

Pre-K for All:

Snapshot of Student Learning



Submitted to:

The New York City Center for
Economic Opportunity and the New
York City Department of Education

Submitted by:

Westat
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

—and—

Metis Associates
120 Wall Street
New York, New York 10005
(212) 425-8833

—and—

Branch Associates
1628 JFK Blvd., Suite 800
8 Penn Center
Philadelphia, PA 19103
(267) 232-0261

August 2016



Pre-K for All 2014-15 Evaluation Response Memo

Pre-K for All is New York City's historic initiative to provide every four-year-old with access to free, full-day, high-quality pre-kindergarten through a two-year expansion that began in the 2014-15 school year. Before Pre-K for All, only 19,287 four-year-olds were enrolled in full-day pre-K in New York City; as of the 2015-16 school year, enrollment was 68,647.

The City implemented the rapid, at-scale universal pre-K program within a short two-year timeframe because filling the gap in access to full-day pre-K was crucial—four-year-olds eligible to enroll in September 2014 would not get another chance to attend pre-K. The City secured funding and quickly began to prepare for the 2014-15 school year. In the summer of 2014, the NYC Department of Education (DOE) and partner agencies worked closely with pre-K providers to ensure they were ready to open their doors on the first day of school. At the same time, the City launched an unprecedented grassroots campaign to recruit and enroll families. This included establishing an Outreach Team of dedicated pre-K enrollment specialists to call families and canvas local communities.

The City's comprehensive approach was grounded in creating a sustainable, high-quality, full-day pre-K model. From its inception, the expansion focused not only on ensuring access but also on investing in pre-K quality. The City built a single system of free, full-day, high-quality pre-K and developed a quality infrastructure to support long-term sustainability and quality improvements. The DOE's model provides all pre-K programs with differentiated support at the classroom- and program-level that focuses on implementing research-based instructional and family engagement practices. Some examples include: free and targeted professional learning for leaders, lead teachers, assistant teachers, and paraprofessionals; on-site support for leaders and teachers from Instructional Coordinators (ICs) and Social Workers (SWs); and guidance through online tools and other resources.

The first year of the expansion marked the beginning of a rigorous two-part research study of this work. The DOE, in conjunction with the NYC Center for Economic Opportunity, collaborated with Westat, Metis Associates, and Branch Associates, with supplemental support from the New York University's Institute for Human Development and Social Change, to undertake a study to inform future years of program delivery as well as lay the foundation for long-term research in the future. The Year 1 evaluation of Pre-K for All included an evaluation of the effectiveness of the implementation process and a snapshot of student learning in the first year of the Pre-K for All initiative.

This memo summarizes the findings of the Year 1 evaluation of Pre-K for All and concludes by outlining accomplishments and improvements made in the 2015-16 school year that address many of the report findings.

Year 1 Evaluation Overview

The analysis conducted over the course of the 2014-15 school year was based on surveys, focus groups, interviews, and assessment data from a variety of stakeholders including parents, teachers, principals, site directors, DOE staff, and staff from other City agencies. The evaluation covered seven areas of Pre-K for All's implementation and are captured in separate reports:

1. Family perceptions of the program
2. Family engagement and communication
3. Curriculum and instruction
4. Using data for programmatic and instructional purposes
5. Expansion rollout
6. Program supports
7. Executive functioning and academic skills

Overall, the studies found:

- o 92 percent of surveyed parents rate the quality of their child's pre-K program as "good" or "excellent" and 83 percent of surveyed parents report that Pre-K for All improved their child's learning and behavior "a lot."
- o Sites offering Pre-K for All report using a variety of family engagement and communication practices.
- o Sites offering Pre-K for All report that they feel supported by the DOE in implementing curriculum.
- o Sites offering Pre-K for All report using a wide variety of data to inform instruction and make programmatic decisions.
- o Most providers that applied to offer free, full-day Pre-K for All report that the application process was clear and well-supported.
- o Nearly 80 percent and 88 percent of site leaders report that staff recruitment and staff retention, respectively, did not pose significant challenges.
- o A majority of site leaders and teachers report using each type of support provided by the DOE (*ex: coaching, professional development, etc.*). In general, Pre-K for All sites report that these supports are helpful.
- o A positive impact on students—across income levels, race, and home language status—was seen through their gains in executive functioning skills and academic skills over the course of the study period.

Year 1 Evaluation Report Summaries

1. Report on Family Perceptions

- 92 percent of surveyed families rate the quality of their child's pre-K program as "good" or "excellent," and 83 percent report that Pre-K for All improved their child's learning and behavior "a lot."
- Nearly 80 percent of surveyed families report receiving resources from their Pre-K for All program to support learning at home.
- Families report that the availability of Pre-K for All affected decisions about child care and labor force participation.
 - o Of the families that were surveyed, more than half (56 percent) report a decrease in spending on childcare from 2013-14 to 2014-15. Surveyed families report an average decrease of \$78 per week in spending.

- o Of the families that reported that Pre-K for All affected the number of hours they worked, half report an increase in hours worked, which they attribute to the availability of full-day pre-K.

2. Report on Family Engagement and Communication

- Overall, sites report undertaking a variety of family engagement and communication practices as a component of Pre-K for All. These include, but are not limited to: using face-to-face communication, providing updates on students' achievements, having accessible program staff and multi-lingual staff, translating communications, using family input to make decisions, and providing opportunities for families to be involved with the program.
- Survey and interview findings also demonstrate that site leaders and instructional staff express a commitment to involving families in the education of their children.

3. Report on Curriculum and Instruction

- Overall, Pre-K for All sites report using a range of curricula to meet the needs of their students and that curriculum satisfaction is high among staff.
- The large majority of site staff report that their curriculum is vertically aligned to kindergarten and beyond, either to a moderate or large extent.
- Pre-K for All sites report using their curricula effectively and confidently to meet students' needs; however, program staff also report requests for continued training and support to improve quality.

