal growth of three cohorts of children merged together. The PEELS growth-over-age represents longitudinal growth.

Table 17. Mean W-ability scores and standard deviations for the norming sample on the Woodcock-Johnson Applied Problems subtest

Age	*M*	*SD*
3	375	26.8
4	399	25.5
5	415	24.1
6	441	19.2
7	462	22.8
8	484	22.3
9	494	24.4
10	505	22.5

NOTE: *M* = mean; *SD* = standard deviation
SOURCE: McGrew, K. S., and Woodcock, R. W. (2001). *Technical Manual. Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

Table 18 shows the three model-based growth components for the PEELS sample. Similar to table 13, the first three growth parameters in table 18 (initial status, linear growth, and quadratic growth) determine the average growth profile over time for the whole group, as displayed in figure 4[28]. The t-test for each parameter indicates whether the parameter is different from zero. A significant test result means the parameter improves the model fit and therefore should be included in the description of whole group growth. For Applied Problems, all three growth parameters were significant and were used to define the group means graphed in figure 5. The final status parameter was added to show the model-based mean for the whole group at age 10[29]. The average initial status on the Applied Problems subtest for children at age 3 was 362 (*SE* = 3.1), and the average final status at age 10 was 488 (*SE* = 2.5). The analysis indicated that growth was decelerating, or slowing down, as the children got older, with scores for children at age 3 growing 32.1 points and scores for children at age 10 growing 4.3 points (see table 19). The average for the age-specific growth rates across the 8 year period was 18.2 points. This means that children's ability to analyze and solve math problems, as measured by the Woodcock-Johnson III Applied Problems subtest, grew at a faster rate when children were younger but slowed as the children got older. The convex shape of the trend line, as shown in figure 4, illustrates that the growth was slowing over time.

[28] Initial status, linear growth, quadratic growth, are from a model in which age contrasts are centered at age 3.

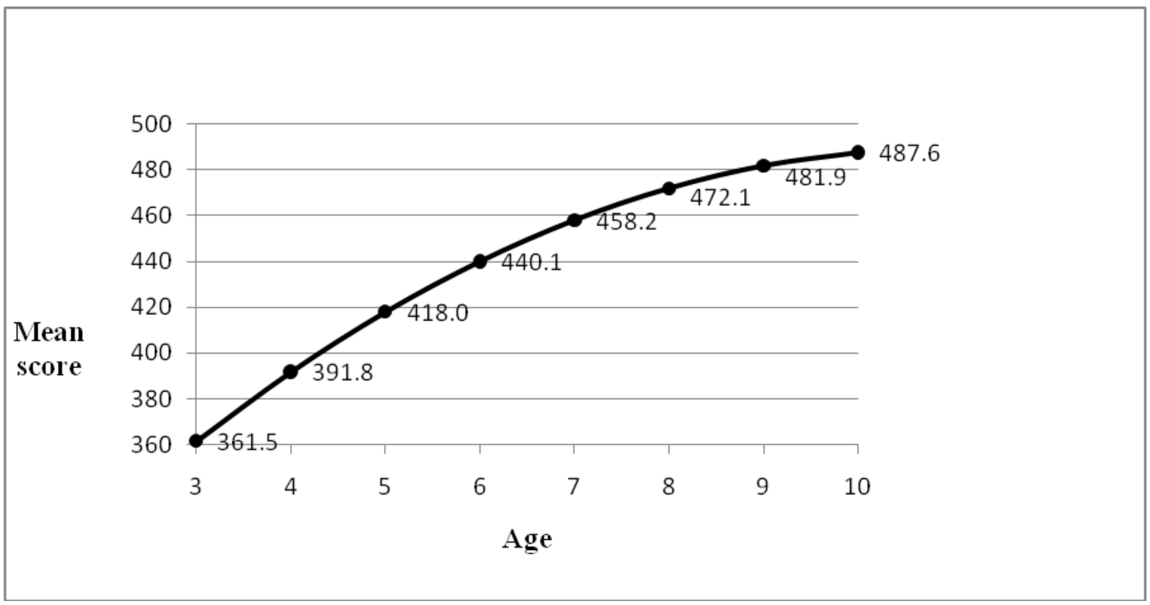
[29] The age 10 final status parameter was from a model in which age contrasts were centered at age 10.

Table 18. Model-based mean initial status at age 3, linear and quadratic growth, and final status at age 10 for young children with disabilities on the Woodcock-Johnson III Applied Problems subtest

Whole group growth components	Adjusted growth	*SE*	*t*	df	Prob
Initial status	361.55	3.89	92.92	205	**<0.001**
Linear growth	32.35	1.77	18.33	1409	**<0.001**
Quadratic growth	-2.05	0.20	-10.04	1409	**<0.001**
Final status	487.60	2.94	165.82	205	**<0.001**

NOTE: *SE* = standard error, df = degrees of freedom. Items in bold were statistically significant at $p < .05$. The growth effects were adjusted for the covariate, household income. The models were estimated twice, once for initial status with age centered at 3 and once for final status with age centered at 10. The linear growth included in the table is from the model run for initial status.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III Applied Problems subtest" (December 2010).

Figure 4. Growth curve for children with disabilities on the Woodcock-Johnson III Applied Problems subtest



SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III Applied Problems" (December 2010).

Table 19. Amount of growth at each age on the Woodcock-Johnson Applied Problems subtest

Age	Age-specific growth	Mean	*SE*	Effect size
3	32.11	361.55	3.89	13.7
4	28.14	391.84	2.83	14.8
5	24.18	418.05	2.42	15.8
6	20.21	440.15	2.45	16.6
7	16.24	458.16	2.58	17.3
8	12.27	472.07	2.68	17.8
9	8.30	481.88	2.79	18.2
10	4.33	487.60	2.94	18.4

NOTE: Effect sizes are the means divided by the standard deviation of the intercept.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III Applied Problems subtest" (December 2010).

Woodcock-Johnson III Applied Problems by Disability

Table 20 presents the Applied Problems model-based means, by age, for the three disability categories: autism, developmental delay, and speech or language impairment. Children with autism, a developmental delay, and a speech or language impairment had mean scores at age 3 of 350, 352, and 366, respectively. At age 10, mean scores for the same subgroups were 492, 472, and 494, respectively.

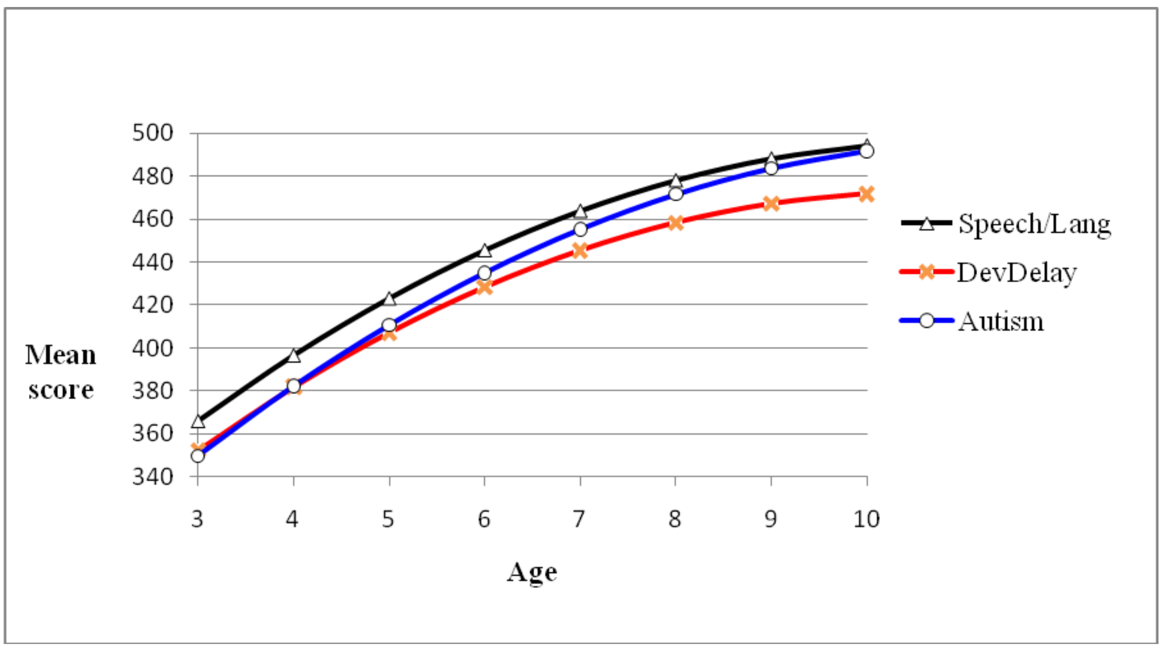
Table 20. Model-based means on the Woodcock-Johnson III Applied Problems subtest, by age and disability category

	Autism			Developmental Delay			Speech or Language Impairment		
Age	Mean	*SE*	Effect size	Mean	*SE*	Effect size	Mean	*SE*	Effect size
3	349.85	8.31	14.2	352.48	3.87	14.3	366.10	4.06	14.9
4	382.29	7.57	15.5	381.88	2.97	15.5	396.68	2.80	16.1
5	410.68	7.41	16.7	407.16	2.69	16.5	423.17	2.18	17.2
6	435.02	7.72	17.7	428.32	2.79	17.4	445.56	2.06	18.1
7	455.30	8.29	18.5	445.37	2.99	18.1	463.87	2.15	18.9
8	471.53	9.02	19.2	458.29	3.17	18.6	478.09	2.24	19.4
9	483.71	9.96	19.7	467.10	3.35	19.0	488.21	2.39	19.8
10	491.84	11.24	20.0	471.78	3.70	19.2	494.25	2.83	20.1

NOTE: Effect sizes are the means divided by the standard deviation of the intercept.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III Applied Problems subtest" (December 2010).