4. Report on Use of Data for Programmatic and Instructional Purposes

- Overall, Pre-K for All sites report using a wide range of data to inform site-level programmatic decisions and classroom-level teaching practices, which include: authentic assessments of children's learning, program quality assessments, and feedback from DOE support staff and families.
- Authentic assessment data is a valuable data source for children's learning and development and 89 percent of sites report using these data for a variety of purposes. However, sites' perceived comfort with the authentic assessment systems vary by site type.
- 89 percent of site leaders report that their site uses data to engage families to a moderate or large extent.

5. Report on Pre-K for All Expansion Rollout

- Most providers that applied to offer full-day pre-K report that the application process was clear, easy to navigate, and well-supported. In general, sites report understanding how to be in compliance with DOE and Department of Health and Mental Hygiene (DOHMH) expectations.
- Key stakeholders report that the engagement of a large number of key players and City agencies, increases in City agency capacity and infrastructure, and collaboration within and between City agencies, demonstrated a high-level of commitment to reach the initiative's goals and were major successes.
- Nearly 80 percent and 88 percent of site leaders report that staff recruitment and staff retention, respectively, did not pose significant challenges.
- On average, lead teachers report having five years of experience in a pre-K educational setting and almost 13 years in any educational setting. Approximately eight out of every ten lead teachers report having the NYS Early Childhood certification and almost all of those who did not have certification were currently pursuing it.

- Overall, the large majority of surveyed pre-K instructional staff report being “satisfied” or “very satisfied” with their pre-K teaching experience in 2014-15.r pre-K teaching experience in 2014-15.

6. Report on Program Supports

- A majority of site leaders and teachers report receiving or using each type of support provided by the DOE (*ex: coaching, professional development, etc.*), and a majority also report that each type of support was “moderately” or “very” helpful.
- Nearly all site leaders (96 percent) report that they or their staff participated in the DOE-sponsored training that took place four times during the year. Overall, the large majority of site leaders and instructional staff (80 percent) report finding each of these professional development opportunities to be helpful.
- Sites report using a variety of resources and materials to support implementation of their Pre-K for All program. Nearly all site leaders (96 percent) report that their site used the DOE website to access units of study, lesson plans, and guidelines for the daily schedule, among other resources. They report the website is useful.

7. Report on Executive Functioning and Academic Skills

- A unique feature of this study is the collection of data on children’s executive functioning skills, a set of skills that includes their inhibitory control, working memory, and ability to shift between pieces of information, that together support children’s self-regulation. Executive functioning skills were measured by two widely-used assessments known as the Pencil Tap task and the Hearts and Flowers task.
- Statistically significant fall-to-spring gains were observed in both measures of executive functioning. The gains in the percentage of correct responses in the Pencil Tap and Hearts and Flowers tasks were 10 percent and 18 percent, respectively.
- Children attending Pre-K for All made statistically significant gains across all academic skills (Letter Recognition, Pre-writing, and Early Math) over the course of a 5.5-month testing window. By the end of this time period, Pre-K for All children outpaced the learning of four-year-olds nationally and were classified as being in the average range across all academic skills.
- This study featured a pre-post design without a comparison group, which means that observed gains in child learning cannot be attributed solely to participation in Pre-K for All. Children naturally learn and develop over time, and the study design means that these gains are confounded with the effects of the Pre-K for All program. Therefore, we cannot estimate the extent to which Pre-K for All was responsible for the children’s learning and development.

2015–16 | Year 2

Updates

The accomplishments and improvements in the second year of the expansion build on the work done in Year 1 to develop a high quality Pre-K for All system. They were informed by the results from the 2014-15 evaluation of Pre-K for All, feedback from Division of Early Childhood Education (DECE) field staff, ongoing program assessments, and partnerships with early childhood education experts.

In the second year of the expansion, the DOE introduced the Pre-K for All Program Quality Standards (PQS), which define the DOE’s vision for high-quality Pre-K for All programs in NYC. The PQS describes the key practices of family engagement, rigorous and developmentally appropriate instruction, professional collaboration, and leadership that support children in gaining the knowledge and skills outlined in

the http://www.p12.nysed.gov/ciai/common_core_standards/pdfdocs/nyslsprek.pdf. The PQS establishes a shared set of expectations for all pre-K programs; the DOE, leaders, educators, and families all use the PQS to understand and advance program quality.

EXPANSION AND POLICY

- The 2015-16 school year marked the first time that every four-year-old in New York City had access to free, full-day, pre-K. As of the 2015-16 school year, 68,647 children were enrolled across all Pre-K for All programs—a number more than triple the number of children who were enrolled before the expansion and larger than the entire school population of major cities like Boston. Enrollment is high across every community, with the highest participation among low-income families.
- The DOE introduced a streamlined pre-K enrollment process for families, which provides one application for families to apply to pre-K programs. Overall, 88 percent of families received a pre-K offer to one of their top three choices through the new streamlined application process.
- The DOE developed and shared critical policy guidance for NYC Early Education Centers (NY-CEECs) to ensure they are adequately supported as they join or continue as Pre-K for All partners.

DIFFERENTIATED SUPPORTS: In 2015-16, the DOE continued to advance its differentiated supports to all programs, tailoring support to each program's needs in order to meet Pre-K for All's Program Quality Standards.

Instructional Tracks and Lanes

- The DOE launched its Pre-K for All Instructional Tracks, providing every pre-K site with differentiated professional learning through a Summer Institute for teachers and leaders and a series of four teacher sessions and three leader sessions during the school year. Based on a variety of factors such as interest, demonstrated need, recommendations from Instructional Coordinators and Social Workers, site quality, and geography, sites were selected to participate in one of the following professional learning tracks and lanes:
 - *NYC Pre-K Explore:* Pre-K sites that participated in the Explore track used the evidence-based Building Blocks math curriculum together with the Pre-K for All Interdisciplinary Units of Study. Paired together, these materials provide a comprehensive, developmentally-appropriate approach to learning in pre-K.
 - *Advancing Social Emotional Development:* Pre-K sites in this lane advanced ways to support pre-K learners in developing social emotional skills needed to build a positive sense of self, form positive relationships, self-regulate, and adapt to change.
 - *Using Data to Inform Instruction:* Pre-K sites in this lane moved each child forward by learning new strategies to identify and meet each learner's needs, using authentic assessments and other data points.
 - *Supporting Linguistically and Culturally Diverse Learners:* In this lane, pre-K sites built on the diverse backgrounds and languages children and families brought to the classroom, with strategies for developing learning environments in which all children can thrive and all families are strong partners.