Likelihood ratio tests were conducted to determine the ability of the models to predict the growth parameters of initial status, linear growth, and quadratic growth based on the inclusion of disability (see Appendix F). The tests indicated that disability was a significant predictor of initial status and linear growth, but not quadratic growth. Figure 5 graphs the model-based means in table 20, and table 21 presents the pairwise contrasts between disability groups on initial status at age 3, linear growth, quadratic growth, and final status at age 10. At age 3, children with a speech or language impairment had a model-based mean score on Applied Problems that was significantly higher than the model-based mean score for children with autism ($t = 2.52$, $p < .05$) or a developmental delay ($t = 5.32$, $p < .05$). The difference in initial status between children with a developmental delay and children with autism was not statistically significant. While disability overall was a significant predictor of linear growth, there were no statistically significant differences in the linear growth rates between disability groups at age 3. The analysis of pairwise contrasts between each set of disability categories for quadratic growth was not conducted because disability was not a significant predictor of quadratic growth (see Appendix F). At age 10, scores for children with a speech or language impairment continued to be significantly higher than scores for children with a developmental delay ($t = 7.68$, $p < .05$).

Figure 5. Growth curve for children with disabilities on the Woodcock-Johnson III Applied Problems subtest, by disability category



SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), “Woodcock-Johnson III Applied Problems” (December 2010).

Table 21. Pairwise contrasts for initial status, growth, and final status on the Woodcock-Johnson III Applied Problems subtest, by disability category

Contrast	Difference in means	*SE*	*t*	Prob	Effect size
Initial status					
Autism – Speech or language	-16.25	6.46	-2.52	**0.01**	-0.66
Developmental delay – Speech or language	-13.62	2.56	-5.32	**<0.001**	-0.55
Developmental delay – Autism	2.63	7.04	0.37	0.71	0.11
Linear growth					
Autism – Speech or language	1.84	1.42	1.30	0.19	0.07
Developmental delay – Speech or language	-1.16	0.59	-1.98	0.05	-0.05
Developmental delay – Autism	-3.01	1.72	-1.75	0.08	-0.12
Quadratic growth	†	†	†	†	†
Final status					
Autism – Speech or language	-2.41	9.88	-0.24	0.81	-0.10
Developmental delay – Speech or language	-22.47	2.93	-7.68	**<0.001**	-0.89
Developmental delay – Autism	-20.06	10.96	-1.83	0.07	-0.80

† Not applicable; pairwise contrasts were not conducted because disability was not a significant predictor of quadratic growth.

NOTE: Differences in means were adjusted for the covariate, household income. Items in bold were statistically significant at $p <$ .05. The models were estimated twice, once for initial status with age centered at 3 and once for final status with age centered at 10. Pairwise contrasts on linear growth are the pairwise differences of the growth at age 3 between groups. Effect size was calculated by dividing the differences in parameter by the standard deviation of the intercept random effect.

SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III Applied Problems subtest" (December 2010).

Summary

Children who received preschool special education services showed growth each year on the measure of receptive vocabulary and the measure of math performance; however, children's growth slowed on both measures as they got older. Children's performance varied across assessments and across subgroups defined by disability. As a starting point, 3-year-olds as a group had a model-based mean score of 61 on the PPVT-III (adapted version) and 362 on the Applied Problems subtest. By age 10, as a group the children had a model-based mean score of 113 on the PPVT-III (adapted version) and 488 on the Applied Problems subtest.

On both assessments, initial performance and final status varied by disability category. On PPVT-III (adapted version), at age 3, children with a speech or language impairment had a significantly higher model-based mean than children with a developmental delay. There were no statistically significant differences in growth rates between disability groups at age 3. At age 10, the gap between children with a speech or language impairment and children with a developmental delay persisted. At age 3, children with a speech or language impairment had a model-based mean score on Applied Problems that was significantly higher than children in the other two disability categories. Similar to the PPVT-III (adapted version), there were no statistically significant differences in growth rates between disability groups at age 3, and at age 10, children with a speech or language impairment continued to have significantly higher mean scores than children with a developmental delay.

REFERENCES

Blackorby, J., Wagner, M., Cameto, R., Levine, P., Davies, E., et al. (2005). *Engagement, Academic, Social Adjustment, and Independence: The Achievements of Elementary and Middle School Students With Disabilities.* Menlo Park, CA: SRI International.

Carlson, E., Daley, T., Bitterman, A., Heinzen, H., Keller, B., Markowitz, J., and Riley, J. (2009). *Early School Transitions and the Social Behavior of Children with Disabilities: Selected Findings From the Pre-Elementary Education Longitudinal Study*. Rockville, MD: Westat.

Center on Education Policy. (2009). *State Test Score Trends Through 2007-08, Part 3: Has Progress Been Made in Raising Achievement for Students with Disabilities?* Washington, DC: Author.

Chatterji, M. (2005). Achievement Gaps and Correlates of Early Mathematics Achievement: Evidence From the ECLS K–First Grade Sample. *Education Policy Analysis Archives, 13*(46).

Clements, D. H. (2004). Major Themes and Recommendations. In D. H. Clements, J. Sarama, and A.-M. DiBiase (Eds.), *Engaging Young Children in Mathematics: Standards for Early Childhood Mathematics Education* (pp. 7-72). Mahwah, NJ: Erlbaum.

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

Denton, K., and West. J. (2002). *Children's Reading and Mathematics Achievement in Kindergarten and First Grade*. Washington, DC: U.S. Department of Education.

Denton, K., West, J., and Walston, J. (2003). *Reading – Young Children's Achievement and Classroom Experiences*. Washington, DC: U.S. Department of Education.

Duncan, S. E., and De Avila, E. A. (1998). *preLAS 2000 Examiner's Manual, English Forms C and D.* Monterey, CA: CTB McGraw-Hill.

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

Dunn, L. M., and Dunn, L. M. (1997b). *Examiner's Manual for the Peabody Picture Vocabulary Test-Third Edition*. Circle Pines, MN: American Guidance Service.

ECRI-MGD (Early Childhood Research Institute on Measuring Growth and Development). (2001, 2004). *Individual Growth and Development Indicators*. Minneapolis, MN: Center for Early Education and Development, University of Minnesota.

Ehri, J. C. (1991). Development of the Ability to Read Words. In R. Barr, M. Kamil, P. Mosenthal, and P. Pearson (Eds.), *Handbook of Reading Research* Vol 2. (pp. 383-417). White Plains, NY: Longman.

Ehri, J. C. (1995). Phases of Development in Learning to Read Words by Sight. *Journal of Research in Reading, 18*: 116-125.

Good, R. H., and Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Guarino, C.M., Hamilton, L.S., Lockwood, J.R., and Rathbun, A.H. (2006). *Teacher Qualifications ,Instructional Practices, and Reading and Mathematics Gains of Kindergartners* (NCES 2006-031). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Hart, B., and Risley, R. T. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore: Paul H. Brookes.

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., and Hedges, L. V. (2006). Preschool Children's Mathematical Knowledge: The Effect of Teacher "Math Talk." *Developmental Psychology, 42*(1): 59-69.

LoGerfo, L., Nichols, A., and Reardon, S. (2006). *Achievement Gains in Elementary and High School*. Washington, DC: Urban Institute.

Markowitz, J., Carlson, E., Frey, W., Riley, J., Shimshak, A., et al. (2006). *Preschoolers' Characteristics, Services, and Results: Wave 1 Overview Report from the Pre-Elementary Education Longitudinal Study (PEELS)*. Rockville, MD: Westat. Available at www.peels.org.

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

Mazzocco, M. M., and Myers, G. F. (2003). Complexities in Identifying and Defining Mathematics Learning Disabilities in Primary School Age Years. *Annals of Dyslexia,* 53(1): 218-253.

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

Miyazaki, Y., and Raudenbush, S.W. (2000). Tests for Linkage of Multiple Cohorts in an Accelerated Longitudinal Design. *Psychological Methods*, 5(1): 44-63.

Morgan, P. L., Farkas, G., and Wu, Q. (2007). Five-Year Growth Trajectories of Kindergarten Children with Learning Disabilities in Mathematics. *Journal of Learning Disabilities*, *42*: 305-321.

Muthén, L. K. and Muthén, B. O. (2002). How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(4): 599-620.

National Early Literacy Panel (2009). *Early Beginnings: Early Literacy Knowledge and Instruction*. National Institute for Literacy: Washington, DC.

National Mathematics Advisory Panel (2008). *Foundations for Success: The Final Report of the National Mathematics Advisory Panel.* U.S. Department of Education: Washington, DC.

Princiotta, D., Flanagan, K. D., and Germino Hausken, E. (2006). *Fifth Grade: Findings from the Fifth Grade Follow-up of the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K).* Washington, DC: U.S. Department of Education.

Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods,* 2nd edition. Newbury Park, CA: Sage.

Raudenbush, S., and Chan, W. (1993). Application of a Hierarchical Linear Model to the Study of Adolescent Deviance in an Overlapping Cohort Design, *Journal of Consulting and Clinical Psychology, 61*(6): 941-951.

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

Steele, F. (2008). Multilevel Models for Longitudinal Data. *Journal of the Royal Statistical Society*, *171*(1): 5-19.

Storch, S. A., and Whitehurst, G. J. (2002). Oral Language and Code-Related Precursors to Reading: Evidence From a Longitudinal Structural Model. *Developmental Psychology, 38*: 934-947.

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.

U.S. Department of Education, National Center for Special Education Research (2010). *Preschoolers with Disabilities: A Look at Parent Involvement*. Washington, DC: Author.