Coaching

- The DOE expanded its cadre of staff to provide on-site support to programs, including Instructional Coordinators and Social Workers.

- To more effectively differentiate support, in the 2015-16 school year, Instructional Coordinators and Social Workers conducted over 1,800 Foundational Support Visits (FSVs) to pre-K sites. Instructional Coordinators and Social Workers used information from these initial visits, ongoing observations, and pre-K program quality assessments to tailor their supports.
- The DOE established partnerships to provide specialized coaching for programs in targeted areas such as the Building Blocks math curriculum and using data to inform instruction.

Interdisciplinary Units of Study

- The DOE created the Pre-K for All research-based Interdisciplinary Units of Study to support student learning in all domains using developmentally appropriate practice. Throughout the year, the DOE released ten interdisciplinary units grounded in the NYS PKFCC.

PROGRAM MEASUREMENT AND USE OF DATA

- Because of its commitment to consistent quality measurement through program assessments, the DOE increased its capacity to provide more frequent program assessments, the Early Childhood Environmental Rating Scale-Revised (ECERS-R), and the Classroom Assessment Scoring System (CLASS). The DOE committed to a three-year cycle for each assessment by the 2016-17 school year for ECERS-R and the 2017-18 school year for CLASS.

OTHER KEY INITIATIVES AND PARTNERSHIPS

- The DOE partnered with researchers at New York University to develop a system of differentiated support that utilizes data on program needs and quality levels; the purpose of this system is to make decisions about the supports each program in our system receives across various aspects of the Pre-K Quality Standards. This is part of an ongoing partnership between DECE and NYU.
- In 2015-2016, the DECE continued its partnership with the Office of Special Education to develop resources and professional learning opportunities so that Instructional Coordinators, Social Workers, teachers, and leaders further strengthen their work to ensure that all children are successfully supported in achieving high expectations for their learning and developmental progress.
- The DOE launched a historic Teacher Incentive Program to support NYCEECs in recruiting and retaining top talent. Through the Pre-K for All Lead Teacher Incentive Program, there are two types signing incentives for certified lead teachers in Pre-K for All classrooms: the Retention Incentive Program for returning certified lead teachers and the New Hire Incentive Program for newly-hired certified lead teachers.

YEAR 2 EVALUATION

The Year 2 evaluation will produce actionable findings that will inform how the DOE can support pre-K programs to advance student learning. The Year 2 evaluation seeks to inform:

- How programs can better support students of different backgrounds and needs and how differentiated supports can serve students with special needs, students whose home language is a language other than English, and students living in poverty.
- The impact of the Pre-K for All's coaching models and professional development to understand how well the DOE is targeting sites for the right kinds and dosage of support based on the areas of growth identified in Year 1 and the Foundational Support Visit.

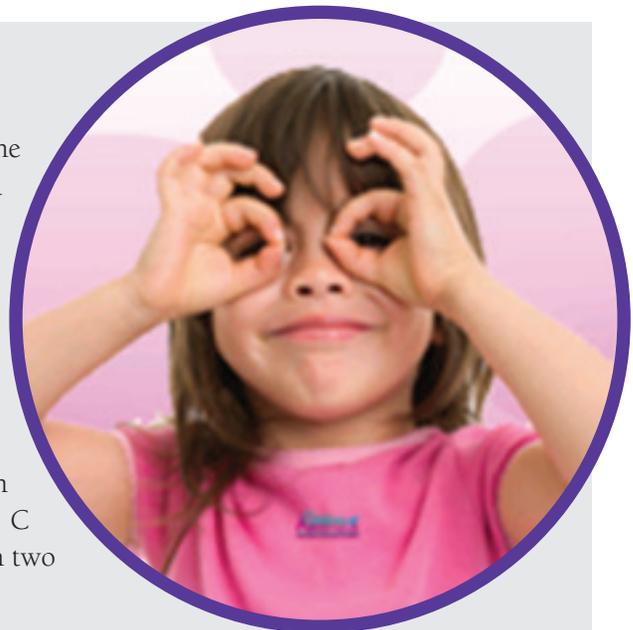
Westat, Metis Associates, and Branch Associates are conducting a comprehensive evaluation to assess the implementation efforts of the *Pre-K for All* initiative in New York City (NYC). As a demonstration of its commitment to learning and quality improvement, the City—the NYC Center for Economic Opportunity (CEO) and the NYC Department of Education (DOE), in cooperation with the NYC Administration for Children’s Services (ACS)—contracted this evaluation in 2014 as a means of gaining actionable information to inform implementation.

This “**Pre-K for All: Snapshot of Student Learning**” report is one in a series designed to provide DOE with information needed to make important policy and programmatic decisions. The study was conducted in the program’s inaugural year in order to give the city a snapshot of student learning to improve the services provided and serve as an initial guidepost to ensure the City’s efforts are in the right direction. Other reports in the series describe the program’s implementation and effects on families.

The academic, social, and emotional benefits of high quality pre-K have already been demonstrated by a substantial body of literature. These studies have shown that high quality pre-K is cost-effective and equips children with the cognitive skills needed for success in elementary school and beyond (Campbell, Ramey, Pungella, Sparling, & Miller-Johnson, 2002; Duncan et al., 2007; Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010; McClland, Acock, & Morrison, 2006; National Early Literacy Panel 2008; Rathbun & Zhang, 2016; Reynolds, Temple, White, Ou, & Robertson, 2011; and Weiland and Yoshikawa, 2013). Therefore, the goal of this study is not to reaffirm these benefits. Instead, this study **describes the first cohort of children to participate in the program, with a focus on the skills children had at entry and their growth over the 2014-2015 study period.** As such, the study is **formative** and **descriptive** in nature—it is designed to inform policy to help maximize child learning and growth.

Organization of the Report

Chapter 1 briefly describes the methodology of the study, including a description of the sample of sites and children participating in the study, and a summary of some important design and data considerations. Chapter 2 presents the results of the analysis of children’s skills, while Chapter 3 presents the results of the executive functioning analysis carried out by our research partners at New York University. Chapter 4 summarizes the key findings. Appendix A presents additional detail on the Woodcock-Johnson achievement assessments, while Appendices B and C provide the results of the analysis of children’s skills in two different score metrics.



Methodology

From the 201 Pre-K for All sites participating in the implementation portion of the study, a citywide representative subsample of 75 sites were recruited to participate in the student learning portion of the study. From these sites, we assessed the early learning skills and executive functioning skills of 1,145 4-year-olds. The instruments used to assess children's early learning skills were selected from the Woodcock-Johnson (Woodcock, McGrew & Mather, 2001) set of assessments. The Woodcock-Johnson is an assessment of academic skills designed for children, adolescents, and adults ranging from age 2 through 90 years (Bradley Johnson, 2004, p.1). It is individually administered by a trained examiner, and is scored using a computer program called Compuscore. The Woodcock-Johnson is widely used in the field of early childhood education, providing valid and reliable¹ information on student skills. In fact, independent reviewers have said *"The Woodcock-Johnson must be considered the premier battery for measuring both the cognitive abilities and school achievement of school-aged children and young adults. The test has been well standardized and thoughtfully developed."* (Cizek & Sandoval, 2003, p. 1027). From the set of available Woodcock-Johnson assessments, the following subtests were selected to assess early literacy and math skills:

**1,145
children
from 75 sites
assessed**

Early literacy

- Letter Recognition (measured by the Letter-Word Identification subtest²)
- Pre-Writing (measured by the fine motor skills/Spelling subtest³)

Early math

- Early Math (measured by the Applied Problems subtest⁴)

These subtests are described in more detail in Appendix A. We also selected a supplementary subtest (Passage Comprehension⁵) to assess conceptual matching skills, which are a precursor to reading. This measure is statistically significantly correlated (at the .01 level) with the other measures used in this study – see Appendix A for the correlation matrix.

To assess children's executive functioning skills, the Westat team partnered with researchers at New York University. Executive functioning is the ability of an individual to avoid distractions, focus attention, hold relevant information in working memory, and regulate impulsive behavior—all of which are important because they set the stage for early learning success. The Pencil Tap and Hearts and Flowers tasks were used to assess these skills.

¹Test-retest reliability is reported to be around .90 for all subtests used in this study (Woodcock, McGrew & Mather, 2001).

²Letter-Word Identification assesses children's letter and word identification ability. Items include identifying and pronouncing letters and words presented to the child.

³The Woodcock-Johnson refers to this subtest as "Spelling." For four-year-olds, Spelling assesses children's pre-writing skills, such as drawing lines and tracing, and writing letters.

⁴Applied Problems assesses children's ability to solve numerical and spatial problems presented verbally with accompanying pictures of objects.

⁵Passage Comprehension assesses children's ability to match conceptually similar pictures with appropriate words that maintain the semantic properties of the stimulus.

Sample

The Pre-K for All sites and the children participating in this study were similar to those who did not participate, as described below.

Sites

As mentioned above, a representative sample of 75 Pre-K for All sites located in each of the five boroughs was recruited to participate in the “Snapshot of Student Learning” portion of the study⁶. Table 1.1 compares these sites to all Pre-K for All sites in each borough and by site type.

Table 1.1 Percentage of sites by borough and site type

Borough	Sites in Study	Pre-K for All Sites
Bronx	21%	18%
Brooklyn	36%	34%
Manhattan	12%	15%
Queens	27%	27%
Staten Island	4%	6%

Note: Differences are not statistically significant

Site Type	Sites in Study	Pre-K for All Sites
ACS NYCEEC*	15%	22%
DOE NYCEEC*	43%	39%
DOE PS**	43%	39%

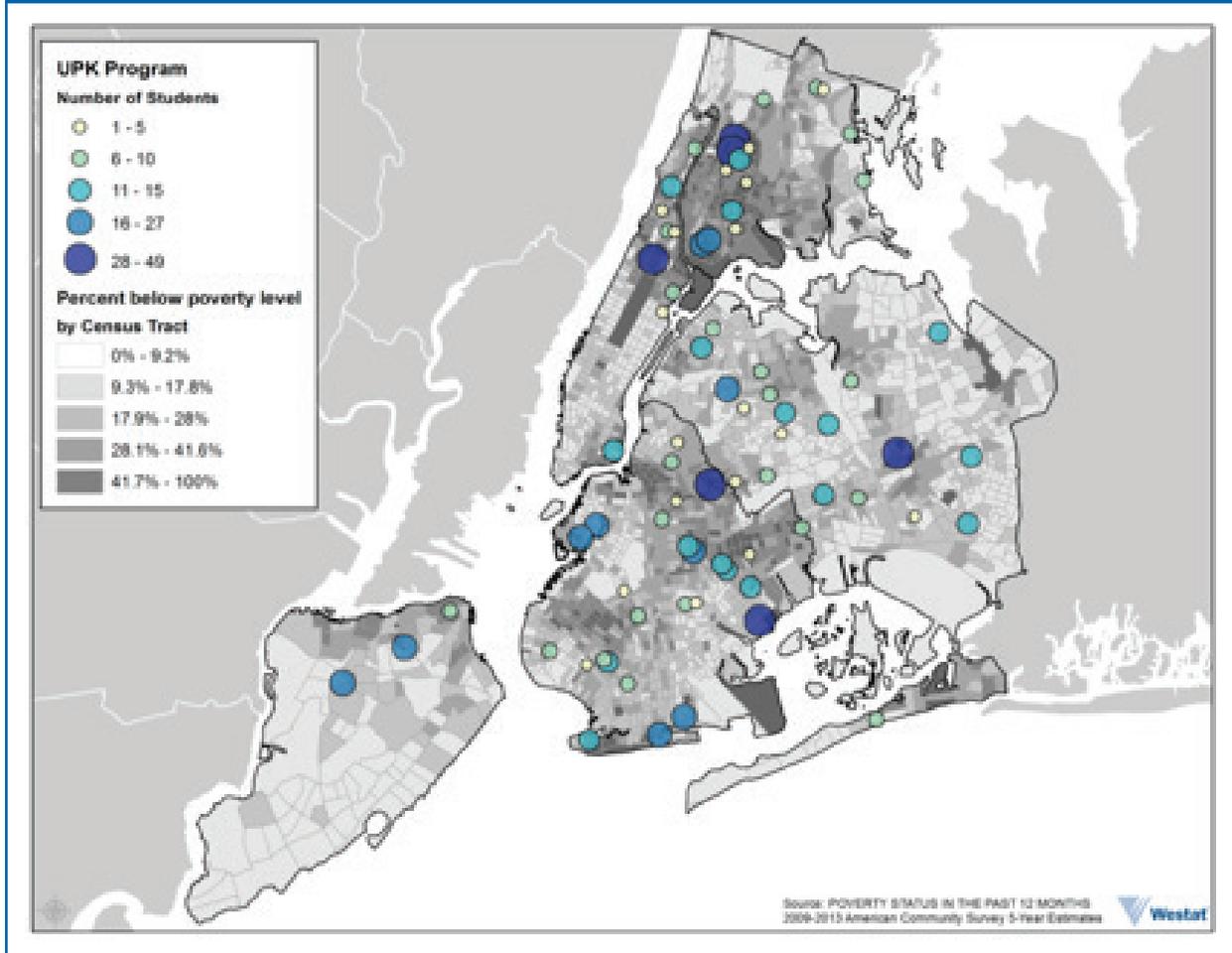
* NYCEECs are New York City Early Education Centers