U.S. Department of Education, National Center for Special Education Research (2006). *Preschoolers with Disabilities: A Look at School Readiness Skills*. Washington, DC: Author.

U.S. Department of Education, National Center for Special Education Research (2010). *Preschoolers with Disabilities: A Look at Social Behavior*. Washington, DC: Author.

U.S. Department of Education, National Center for Special Education Research (2010). *Preschoolers with Disabilities: A Look at Transitions from Preschool to Kindergarten*. Washington, DC: Author.

U.S. Department of Education, National Center for Special Education Research (2010). *Preschoolers with Disabilities: Early Math Performance*. Washington, DC: Author.

U.S. Department of Education, National Center for Special Education Research (2010). *Preschoolers with Disabilities: Reclassification Across Disability Categories*. Washington, DC: Author.

U.S. Department of Education, National Center for Special Education Research (2010).*Young Children with Disabilities: Access to Community Activities.* Washington, DC: Author.

U.S. Department of Education, National Center for Special Education Research (2010).*Young Children with Disabilities: Access to Educational Activities in Kindergarten.* Washington, DC: Author.

Vukovic, R. K., and Siegel, L. S. (2010). Academic and Cognitive Characteristics of Persistent Mathematics Difficulty from First Through Fourth Grade. *Learning Disabilities Research & Practice, 25*(1): 25-38.

Walston, J.T., and West, J. (2004). *Full-day and Half-day Kindergarten in the United States: Findings from the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99* (NCES 2004–078). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

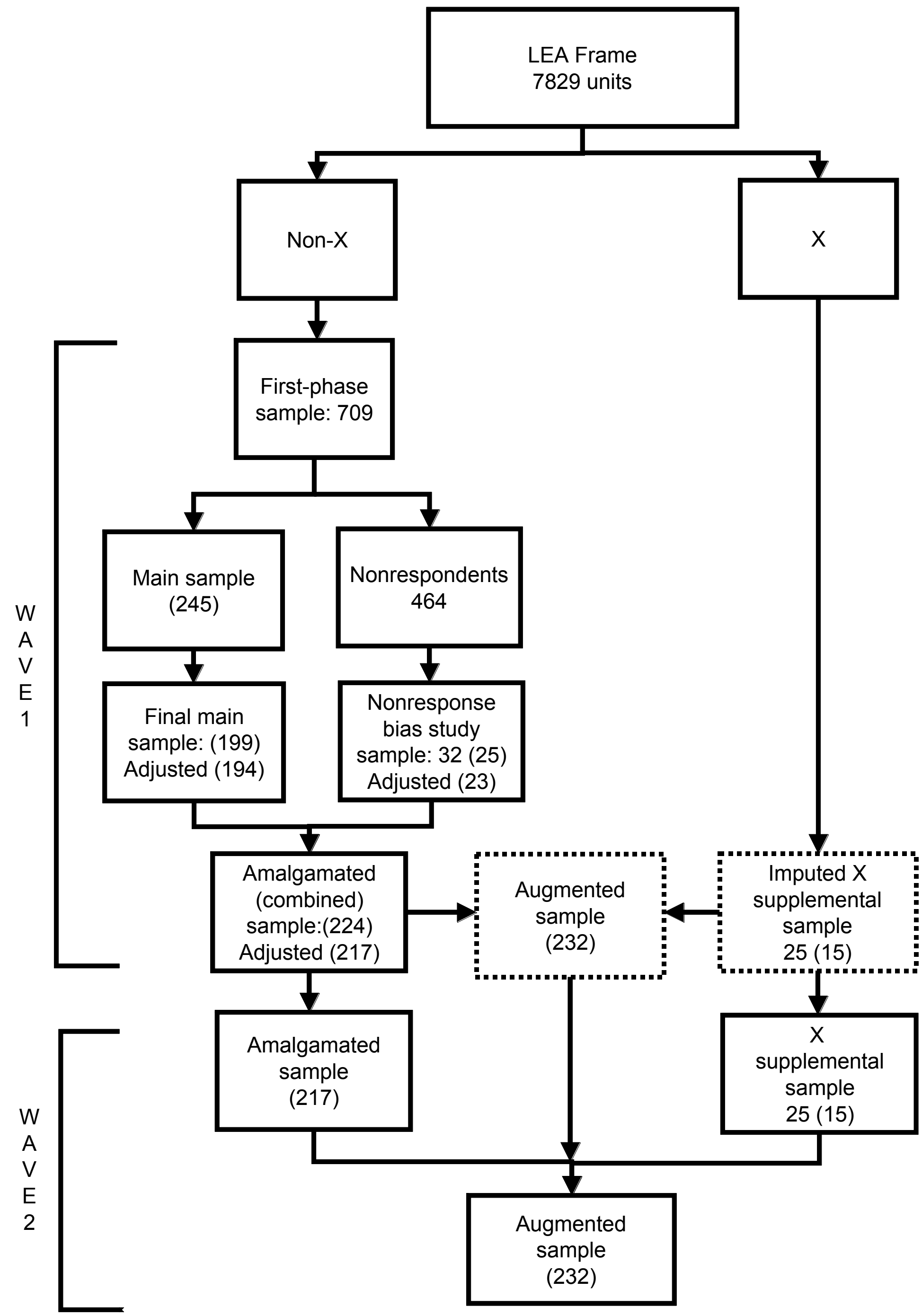
Wagner, M., Newman, L., Cameto, R., and Levine, P. (2006). *The Academic Achievement and Functional Performance of Youth With Disabilities. A Report From the National Longitudinal Transition Study-2 (NLTS2)*. (NCSER 2006-3000). Menlo Park, CA: SRI International.

Westat (2002). *WesVar User Guide*. Rockville, MD: Author.

Williams (1997). *Technical References to the Peabody Picture Vocabulary Test- Third Edition*. Circle Pines, MN: American Guidance Services, Inc.

Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. 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 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 that 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.[1] 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 nonsupplemental 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.[2] Due to restrictions in the Data Analysis 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.

---

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

2 For additional information on WesVar's variance estimation and other technical characteristics, we refer the reader to the documentation in the user's guide (Westat 2002).

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 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

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

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:

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

and

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

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.

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.

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, however, 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 nonparticipation. These and 1, 2, 8, and 12 are treated as not having passed in the sampling system. Status codes 60, 61, and 62 are relevant only to the children in Wave 2.

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 were the real strata from which they were selected. Some children 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. Ninety-six percent of the enrolled children had a parent interview for Wave 1. In Wave 2, a total of 93 percent of parents responded, while 91 percent of the parents responded in both waves. The parent interview response rate was 88 percent in Wave 3, whereas 83 percent of parents responded in all three waves. The corresponding cross-sectional and longitudinal response rates in Wave 4 were 80 percent and 73 percent, respectively.

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. Ninety-six percent of the enrolled children were assessed in Wave 1; a total of 95 percent were assessed in Wave 2, and 92 percent were assessed in both waves. In Wave 3, a total of 93 percent of children were assessed, while 83 percent were assessed in all three waves. In Wave 4, a total of 85 percent were assessed, and 73 percent were assessed in all four waves. In Wave 5, 81 percent were assessed, and 66 percent were assessed in all five 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. Seventy-five percent of the children's teachers responded in Wave 1; a total of 83 percent responded in Wave 2; and a total of 65 percent responded in both waves; 81 percent responded in Wave 3; and 87 percent were included in the teacher longitudinal data with the relaxed condition of responding in any two waves. The Wave 4 cross-sectional response rate was 81 percent, and 80 percent responded in at least three of four 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. A total of 92 percent of the children had both a child assessment and parent interview in Wave 1; a total of 89 percent had both a child assessment and parent interview in Wave 2; a total of 85 percent had both a child assessment and parent interview in both waves; a total of 85 percent had both a child assessment and parent interview in Wave 3; a total of 72 percent had both a child assessment and parent interview in all three waves; 72 percent had both a child assessment and parent interview in Wave 4; and 58 percent had a child assessment and parent interview in all four waves and 53 percent had a child assessment in all five waves and a parent interview in Waves 1-4.

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-teacher 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. Seventy percent of the children had a child assessment, parent interview, and teacher questionnaire in Wave 1; a total of 76 percent had a child assessment, parent interview, and teacher questionnaire in Wave 2; a total of 57 percent had completed instruments for all three in both waves; 72 percent had a child assessment, parent interview, and teacher questionnaire in Wave 3; a total of 65 percent had a child assessment and parent interview in all three waves and teacher responses in any two of three waves; 64 percent had a child assessment, parent interview, and teacher questionnaire in Wave 3; 50 percent had a child assessment and parent interview in all four waves and teacher responses in any three of four waves; and 38 percent had a child assessment in all five waves, a parent interview in Waves 1-4, and teacher responses in any three of four waves.

Use of Weights in Analysis

Table B-4 provides a description of each weight and the analyses for which it is used.