** Includes 1 charter school

Figure 1.1. shows the geographic distribution and size of the participating sites across the five boroughs. Moreover, the figure also illustrates site location by neighborhood poverty, with the darker gray areas indicating higher proportions of households living below the poverty level.

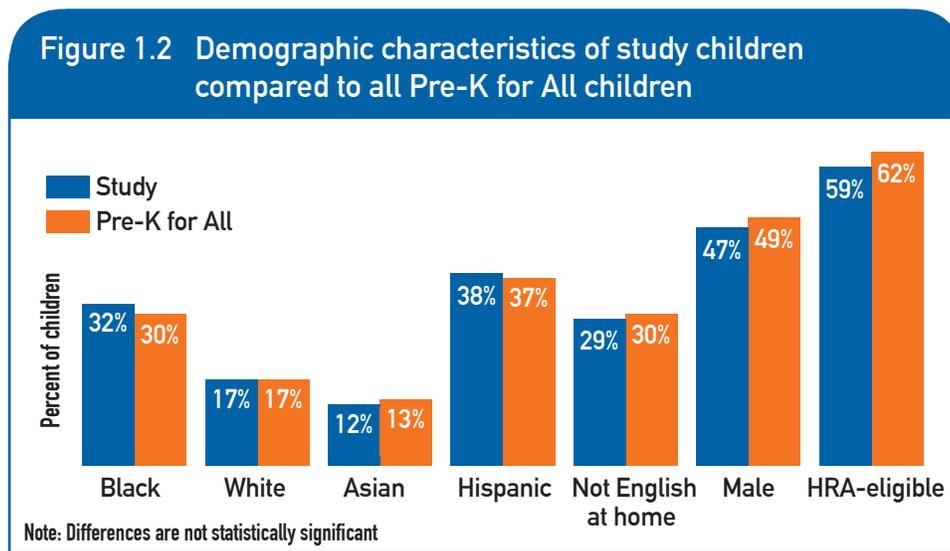
⁶201 sites participated in the implementation portion of the study.

Figure 1.1 Location and size of participating sites



Children

The 1,145 children who participated in our study⁷ were demographically equivalent to other children enrolled in the Pre-K for All program in terms of race/Ethnicity, home language⁸, gender, and HRA eligibility⁹. Figure 1.2 presents the demographic characteristics of the children in our study compared to all Pre-K for All children. The two groups of children were statistically equivalent across all demographic variables, which allows study findings to be generalized to the larger Pre-K for All population.



It is important to remember that the demographic categories displayed in Figure 1.2 are not mutually exclusive. For example, of the children whose home language was a language other than English, 60 percent were Hispanic, while another 25 percent were Asian. Figures 1.3 and 1.4 present the proportions of children in each racial/ Ethnic group whose home language was a language other than English and who lived in households eligible for HRA services.

⁷Active parental consent was obtained before children could participate in the study.

⁸Home language is a binary variable for which children either lived in an English-speaking household, or in a household where the primary language spoken at home was a language other than English. We do not know the level of English proficiency of these children, as that is not officially assessed until they are in Kindergarten or first grade.

⁹Students are HRA eligible if they qualify for free meals based on their families' receipt of Temporary Assistance to Needy Families (TANF) or Food Stamp benefits.

Figure 1.3
The proportion of children whose home language was a language other than english, by race/Ethnicity

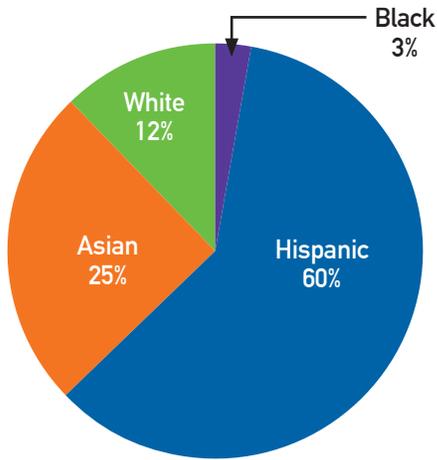
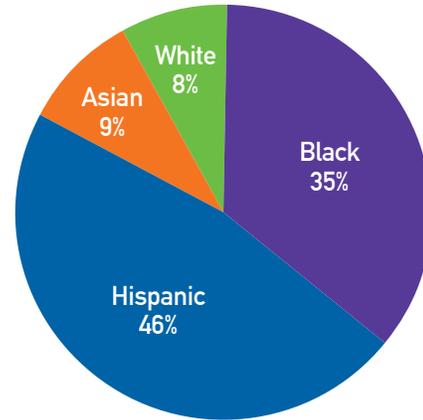


Figure 1.4
The proportion of children living in households that were eligible for HRA services, by race/Ethnicity



Considerations

As with any study, this study has a number of considerations that must be acknowledged. These are divided into two categories: considerations associated with the design itself, and considerations associated with the data.