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

Description	Uses
Cross-sectional Wave 1 assessment weight	Analyses using only data from the Wave 1 assessment
Cross-sectional Wave 2 assessment weight	Analyses using only data from the Wave 2 assessment
Cross-sectional Wave 3 assessment weight	Analyses using only data from the Wave 3 assessment
Cross-sectional Wave 4 assessment weight	Analyses using only data from the Wave 4 assessment
Cross-sectional Wave 5 assessment weight	Analyses using only data from the Wave 5 assessment
Longitudinal assessment weight for Waves 1 and 2	Analyses using only assessment data, from Waves 1 and 2
Longitudinal assessment weight for Wave 1, Wave 2, and Wave 3	Analyses using only assessment data, from Waves 1 and 3, or Waves 2 and 3, or all three Waves
Longitudinal assessment weight for Wave 1, Wave 2, Wave 3, and Wave 4	Analyses using only assessment data from Waves 1 and 4, 2 and 4, 3 and 4, or 1, 2, and 4, 1, 3, and 4, 2, 3, and 4, or all four Waves
Longitudinal assessment weight for Wave 1, Wave 2, Wave 3, Wave 4, and Wave 5	Analyses using only assessment data from any combinations of or subsets of Waves 1 through 5
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
Cross-sectional Wave 3 parent interview weight	Analyses using only data from the Wave 3 parent interview file
Cross-sectional Wave 4 parent interview weight	Analyses using only data from the Wave 4 parent interview file
Longitudinal parent weight for Waves 1 and 2	Analyses using only parent file data, from Waves 1 and 2
Longitudinal parent weight for Wave 1, Wave 2, and Wave 3	Analyses using only parent file data, from Waves 1 and 3, or Waves 2 and 3, or all three Waves
Longitudinal parent weight for Wave 1, Wave 2, Wave 3, and Wave 4	Analyses using only parent interview data from Waves 1 and 4, 2 and 4, 3 and 4, or 1, 2, and 4, 1, 3, and 4, 2, 3, and 4, or all four Waves
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
Cross-sectional Wave 3 teacher weight	Analyses using only data from the Wave 3 teacher files
Cross-sectional Wave 4 teacher weight	Analyses using only data from the Wave 4 teacher files
Longitudinal teacher weight for Waves 1 and 2	Analyses using only teacher file data, from Waves 1 and 2
Longitudinal teacher weight for Wave 1, Wave 2, and Wave 3	Analyses using only teacher file data, from Waves 1 and 3, or Waves 2 and 3, or all three Waves
Longitudinal teacher weight for Wave 1, Wave 2, Wave 3, and Wave 4	Analyses using only teacher data from Waves 1 and 4, 2 and 4, 3 and 4, or 1, 2, and 4, 1, 3, and 4, 2, 3, and 4, or all four Waves
Cross-sectional Wave 1 program director/principal weight	Analyses using only data from the Wave 1 program director or principal files

Table B-4. Description and uses of Waves 1-4 cross-source and longitudinal weight variables used in this report (continued)

Description	Uses
Cross-sectional Wave 2 program director/principal weight	Analyses using only data from the Wave 2 program director or principal files
Cross-sectional Wave 3 program director/principal weight	Analyses using only data from the Wave 3 program director or principal files
Cross-sectional Wave 4 program director/principal weight	Analyses using only data from the Wave 4 program director or principal 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 3 parent/assessment weight	Analyses using data from the Wave 3 parent interview and Wave 3 assessment files
Cross-sectional Wave 4 parent/assessment weight	Analyses using data from the Wave 4 parent interview and Wave 4 assessment files
Longitudinal parent/assessment weight for Waves 1 and 2	Analyses using data from parent **and** assessment files, from Waves 1 and 2
Longitudinal parent/assessment weight for Wave 1, Wave 2, and Wave 3	Analyses using data from parent **and** assessment files, from Waves 1 and 3, or Waves 2 and 3, or all three Waves
Longitudinal parent/assessment weight for Wave 1, Wave 2, Wave 3, and Wave 4	Analyses using data from parent **and** assessment files, from Waves 1 and 4, 2 and 4, 3 and 4, or 1, 2, and 4, 1, 3, and 4, 2, 3, and 4, or all four Waves
Longitudinal parent/assessment weight for Wave 1, Wave 2, Wave 3, Wave 4, and Wave 5	Analyses using data from parent and assessment files, from waves 1 and 4, 2 and 4, 3 and 4, 4 and 5 or 1, 2, and 4, 1, 3, and 4, 2, 3, and 4, 3, 4, and 5, or all five waves
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
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
Cross-sectional Wave 3 parent/assessment/teacher weight	Analyses using data from the Wave 3 parent interview, Wave 3 assessment, and Wave 3 teacher files
Cross-sectional Wave 4 parent/assessment/ teacher weight	Analyses using data from the Wave 4 parent interview, Wave 4 assessment, and Wave 4 teacher files
Longitudinal parent/assessment/teacher weight for Waves 1 and 2	Analyses using data from parent, assessment, **and** child files, from Waves 1 and 2
Longitudinal parent/assessment/teacher weight for Wave 1, Wave 2, and Wave 3	Analyses using data from parent, assessment, **and** child files, from Waves 1 and 3, or Waves 2 and 3, or all three Waves
Longitudinal parent/assessment/teacher weight for Wave 1, Wave 2, Wave 3, and Wave 4	Analyses using data from parent, assessment, **and** child files, from Waves 1 and 4, 2 and 4, 3 and 4, or 1, 2, and 4, 1, 3, and 4, 2, 3, and 4, or all four Waves
Longitudinal parent/assessment/teacher weight for Wave 1, Wave 2, Wave 3, Wave 4, and Wave 5	Analyses using data from parent, assessment, and child files, from waves 1 and 4, 2 and 4, 3 and 4, 4 and 5 or 1, 2, and 4, 1, 3, and 4, 2, 3, and 4, 3, 4, and 5, or all five waves

NOTE: Data from the demographics files may be used in conjunction with data from other files without changing the weight.

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 PEELS (OMB #1820-0656) required the U.S. Department of Education's Office of Special Education Programs (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 percent) actually followed through with their participation, meaning they successfully recruited one or more families. 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	‡
Large	629	32	‡
Medium	1,897	43	6
Small	5,175	86	10

‡ Reporting standards not met.

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 are:

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 in PEELS?

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 *combined* 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 combined sample, which includes the main plus nonresponse study samples, with proper weighting, will provide unbiased estimates because the combined 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 *combined-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 involves a series of logistic regressions in which participation status (main or initial respondents v. initial nonrespondents) is predicted using child age, disability category, and assessment scores. Significant coefficients for the assessment scores will 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 combined-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-III adapted version), 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-III (adapted version), 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 combined sample estimates of child assessment scores, we assumed that the estimates obtained from the combined 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 combined sample and the main sample. If a test result was significant for a variable, we interpreted the result as evidence to indicate a potential for bias in the main sample estimates for the variable. A nonsignificant result indicated a lack of such evidence. Tables C-2 through C-4 present the test results for nine outcome performance score variables[1] and eight additional demographic variables, including age, sex, and disability category.

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

---

[1] 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.

Comparisons Between the Weighted Main and Combined Samples

First, we looked at the age and sex distributions and also the distribution of disability categories as presented in table C-2. The combined sample estimate of male percentage is 71.5 percent, which is slightly higher than the main sample estimate of 69.8 percent. The difference is not significant, with 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 statistically 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 nonsignificant. 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 combined-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 statistically 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 of the eight across-age comparisons, WJAPSCORE, revealed a significant difference. Among the eight across-age comparisons and the18 by-age comparisons, three 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 in large part to remove such bias.

Grouped Overall Comparisons

If we look at the results from the viewpoint 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 combined 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.

<u>Logistic Regression Results</u>

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 C-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 combined sample comparison of sex, age, and disability categories

	Main		Combined		Difference on main and combined sample est					
Variable name	***N***	**est**	***N***	**est**	**est**	***SE***	**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

NOTE: *N* = number of cases in the full sample; est = estimate; *SE* = standard error; C.L. = confidence level; SEX = Child's gender; AGE = Child's age; and DDCAT = Child's disability.

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

	Main		Combined		Difference					
Variable name	***N***	**est**	***N***	**est**	**est**	***SE***	**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

NOTE: *N* = number of cases in the full sample; est = estimate; *SE* = standard error; C.L. = confidence level; WJQCCScore = Woodcock-Johnson Quantitative Concepts: Concepts; WJQCNSScore = Woodcock-Johnson Quantitative Concepts: Number Series; WJAPScore = Woodcock-Johnson Applied Problems,;WJLWScore = Woodcock-Johnson Letter Word; IGDIPNScore = Individual Growth and Development Indicators: Picture Naming; IGDIAScore = Individual Growth and Development Indicators: Alliteration; IGDIRScore = Individual Growth and Development Indicators: Rhyming; IGDISBScore = Individual Growth and Development Indicators: Segment Blending; ATTEN3 = Leiter-R: Attention Sustained, age 3; ATTEN4 = Leiter-R Attention Sustained, age 4; and ATTEN5 = Leiter-R Attention Sustained, age 5.

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

		Main		Combined		Difference					
Variable name	Age group	*N*	est	*N*	est	est	*SE*	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

NOTE: *N* = number of cases in the full sample; est = estimate; *SE* = standard error; C.L. = confidence level; WJAPScore = Woodcock-Johnson Applied Problems,;WJLWScore = Woodcock-Johnson Letter Word; IGDIPNScore = Individual Growth and Development Indicators: Picture Naming; IGDIAScore = Individual Growth and Development Indicators: Alliteration; IGDIRScore = Individual Growth and Development Indicators: Rhyming; IGDISBScore = Individual Growth and Development Indicators: Segment Blending;

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

	Main		Nonresponse		Difference of main and nonresponse sample est					
Variable name	***N***	**est**	***N***	**est**	**est**	***SE***	**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

NOTE: *N* = number of cases in the full sample; est = estimate; *SE* = standard error; C.L. = confidence level; WJAPScore = Woodcock-Johnson Applied Problems,;WJLWScore = Woodcock-Johnson Letter Word; IGDIPNScore = Individual Growth and Development Indicators: Picture Naming; IGDIAScore = Individual Growth and Development Indicators: Alliteration; IGDIRScore = Individual Growth and Development Indicators: Rhyming; IGDISBScore = Individual Growth and Development Indicators: Segment Blending; ATTEN3 = Leiter-R: Attention Sustained, age 3; ATTEN4 = Leiter-R Attention Sustained, age 4; and ATTEN5 = Leiter-R Attention Sustained, age 5.