Design Considerations

This study features a **pre-post design without a comparison group**¹⁰, which means that observed child growth cannot be attributed solely to participation in Pre-K for All. Children naturally learn and grow over time, and the study design means that this growth is confounded with the effects of Pre-K for All and other factors. Therefore, we cannot estimate the extent to which Pre-K for All was responsible for the children's learning and growth.

Another consideration is that the timing of testing may have influenced growth. Three aspects are important here. First, the study was conducted in the **program's inaugural year** in order to give the city a snapshot of student learning to improve the services provided. Second, the **average time between the fall and spring testing** was five and a half months, which is less than a full school year. As a result of

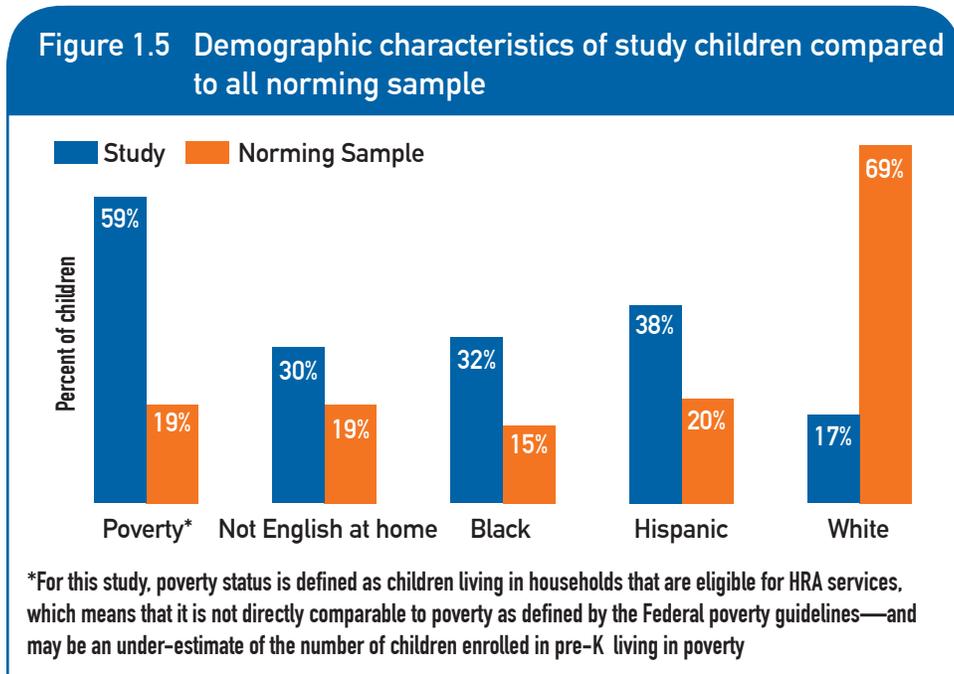
¹⁰A pre-post design examines whether participants in a program improve or become worse off during the course of the program, but any such improvement or deterioration cannot be directly attributed to the program.

these two aspects, the findings likely do not reflect the effects of a full year of participation in Pre-K for All on children’s academic and executive functioning skills. Third, the **pretests were administered after the start of the school year** (November/December), so children had already been enrolled in the program for a few months prior to taking the pretest. This might have affected children’s performance on the pretests.

Data Considerations

Although study sites and children were demographically representative of the Pre-K for All population, study participation was not mandatory. Site participation was voluntary and a small incentive was provided. Similarly, active parental consent was obtained for all assessed children. As a result, those sites and parents who opted not to participate in the study may have been different in some unmeasured way from those who agreed to participate.

Also, while the study children were representative of the population of children enrolled in Pre-K for All, the **demographic characteristics of children in the study were significantly different from those of the Woodcock-Johnson national norming sample**. The Woodcock-Johnson was normed on 8,818 children and adults (4,783 in grades kindergarten through 12) in a well-designed, nationally representative sample. While this sample accurately reflects the nation as a whole, it does not reflect the specific demographics of New York City, as shown in Figure 1.5. As such, readers should use caution when interpreting the performance of study children compared to the Woodcock-Johnson national norming sample.



Although the norming sample was not reflective of the demographics of the children tested for this study, the reliability of the Woodcock-Johnson is not called into question. In fact, test-retest reliability is reported to be around 0.9¹¹ for all subtests selected for this study (Woodcock, McGrew & Mather, 2001). Moreover, the Woodcock-Johnson has been found to be free of cultural and racial bias (Edwards & Oakland, 2006; McGrew & Woodcock, 2001). Lastly, the Woodcock-Johnson is widely used in other jurisdictions where the characteristics of the tested population diverge from that of the norming sample, with the Boston and Tulsa pre-K program evaluations being just two examples.

That said, it is important to point out that the Woodcock-Johnson **norming sample is more than two decades old**, having first been collected in 1996 and re-weighted in 2005. The age of the norming sample can distort the interpretation of results because people's performance on achievement tests has been found to improve over time (Silverstein & Nelson, 2000). That is, 4-year-olds in 2015 can be expected to score higher on a test that was normed 20 years ago than one that was normed more recently – this is called the “Flynn Effect” (Flynn, 2009).