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**	***SE***	**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

NOTE: *N* = number of cases in the full sample; est = estimate; *SE* = standard error; C.L. = confidence level; WJAPScore = Woodcock-Johnson Applied Problems,;WJLWScore = Woodcock-Johnson Letter Word; IGDIPNScore = Individual Growth and Development Indicators: Picture Naming; IGDIAScore = Individual Growth and Development Indicators: Alliteration; IGDIRScore = Individual Growth and Development Indicators: Rhyming; and IGDISBScore = Individual Growth and Development Indicators: Segment Blending

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

NOTE: Num = number; df = degrees of freedom; HO = null hypothesis; Denom = denominator; WJQCCScore = Woodcock-Johnson Quantitative Concepts: Concepts; WJQCNSScore = Woodcock-Johnson Quantitative Concepts: Number Series; and ddiscat = Child's disability.

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

NOTE: Num = number; df = degrees of freedom; HO = null hypothesis; Denom = denominator; WJLWScore = Woodcock-Johnson Letter Word; WJAPScore = Woodcock-Johnson Applied Problems IGDIPNScore = Individual Growth and Development Indicators: Picture Naming; CHLDAGE = child's age; and ddiscat = Child's disability.

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

NOTE: Num = number; df = degrees of freedom; HO = null hypothesis; Denom = denominator; IGDIAScore = Individual Growth and Development Indicators: Alliteration; CHLDAGE = child's age; and ddiscat = Child's disability.

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	

NOTE: Num = number; df = degrees of freedom; HO = null hypothesis; Denom = denominator; IGDIRScore = Individual Growth and Development Indicators: Rhyming; CHLDAGE = child's age; and ddiscat = Child's disability.

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

NOTE: Num = number; df = degrees of freedom; HO = null hypothesis; Denom = denominator; IGDISBScore = Individual Growth and Development Indicators: Segment Blending; CHLDAGE = child's age; and ddiscat = Child's disability.

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

NOTE: Num = number; df = degrees of freedom; HO = null hypothesis; Denom = denominator; ATTEN3 = Leiter-R Attention Sustained, age 3; and ddiscat = Child's disability.

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

NOTE: Num = number; df = degrees of freedom; HO = null hypothesis; Denom = denominator; ATTEN4 = Leiter-R Attention Sustained, age 4; and ddiscat = Child's disability.

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	Num. 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	

NOTE: Num = number; df = degrees of freedom; HO = null hypothesis; Denom = denominator; ATTEN5 = Leiter-R Attention Sustained, age 5; and ddiscat = Child's disability.

APPENDIX D: NUMBER OF CHILDREN WHO HAD TEST ACCOMMODATIONS

Table D-1. Unweighted number of children who had various test accommodations in the PEELS Wave 1 direct assessment, by gender: School year 2003-04

	Male	Female
Abacus	‡	‡
Adaptive furniture	11	8
Communication device	6	3
Enlarged print	‡	‡
Familiar person administered test	‡	‡
Familiar person present	125	49
Multiple test sessions	68	33
Person to help child respond	10	4
Sign language interpreter	‡	‡
Other	10	4

‡ Reporting standards not met.
NOTE: This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-2. Unweighted number of children who had various test accommodations in the PEELS Wave 1 direct assessment, by race/ethnicity: School year 2003-04

	Black	Hispanic	White
Abacus	‡	‡	‡
Adaptive furniture	5	‡	11
Communication device	4	‡	3
Enlarged print	‡	‡	‡
Familiar person administered test	‡	‡	‡
Familiar person present	38	10	115
Multiple test sessions	12	7	72
Person to help child respond	4	‡	8
Sign language interpreter	‡	‡	‡
Other	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. This table includes children in Cohorts A, B, and C.SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-3. Unweighted number of children who had various test accommodations in the PEELS Wave 1 direct assessment, by primary disability: School year 2003-04

	AU	DD	ED	LD	MR	OI	OHI	SLI	LI
Abacus	‡	‡	‡	‡	‡	‡	‡	‡	‡
Adaptive furniture	‡	7	‡	‡	‡	6	‡	‡	4
Communication device	‡	‡	‡	‡	‡	‡	‡	‡	6
Enlarged print	‡	‡	‡	‡	‡	‡	‡	‡	‡
Familiar person administered test	‡	‡	‡	‡	‡	‡	‡	‡	‡
Familiar person present	16	48	‡	7	5	‡	5	76	8
Multiple test sessions	6	37	3	3	‡	4	‡	36	8
Person to help child respond	3	‡	‡	‡	‡	‡	‡	7	‡
Sign language interpreter	‡	‡	‡	‡	‡	‡	‡	‡	‡
Other	‡	4	‡	‡	‡	‡	‡	7	‡

‡ Reporting standards not met.
NOTE: AU = Autism; DD = ; 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. This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-4. Unweighted number of children who had various test accommodations in the PEELS Wave 1 direct assessment, by age cohort: School year 2003-04

	Cohort A (age 3)	Cohort B (age 4)	Cohort C (age 5)
Abacus	‡	‡	‡
Adaptive furniture	4	9	6
Communication device	‡	‡	6
Enlarged print	‡	‡	‡
Familiar person administered test	‡	‡	‡
Familiar person present	58	65	51
Multiple test sessions	35	39	27
Person to help child respond	3	3	8
Sign language interpreter	‡	‡	‡
Other	5	3	6

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

Table D-5. 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.
NOTE: This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-6. 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. This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-7. 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 = ; 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. This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-8. 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 4)	Cohort B (age 5)	Cohort C (age 6)
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).

Table D-9. Unweighted number of children who had various test accommodations in the PEELS Wave 3 direct assessment, by gender: School year 2005-06

	Male	Female
Abacus	‡	‡
Adaptive furniture	10	6
Communication device	6	‡
Enlarged print	3	‡
Familiar person administered test	‡	‡
Familiar person present	38	7
Multiple test sessions	28	10
Person to help child respond	5	‡
Sign language interpreter	‡	‡
Other	16	6

‡ Reporting standards not met.
NOTE: This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-10. Unweighted number of children who had various test accommodations in the PEELS Wave 3 direct assessment, by race/ethnicity: School year 2005-06

	Black	Hispanic	White
Abacus	‡	‡	‡
Adaptive furniture	‡	4	11
Communication device	‡	‡	5
Enlarged print	‡	‡	‡
Familiar person administered test	‡	‡	‡
Familiar person present	5	14	25
Multiple test sessions	‡	9	26
Person to help child respond	‡	‡	‡
Sign language interpreter	‡	‡	3
Other	‡	6	11

‡ 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. This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-11. Unweighted number of children who had various test accommodations in the PEELS Wave 3 direct assessment, by Wave 1 primary disability: School year 2005-06

	AU	DD	ED	LD	MR	OI	OHI	SLI	LI
Abacus	‡	‡	‡	‡	‡	‡	‡	‡	‡
Adaptive furniture	‡	4	‡	‡	‡	5	4	‡	‡
Communication device	‡	‡	‡	‡	‡	‡	‡	‡	5
Enlarged print	‡	‡	‡	‡	‡	‡	‡	3	‡
Familiar person administered test	‡	‡	‡	‡	‡	‡	‡	‡	‡
Familiar person present	12	7	‡	‡	3	‡	3	13	5
Multiple test sessions	6	12	‡	‡	‡	‡	‡	15	3
Person to help child respond	‡	3	‡	‡	‡	‡	‡	5	6
Sign language interpreter	‡	‡	‡	‡	‡	‡	‡	‡	3
Other	3	4	‡	‡	‡	‡	‡	5	6

‡ Reporting standards not met.
NOTE: AU = Autism; DD = ; 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. This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-12. Unweighted number of children who had various test accommodations in the PEELS Wave 3 direct assessment, by age cohort: School year 2005-06

	Cohort A (5 years old)	Cohort B (6 years old)	Cohort C (7 years old)
Abacus	‡	‡	‡
Adaptive furniture	11	4	‡
Communication device	‡	3	3
Enlarged print	‡	‡	‡
Familiar person administered test	‡	‡	‡
Familiar person present	18	13	14
Multiple test sessions	10	15	13
Person to help child respond	4	‡	‡
Sign language interpreter	‡	‡	‡
Other	8	6	8

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

Table D-13. Unweighted number of children who had various test accommodations in the PEELS Wave 4 direct assessment, by gender: School year 2006-07

	Male	Female
Abacus	‡	‡
Adaptive furniture	7	5
Communication device	6	3
Enlarged print	4	5
Familiar person administered test	‡	‡
Familiar person present	30	15
Multiple test sessions	12	4
Person to help child respond	10	‡
Sign language interpreter	‡	4
Other	16	6

‡ Reporting standards not met.
NOTE: This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-14. Unweighted number of children who had various test accommodations in the PEELS Wave 4 direct assessment, by race/ethnicity: School year 2006-07

	Black	Hispanic	White
Abacus	‡	‡	‡
Adaptive furniture	‡	‡	8
Communication device	‡	‡	6
Enlarged print	‡	‡	7
Familiar person administered test	‡	‡	‡
Familiar person present	3	14	27
Multiple test sessions	‡	4	9
Person to help child respond	‡	3	7
Sign language interpreter	‡	‡	3
Other	‡	5	15

‡ 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. This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).).

Table D-15. Unweighted number of children who had various test accommodations in the PEELS Wave 4 direct assessment, by Wave 1 primary disability: School year 2006-07

	AU	DD	ED	LD	MR	OI	OHI	SLI	LI
Abacus	‡	‡	‡	‡	‡	‡	‡	‡	‡
Adaptive furniture	‡	4	‡	‡	‡	3	‡	‡	‡
Communication device	‡	‡	‡	‡	‡	‡	‡	‡	6
Enlarged print	‡	3	‡	‡	‡	‡	‡	‡	‡
Familiar person administered test	‡	‡	‡	‡	‡	‡	‡	‡	‡
Familiar person present	8	16	‡	‡	‡	‡	‡	12	4
Multiple test sessions	5	3	‡	‡	‡	‡	‡	7	‡
Person to help child respond	‡	4	‡	‡	‡	‡	‡	‡	3
Sign language interpreter	‡	‡	‡	‡	‡	‡	‡	‡	5
Other	3	6	‡	‡	‡	‡	3	4	‡

‡ Reporting standards not met.
NOTE: AU = Autism; DD = ; 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. This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-16. Unweighted number of children who had various test accommodations in the PEELS Wave 4 direct assessment, by age cohort: School year 2006-07

	Cohort A (6 years old)	Cohort B (7 years old)	Cohort C (8 years old)
Abacus	‡	‡	‡
Adaptive furniture	8	‡	‡
Communication device	‡	‡	5
Enlarged print	4	3	‡
Familiar person administered test	‡	‡	‡
Familiar person present	22	14	9
Multiple test sessions	6	4	6
Person to help child respond	5	4	3
Sign language interpreter	‡	‡	3
Other	9	7	6

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

Table D-17. Unweighted number of children who had various test accommodations in the PEELS Wave 5 direct assessment, by gender: School year 2008-09

	Male	Female
Abacus	‡	‡
Adaptive furniture	3	‡
Communication device	5	‡
Enlarged print	‡	4
Familiar person administered test	‡	‡
Familiar person present	41	14
Multiple test sessions	7	‡
Person to help child respond	13	6
Sign language interpreter	‡	4
Other	20	7

‡ Reporting standards not met.
NOTE: This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-18. Unweighted number of children who had various test accommodations in the PEELS Wave 5 direct assessment, by race/ethnicity: School year 2008-09

	Black	Hispanic	White
Abacus	‡	‡	‡
Adaptive furniture	‡	‡	3
Communication device	‡	‡	7
Enlarged print	‡	‡	4
Familiar person administered test	‡	‡	‡
Familiar person present	3	10	38
Multiple test sessions	‡	‡	7
Person to help child respond	‡	3	14
Sign language interpreter	‡	‡	‡
Other	‡	7	14

‡ 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. This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).).

Table D-19. Unweighted number of children who had various test accommodations in the PEELS Wave 5 direct assessment, by Wave 1 primary disability: School year 2008-09

	AU	DD	ED	LD	MR	OI	OHI	SLI	LI
Abacus	‡	‡	‡	‡	‡	‡	‡	‡	‡
Adaptive furniture	‡	‡	‡	‡	‡	‡	‡	‡	‡
Communication device	‡	‡	‡	‡	‡	‡	‡	‡	4
Enlarged print	‡	‡	‡	‡	‡	‡	‡	‡	‡
Familiar person administered test	‡	‡	‡	‡	‡	‡	‡	‡	‡
Familiar person present	18	19	‡	‡	3	‡	‡	11	3
Multiple test sessions	3	‡	‡	‡	‡	‡	‡	‡	‡
Person to help child respond	3	8	‡	‡	‡	‡	‡	‡	3
Sign language interpreter	‡	‡	‡	‡	‡	‡	‡	‡	4
Other	8	10	‡	‡	‡	‡	‡	4	‡

‡ Reporting standards not met.
NOTE: AU = Autism; DD = ; 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. This table includes children in Cohorts A, B, and C.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS).

Table D-20. Unweighted number of children who had various test accommodations in the PEELS Wave 5 direct assessment, by age cohort: School year 2008-09

	Cohort A (8 years old)	Cohort B (9 years old)	Cohort C (10 years old)
Abacus	‡	‡	‡
Adaptive furniture	‡	‡	‡
Communication device	‡	3	4
Enlarged print	‡	3	‡
Familiar person administered test	‡	‡	‡
Familiar person present	25	15	15
Multiple test sessions	4	3	‡
Person to help child respond	7	4	8
Sign language interpreter	‡	‡	‡
Other	8	9	10

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

APPENDIX E: FINAL AUGMENTED LEA SAMPLE SIZE

Table E-1. Final augmented LEA sample size by district size and region

	Size				
Region	Total	Very large	Large	Medium	Small
Total	232	39	42	51	100
Northeast	66	9	13	14	30
Southeast	56	16	10	16	14
Central	63	3	8	15	37
West/Southwest	47	11	11	6	19

Note: District size was obtained through the LEA Policies and Practices Questionnaire and was based on report of total district enrollment. Using cutoffs from the National Center for Education Statistics (NCES) Common Core of Data, the districts were categorized as *small* if they had 300-2,500 students, *medium* if they had 2,501-10,000 students, *large* if they had 10,001-25,000 students, and *very large* if they had more than 25,000 students.

Table E-2. Final augmented LEA sample size by district size and wealth

	Size				
District wealth	Total	Very large	Large	Medium	Small
Total	232	39	42	51	100
High	67	4	10	15	38
Medium	67	8	14	14	31
Low	59	12	9	15	23
Very low	39	15	9	7	8

Note: District size was obtained through the LEA Policies and Practices Questionnaire and was based on report of total district enrollment. Using cutoffs from the National Center for Education Statistics (NCES) Common Core of Data, the districts were categorized as *small* if they had 300-2,500 students, *medium* if they had 2,501-10,000 students, *large* if they had 10,001-25,000 students, and *very large* if they had more than 25,000 students. District wealth was defined as a percentage of the district's children falling below the federal government poverty guidelines, where *high wealth* was 0-12 percent, *medium wealth* was 13-34 percent, *low wealth* was 35-40 percent, and *very low wealth* was more than 40 percent.

Table E-3. Final augmented LEA sample size by district region and wealth

	Region				
District wealth	Total	Northeast	Southeast	Central	West/Southwest
Total	232	66	56	63	47
High	67	31	5	19	12
Medium	67	13	13	29	12
Low	59	11	26	12	10
Very low	39	11	12	3	13

Note: District wealth was defined as a percentage of the district's children falling below the federal government poverty guidelines, where *high wealth* was 0-12 percent, *medium wealth* was 13-34 percent, *low wealth* was 35-40 percent, and *very low wealth* was more than 40 percent.

Table E-4. Participating LEA sample size by three stratification variables

	Size			
Total	Very large	Large	Medium	Small
223	39	42	51	91
	Region			
	Northeast	Southeast	Central	West/Southwest
223	63	55	59	46
	District wealth			
	High	Medium	Low	Very low
223	62	65	57	39

Note: District size was obtained through the LEA Policies and Practices Questionnaire and was based on report of total district enrollment. Using cutoffs from the National Center for Education Statistics (NCES) Common Core of Data, the districts were categorized as *small* if they had 300-2,500 students, *medium* if they had 2,501-10,000 students, *large* if they had 10,001-25,000 students, and *very large* if they had more than 25,000 students. District wealth was defined as a percentage of the district's children falling below the federal government poverty guidelines, where *high wealth* was 0-12 percent, *medium wealth* was 13-34 percent, *low wealth* was 35-40 percent, and *very low wealth* was more than 40 percent.

APPENDIX F: LIKELIHOOD RATIO TESTS FOR PREDICTION MODELS

For the analyses presented in this report, likelihood ratio tests were conducted to determine the ability of the models to predict the growth parameters of initial status, linear growth, and quadratic growth based on the inclusion of disability. If disability was a statistically significant predictors of initial achievement level, linear growth, or quadratic growth, *t*-tests were performed to judge whether growth parameters between each pair of disability categories was significantly different from one another. If disability was not a statistically significant predictor, no such *t*-tests were performed.

Likelihood Tests for Peabody Picture Vocabulary Test-III (PPVT-III adapted version)

Table F-1 compares the likelihood for growth models in which the growth parameters are predicted from membership in three disability categories: autism, speech or language impairment, and developmental delay. The first growth model excludes disability category as a predictor. In the second through fourth models, growth parameters are predicted by disability category. Table F-1 shows that predicting initial status by disability category increases the likelihood of the model (X^2 = 74.05, $p < .001$). The same is true for predicting linear growth (X^2 = 13.20, p = .002). However, adding disability did not increase the likelihood for predicting quadratic growth (X^2 = 3.01, p =.220).

Table F-1. Likelihood ratio tests for PPVT-III (adapted version) prediction models by disability

	Log likelihood	Sequence	X^2 This model vs. previous model	df	Prob
Growth parameters: Initial status, linear and quadratic growth	-26773.95	1			
Disability predicts initial status	-26736.93	2	74.05	2	<0.001
Disability predicts initial status and linear growth	-26730.32	3	13.20	2	0.002
Disability predicts initial status, linear growth, and quadratic growth	-26728.82	4	3.01	2	0.220

NOTE: df = degrees of freedom.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Peabody Picture Vocabulary Test-III" (December 2010).

Likelihood Tests for Woodcock-Johnson III: Applied Problems

Table F-2 shows models predicting the growth parameters of initial status, linear growth, and quadratic growth based on membership in three disability categories: autism, speech or language impairment, and developmental delay. The first model excludes disability as a predictor. The second through fourth models predict growth using disability. Table F-2 shows that adding disability to the model increased the likelihood for predicting initial status (X^2 = 167.06, $p < .001$). Similarly, adding disability significantly increased the likelihood for predicting linear growth (X^2 = 33.56, $p < .001$), but it did not increase the likelihood for predicting quadratic growth (X^2 = 4.00, p = 0.133).