As with any pre-post study, there was some **attrition** from fall to spring (12 percent) and some **missing data**. In cases where we had spring test scores but were missing fall scores,¹² we used multiple imputation, a statistical technique, that allowed us to infer missing fall scores.¹³ Further, we analyzed the data with and without the statistical adjustment for missing data and found no significant differences in the results.

Lastly, the **sample size** for the subgroup analyses was considerably smaller than the total sample size of 1,145 children. With smaller samples, descriptive statistics become less reliable. Therefore, readers should use caution when interpreting subgroup results.

¹¹Test-retest reliability coefficients range from 0 to 1, and they indicate the degree to which the results of a given test are consistent over time. A reliability coefficient of 1 means that the same results would be expected each time a test was taken.

¹²Between 35 and 54 cases, depending on the subtest.

¹³Multiple imputation is consistent with best practice in the field (Graham, 2009).

¹⁴We conducted a sensitivity analysis, analyzing the data with and without the imputed cases, and found no significant differences in the results.

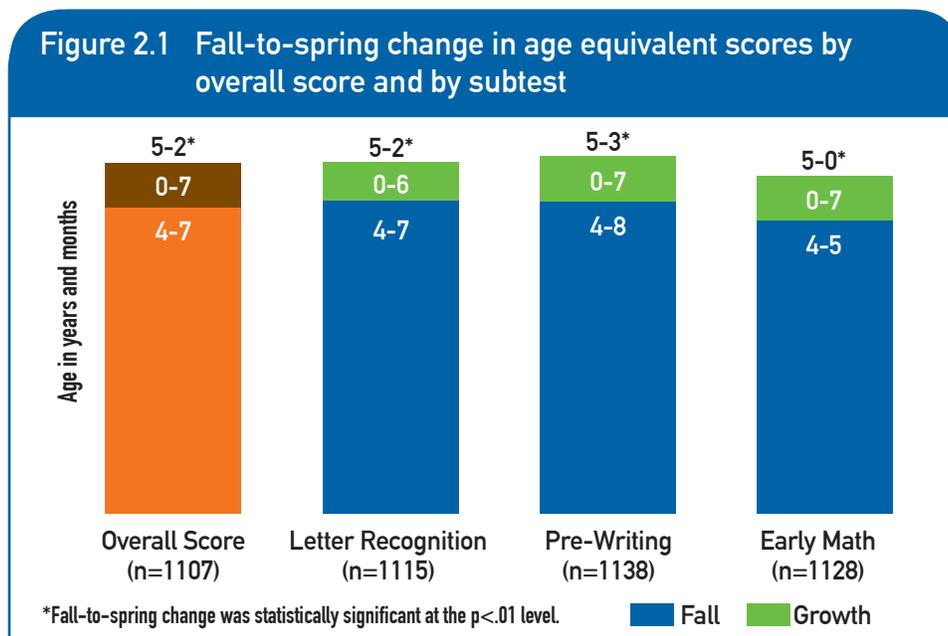
Findings

In this chapter we report the overall growth of Pre-K for All children in the tested academic skills (Letter Recognition, Pre-Writing, and Early Math) compared to their expected age-based level of development. The chapter concludes with the findings broken out by subgroup. Before presenting the findings, it is important to note that the Woodcock-Johnson provides a variety of score options for interpreting performance. These scores include: (1) degree of proficiency (W scores), (2) comparison with peers (scale scores), and (3) the level of development (age equivalent scores). In this report, we present the findings using age equivalent scores, while the W and scale score findings are presented in Appendix B. Age equivalent (AE) scores are highlighted in this chapter because they provide an “estimate of the chronological age at which typically developing children demonstrate the skills exhibited by the child being assessed” (Gengoux, 2013). For example, if Michael’s chronological age is 4 years and 8 months (4-8) at the time of testing, and he receives an age-equivalent score of 5-2, his performance is comparable to that of an average child of 5 years and 2 months (Jaffe, 2009). According to the assessment author, these scores can be used for forming recommendations regarding instructional level and materials, and can also be used in placement decisions (Jaffe, 2009).