Table F-2. Likelihood ratio tests for Woodcock-Johnson III Applied Problems prediction models by disability

	Log likelihood	Sequence	X^2 This model vs. previous model	df	Prob
Growth parameters: Initial status, linear and quadratic growth	-34634.45	1			
Disability predicts initial status	-34550.92	2	167.06	2	<0.001
Disability predicts initial status and linear growth	-34534.14	3	33.56	2	0.001
Disability predicts initial status, linear growth, and quadratic growth	-34532.14	4	4.00	2	0.133

NOTE: df = degrees of freedom.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), "Woodcock-Johnson III Applied Problems," (December 2010).

APPENDIX G: DETAILS OF THE LIKELIHOOD RATIO TESTS FOR THE FIT OF THE MERGED-COHORT MODELS

This appendix describes the variables and models that were used to test the fit of the merged-cohort models used in this report. Following Miyazaki and Raudenbush (2000), two models were compared:

a. A separate-cohort model in which there were cohort-by-age interactions and cohort-by-age by subgroup interactions. This model assumed that the growth profile over age groups differed by cohort. The cohort-by-age by subgroup interactions enabled analysts to perform a significance test to determine if the growth of disability subgroups differed by cohort.

b. A merged-cohort model in which data from different cohorts were merged. This allowed for growth profiles that included a wider range of ages than were covered by any single cohort. The merged-cohort model, also called an accelerated longitudinal design, assumes that a single growth profile describes growth for all cohorts (i.e., there are no cohort-by-age interactions). In addition, when considering subgroups such as disability classification, the merged-cohort model assumes that there are no cohort-by-subgroup growth interactions. For the whole merged sample, however, there can be different growth profiles for the modeled subgroups (i.e., disability classification).

G.1. Definition of dummy variables in the separate and merged-cohort models

Table G-1. Definition of cohort indicator variable: Cohort A, B, and C

Age cohort	Cohort B	Cohort C
Cohort A (reference group)	0	0
Cohort B	1	0
Cohort C	0	1

Table G-2. Definition of disability indicator variables: Cohort A, B and C

Disability	DevDelay	SpchLang
Autism (reference group)	0	0
Developmental Delay	1	0
Speech or Language	0	1

G.2. Cohort by Disability Interactions

Six dummy variables, with values of 0, 1 or -1, define the cohort by disability group interactions. There are eight cohort by disability group combinations, and accordingly there are eight different configurations of the three dummy variables, so including the dummy variables in the model accounts for the effect of all interactions.

Table G-3. Cohort by child's disability interactions

	Variable			
Value Labels	**CohXDis1**	**CohXDis2**	**CohXDis3**	**CohXDis4**
Cohort A, Autism	1	1	1	1
Cohort A, Developmental delay	-1	0	-1	0
Cohort A, Speech language	0	-1	0	-1
Cohort B, Autism	-1	-1	0	0
Cohort B, Developmental delay	1	0	0	0
Cohort B, Speech language	0	1	0	0
Cohort C, Autism	0	0	-1	-1
Cohort C, Developmental delay	0	0	1	0
Cohort C, Autism	0	0	0	1

Similar to the interactions described above, there are two dummy variables, with values of 0, 1 or -1, that define the nine cohort by income group interactions.

G.3. Models For the Separate-Cohort vs. Merged-Cohort Likelihood Ratio Tests

I. PPVT-III (adapted version)

1. Separate-Cohort Model for Disability Groups

This model assumes a separate growth profile for each disability group for each cohort. The initial status, linear, and quadratic growth parameters from level 1 are modeled at level 2 by main cohort and disability group effects and by cohort-by-disability group interactions. For child i in district j at time k:

Level-1 Model (time within child)

$$\text{PPVT}_{ijk} = \pi_{0ij} + \pi_{1ij} * \text{Age}^*_{ijk} + \pi_{2ij} * \text{AgeQ}^*_{ijk} + e_{ijk}$$

Level-2 Model (child within district)

$$\pi_{0ij} = \beta_{00j} + \beta_{01} * CohortB_{ij} + \beta_{02} * CohortC_{ij} +$$
$$\beta_{03} * DevDelay_{ij} + \beta_{04} * SpchLang_{ij} +$$
$$\beta_{05} * CohXDis1_{ij} + \beta_{06} * CohXDis2_{ij} + \beta_{07} * CohXDis3_{ij} + \beta_{08} * CohXDis4_{ij} + r_{0ij}$$
$$\pi_{1jk} = \beta_{10j} + \beta_{11} * CohortB_{ij} + \beta_{12} * CohortC_{ij} +$$
$$\beta_{13} * DevDelay_{ij} + \beta_{14} * SpchLang_{ij} +$$
$$\beta_{15} * CohXDis1_{ij} + \beta_{16} * CohXDis2_{ij} + \beta_{17} * CohXDis3_{ij} + \beta_{18} * CohXDis4_{ij} + r_{1ij}$$
$$\pi_{2ij} = \beta_{20} + \beta_{21} * CohortB_{ij} + \beta_{22} * CohortC_{ij} +$$
$$\beta_{23} * DevDelay_{ij} + \beta_{24} * SpchLang_{ij} +$$
$$\beta_{25} * CohXDis1_{ij} + \beta_{26} * CohXDis2_{ij} + \beta_{27} * CohXDis3_{ij} + \beta_{28} * CohXDis4_{ij} + r_{2ij}$$

Level-3 Model (district)

$$\beta_{00j} = \gamma_{000} + u_{00j}$$

Where,

Age^*_{ijk} and AgeQ^*_{ijk} are linear and quadratic contrasts for age, deviated from cohort medians, $\pi_{0ij}, \pi_{1ij}, \pi_{2ij}$ are the initial status, linear, and quadratic growth parameters for each child (level 1), e_{ijk} is the age within child error term,

CohortB and CohortC are the cohort membership indicators,

DevelDelay, and SpchLang are the disability category indicator variables,

CohXDis1, CohXDis2, CohXDis3, and CohXDis4 are the cohort by disability interaction variables,.

The β's at level 2 are the regression parameters for predictors of the growth parameters at the child level.

r_{0ij}, r_{1ij}, r_{2ij} are the child level random effects for initial status, linear, and quadratic growth parameters, γ_{000} is the grand mean for the outcome, and

u_{00j} is the random effect for district j.

Some of the predictors in the level 1 model on page G-3 were not significant and were not included in the separate cohort models. The same holds true for the analysis of the Applied Problems measure.

2. Merged-Cohort Model- for disability groups

This merged-cohort model assumes that all cohorts can be merged together by age. The whole-group age contrast variables are included (Age, AgeQ) as well as the cohort-centered age contrasts (Age^*, AgeQ^*). The latter two contrasts are included in accordance with the Miyazaki and Raudenbush (2000) formulation, so the error structure of the separate cohort model is preserved, and the merged-cohort model is nested within the separate-cohort model. If the likelihood ratio test shows that the likelihood for this model is greater than or equal to the separate-cohort model, then the cohorts can be merged in a final analysis. It should be noted that this configuration of the model is only for the purposes of conducting the likelihood ratio test. The final models that describe growth by subgroup for the merged-cohort model do not include cohort-centered age variables.

Also note that although the merged cohort model specifies that disability groups have different growth profiles, the growth profiles for each group is the same for all three cohorts.

It was found that the merged-cohort model fit the data as well or better than the separate-cohort model only if income was included as a covariate at level two. This was true for both outcomes, PPVT-III (adapted version) and Applied Problems.

In selecting covariates, we considered two points: the concomitant reduction in degrees of freedom and possible measurement error in the covariates. To limit the loss in degrees of freedom, we wanted to enter as few covariates as possible while enhancing the model fit. To minimize measurement error, we considered use of demographic variables, which typically demonstrate high reliability. Age was already included as part of the model, and gender was not highly correlated with outcomes in previous PEELS reports (Markowitz et al. 2006, Carlson et al. 2009). Researchers had previously documented the correlation between household income and both PPVT and Applied Problems scores in PEELS (Markowitz et al. 2006) as well as in other studies of children and youth with disabilities (Wagner, Newman, Cameto, and Levine 2006). Household income is a common covariate in studies of educational performance (see, for example, Guarino, Hamilton, Lockwood, and Rathbun 2006; Walston and West 2004). In this analysis, household income enhanced model fit adequately and was used as the sole covariate.

As a result, all of the final models for growth include income as a covariate (see appendix H for details).

Merge Cohort Model:

Level-1 Model (time within child)

$$\mathrm{PPVT}_{ijk} = \pi_{0jk} + \pi_{1ij} * \mathrm{Age}_{ijk} + \pi_{2ij} * \mathrm{AgeQ}_{ijk} + \pi_{3ij} * \mathrm{Age}^{*}_{ijk} + \pi_{4ij} * \mathrm{AgeQ}^{*}_{ijk} + e_{ijk}$$

Level-2 Model (child within district)

$$\pi_{0jk} = \beta_{00j} + B_{01} * IncomeHigh_{ij} + \beta_{02} * DevDelay_{ij} + \beta_{03} * SpchLang_{ij} + r_{0ij}$$

$$\pi_{1jk} = \beta_{10} + B_{11} * IncomeHigh_{ij} + \beta_{12} * DevDelay_{ij} + \beta_{13} * SpchLang_{ij} + \beta_{13} * Other_{ij}$$

$$\pi_{2jk} = \beta_{20} + B_{21} * IncomeHigh_{ij} + \beta_{22} * DevDelay_{ij} + \beta_{23} * SpchLang_{ij} + \beta_{23} * Other_{ij}$$

$$\pi_{3jk} = r_{3ij}$$

$$\pi_{4jk} = r_{4ij}$$

Level-3 Model (district)

$$\beta_{00j} = \gamma_{000} + u_{00j}$$

Where,

Age^*_{ijk} and AgeQ^*_{ijk} are linear and quadratic contrasts for age, deviated from cohort medians,

Age_{ijk} and AgeQ_{ijk} are linear and quadratic contrasts for age not mean deviated,

DevelDelay, SpchLang, Other are the disability category dummies,

$\pi_{0ij}, \pi_{1ij}, \pi_{2ij}, \pi_{3ij}, \pi_{4ij}$ are the growth parameters associated with the intercept, linear age, quadratic age, linear cohort-deviated age, and quadratic cohort deviated age, respectively.

$\beta_{00j}, \beta_{01}, \beta_{02}, \beta_{03}$ are the intercept and disability group regression effects associated with the initial status growth parameter,

$\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}$ are the intercept and disability group regression effects associated with the linear growth parameter,

$\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}$ are the intercept and disability group regression effects associated with the quadratic growth parameter.

r_{0ij}, r_{1ij}, r_{2ij} are the child level random effects for initial status, linear, and quadratic growth parameters,

γ_{000} is the grand mean for the outcome, and

u_{00j} is the random effect for district j.

Models for Applied Problems

The separate-cohort and merged-cohort models for disability groups are the same for the PPVT-III (adapted version) and Applied Problems measures.

G.4. Results of likelihood ratio tests for PPVT-III (adapted version), Letter-Word Identification and Applied Problems

Table G-4 summarizes the likelihood ratio test results comparing the likelihood of the merged-cohort model (income included as a covariate) with the likelihood of the separate- cohort model for the full-cohort sample. For both PPVT- III (adapted version) and Applied Problems outcomes, the likelihood test is significant, but the merged-cohort model fits better (has a higher likelihood) than the separate-cohort model. For Letter Word Identification, the merged-cohort model likelihood test is significant, *but* the merged-cohort model fits worse than the separate cohort model (has a lower likelihood).

Table G-4. Fit of Merged-Cohort Model for Disability Groups With Income as a Covariate

	Chi-Square test of difference in models	Which model fits the best
PPVT-III (adapted version)	41.95 (<.001)	Merged-cohort
Applied Problems	54.92 (<.001)	Merged-cohort
Letter Word Identification	589.09 (<.001)	Separate-cohort

APPENDIX H: HIERARCHICAL LINEAR MODELS USED IN THE ANALYSIS

The hierarchical linear model used in this report can be characterized as a series of ordinary regressions specified at each of the three levels of aggregation. In the first level, the child's score is modeled as a function of initial status, linear growth, and quadratic growth:

Level 1 Model, Repeated Measures Within Child:

$$Achievement_{ija} = \pi_{0ij} + \pi_{1ij} * \text{Age}_{ija} + \pi_{2ij} * \text{AgeSq}_{ija} + e_{ija}$$

for age *a* of child *i*, within LEA *j*. $Achievement_{ija}$ is the achievement outcome (either PPVT-III (adapted version) or Applied Problems). π_{0ij} is the initial status of child *i*. Age is the linear age contrast of the child at a particular wave of data collection. Although there were 5 waves of data collection possible for each child, the children's ages ranged from 3 to 10, depending on the cohort and wave of observation.[1] The predictor, AgeSq, is the square of the Age contrast.[2] π_{1ij} and π_{2ij} are the linear and quadratic growth parameters for child *i*.

At Level 2, child nested within LEA, the growth parameters for each child are modeled by income and the disability group the child belongs to. There are three disability groups the child could belong to, Autism, Developmental Disability, or Speech or Language Impairment. The level two model for growth parameters is:

Level-2 Model (child within district)

$$\pi_{0ij} = \beta_{00j} + \beta_{01} * IncomeHigh_{ij} + \beta_{02} * DevDelay_{ij} + \beta_{03} * SpchLang_{ij} + r_{0ij}$$
$$\pi_{1ij} = \beta_{10} + \beta_{11} * IncomeHigh_{ij} + \beta_{12} * DevDelay_{ij} + \beta_{13} * SpchLang_{ij} + r_{1ij}$$
$$\pi_{2ij} = \beta_{20} + \beta_{21} * IncomeHigh_{ij} + r_{2ij}$$

where r_{0ji}, r_{1ji}, and r_{2ji} are random student effects. In this report income is an indicator of whether the child is from a high income household (0 for low and 1 for high). The grouping variables, DevDelay and SpchLang are indicators of whether the student is in the developmental delay or speech language groups. These are 0,1 indicator variables. As a result, if DevDelay and SpchLang are both zero, the child belongs to the reference group, Autism.

In the first equation, which predicts initial status (π_{0ij}), the regression coefficients, $\beta_{00j}, \beta_{01}, \beta_{02}, \text{and}, \beta_{03}$ are the mean status[3], the income effect, the effect of being in the development disability group and the effect of being in the speech language group (respectively), on initial status.

---

[1] Age 3 is the reference category, so the linear age contrast for age 3 is zero. For ages 3 through 10, linear age contrasts range from zero to 7.

[2] With this definition of age contrasts Age and AgeSq at age 3 is zero. This means that the intercept is the child's initial status.

[3] When all predictors are zero.

Similarly, for the 2nd equation, which predicts or linear growth for student i (π_{1ij}), $\beta_{10}, \beta_{11}, \beta_{12}$, and β_{13} are the mean linear growth[3], the income effect, the effect of being in the development disability group and the effect of being in the speech language group (respectively), on linear growth.

In the third equation, which predicts quadratic growth,(π_{2ij}), the first term, β_{20}, is the mean quadratic growth[3]. The second term, β_{21}, is the effect of income on quadratic growth. Note that the likelihood ratio tests indicated that disability group was not a significant predictor of quadratic growth (see appendix F).

In the level 3 (LEA) model, only a grand mean, γ_{000}, and random effect, u_{00j}, are included. The purpose of this part of the model is to account for the cluster effect of child means within LEA.

Level-3 Model (LEA)

$$\beta_{00j} = \gamma_{000} + u_{00j}$$

Centering options

At level 1 the linear and quadratic age contrasts were centered at age 3 for most of the analyses. For reporting means and standard errors for other ages, the age contrasts were centered at the respective ages.

For the level 2 income and disability predictors, the indicator variables were entered without centering. This produces the same group and covariate effects and standard errors as grand-mean centering, which is recommended by Raudenbush and Bryk 2001[4]. Since all of the level 2 covariates are indicator variables of group membership, it was thought that interpretability of the analysis would be simpler without centering.

Use of sampling weights in the HLM analysis

For continuous normal outcomes, Pfeffermann, Skinner, Goldstein, and Rasbash 1998)[5] recommend partitioning the sampling weight for individuals into two components: a component for individuals within groups and a component for groups. The current analysis used a 3-level HLM model with the levels corresponding to repeated measures within student, students within LEA, and LEA. Since there was no weighting component for repeated measures, we used method 2 recommended by Pfeffermann, et al., where separate weight components are defined for students and LEAs.

---

[4] Raudenbush, R. and Bryk, A. (2001). *Hierarchical Linear Models*, (p. 33). Thousand Oaks, CA: Sage.

[5] Pfeffermann, D., Skinner, C., Goldstein, H. and Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models, *Journal of the Royal Statistical Society, B, 60*:23-40.

Most of the commercially available programs specialized to do multilevel analysis, including HLM which was used for the analysis presented in this report, implement the method 2 form of the multilevel weights suggested by Pfeffermann, et al. With this method, the conditional individual-within-group weight is normalized to the number of units actually observed in each group.

Variance components of the three-level hierarchical linear models

Multilevel analyses partitions the total variance of the outcome into components corresponding to each level of the model. In the analyses for this report there is a variance component for repeated measures within child, for effects between children within LEA and for effects between LEAs.

Table H-1 presents the variance components for the 3-level PPVT-III (adapted version) analysis. For the PPVT-III (adapted version) outcome, 16% of the variance was at level 1, measures within children. The remainder of the variance, 78%, was at level 2, between students. Five percent of the variance was at level 3, between LEAs.

Table H-1. PPVT –III (adapted version) variance components

Source	Description	Variance	Df	Chi-Square	Prob	% of Total
Level 1	Within-child	26.98				16%
Level 2	Between children					
Intercept	Average child outcome	124.09	1413	4477.24	<0.001	75%
Linear	Linear growth component	4.96	1618	2315.70	<0.001	3%
Quadratic	Quadratic growth component	0.02	1618	2043.39	<0.001	0%
Level 3	Between LEA variance	9.06	205	577.01	<0.001	5%

NOTE: Df = degrees of freedom.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), “Peabody Picture Vocabulary Test-III” (December 2010).

Table H-2 presents the variance components for the 3-level Applied Problems analysis. Similar to the PPVT-III (adapted version), 15% of the variance was at level 1, measures within student. Seventy-five percent was at level 2, between students. Ten percent of the variance was at level 3, between LEAs.

Table H-2. Applied Problems variance components

Source	Description	Variance	Df	Chi-Square	Prob	% of Total
Level 1	Within child	134.67				15%
Level 2	Between children					
Intercept	Average child outcome	605.29	1411	4868.39	<0.001	66%
Linear	Linear growth component	86.20	1616	2970.43	<0.001	9%
Quadratic	Quadratic growth component	1.14	1618	2572.52	<0.001	0%
Level 3	Between LEA variance	95.25	205	597.23	<0.001	10%

NOTE: df = degrees of freedom.
SOURCE: U.S. Department of Education, National Center for Special Education Research, Pre-Elementary Education Longitudinal Study (PEELS), “Woodcock-Johnson III Applied Problems subtest” (December 2010).

www.ed.gov ies.ed.gov

33242.0711.80890615