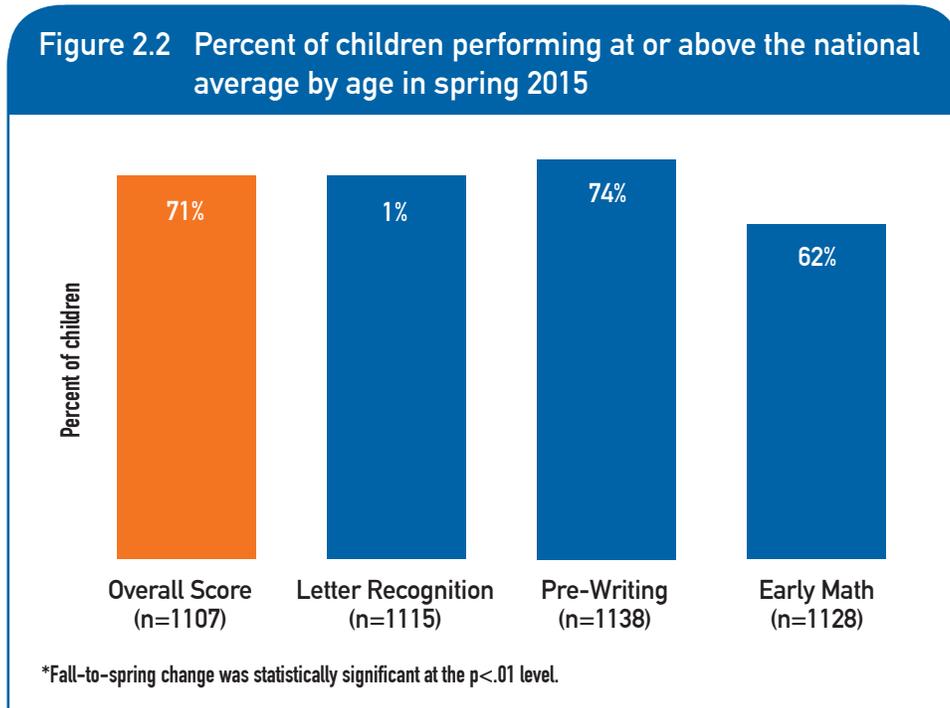
2.1: Overall findings

Early Literacy and Math

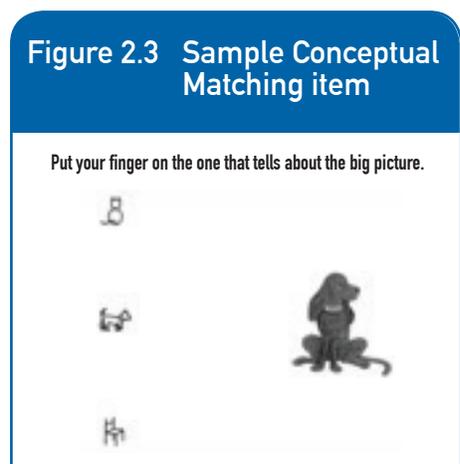
Children in the study showed statistically significant fall-to-spring gains in all tested academic skills. Looking at the Overall Score in Figure 2.1, in the fall, these children performed at the level of an average child of 4 years and 7 months. At the time of the spring assessment, children attending Pre-K for All performed at the level of an average child of 5 years and 2 months. That is, over the five-and-a-half months between the fall and spring assessments, study children gained an average of seven months of learning—a month-and-a-half more than is expected based on the national norming sample.



Looking at this information from another perspective, we found that by the spring over 70 percent of Pre-K for All children in the study performed at or above the age-based national averages in early literacy¹⁷ (as shown in Figure 2.2). The corresponding percentage for Early Math was 62 percent.



Supplementary Subtest



In addition to the early literacy and early math skills described above, we administered a supplementary subtest to obtain information on Conceptual Matching skills (assessed using the Passage Comprehension subtest). This subtest measures the extent to which children can match conceptually similar pictures, in which one picture is more realistically portrayed and the other is more abstract. A sample item is presented in Figure 2.3. This subtest is not as widely used to assess pre-K children as the others, but it is statistically significantly correlated with the other three subtests used in this study. This means that there is a positive linear relationship between the conceptual matching subtest and the other subtests, so that if a child did well in one subtest, she could be expected to do well in the other subtests.

¹⁷Using the age-equivalent scores. This is the same methodology used by the University of Minnesota's study of pre-K in Chicago (Reynolds et al., 2014).

For this subtest, in the fall, children in the study performed at the level of a child of 4 years and 7 months. Over the five-and-a-half months between the fall and spring assessments, these children gained an average of four months of learning—a statistically significant gain at the $p < .05$ level. However, this gain is a month-and-a-half less than might be expected based on the national norming sample.

In the next section findings are broken out by race/Ethnicity, home language, and HRA eligibility. It is important to note that the subgroup results should be interpreted with caution due to small sample sizes in many cases.

2.2: Subgroup findings

Race/Ethnicity

The tables presented in this section provide two pieces of information. They first compare the growth of children in each racial/Ethnic subgroup to that of the White children in our sample. They next compare the fall-to-spring growth of each subgroup to that of the norming sample, to see if the growth is greater or less than what is expected for the nation as a whole.

In **Letter Recognition**, White children performed at the level of an average child of 4 years and ten months in the fall. While Asian and Black children performed similarly to White children, Hispanic children began the school year statistically significantly behind the other subgroups. In the spring, Hispanic children still scored statistically significantly lower than White children (see Table 2.1). There were no other statistically significant differences for Asian or Black children when compared to White children.

Although Hispanic children scored significantly lower than White children in both the fall and spring, their fall-to-spring growth was noteworthy. They grew by 9 months, which was 3.5 months more than expected given the timeframe of test administrations. In contrast, White children grew one-half month more than might be expected, while Asian and Black children did not meet expectations for growth. All subgroups of children experienced statistically significant fall-to-spring growth, at the $p < .01$ level.

Table 2.1 Age equivalent-score comparisons by racial and Ethnic group in Letter Recognition

Letter Recognition	n	Age: Years–Months		Fall-to-spring growth	Was growth greater than expected?^
		Fall	Spring		
White	162	4-10	5-4	6 months	Yes: 0.5 months
Asian	114	5-0 (NS)	5-3 (NS)	3 months	No
Black	296	5-0 (NS)	5-5 (NS)	5 months	No
Hispanic	350	4-2 (Sig)	4-11 (Sig)	9 months	Yes: 3.5 months

NS = not significantly different from the score obtained by White children

Sig = statistically significantly different from the score obtained by White children

^5.5 months is the elapsed time between the fall and spring test administrations. Growth that is larger than 5.5 months exceeds national expectations based on the Woodcock-Johnson norming sample

In **Pre-Writing**, Hispanic children again scored statistically significantly lower than White children in the fall, and the gap remained statistically significant in the spring (see Table 2.2). There were no statistically significant differences for Black or Asian children when compared to White children. Unlike in Letter Recognition, all racial and Ethnic subgroups exhibited greater than expected fall-to-spring growth in Pre-Writing. Asian children exhibited the most fall-to-spring growth, with nearly double the expected 5.5 months of growth. All subgroups of children experienced statistically significant fall-to-spring growth, at the $p < .01$ level.

Table 2.2 Age equivalent-score comparisons by racial and Ethnic group in Pre-Writing

Pre-Writing	n	Age: Years-Months		Fall-to-spring growth	Was growth greater than expected?^
		Fall	Spring		
White	165	4-8	5-5	9 months	Yes: 3.5 months
Asian	114	4-9 (NS)	5-7 (NS)	10 months	Yes: 4.5 months
Black	304	4-10 (NS)	5-5 (NS)	7 months	Yes: 1.5 months
Hispanic	360	4-6 (Sig)	5-1 (Sig)	7 months	Yes: 1.5 months

NS = not significantly different from the score obtained by White children

Sig = statistically significantly different from the score obtained by White children

^5.5 months is the elapsed time between the fall and spring test administration. Growth that is larger than 5.5 months exceeds expectations based on the Woodcock-Johnson norming sample

In **Early Math**, White children had statistically significantly higher age equivalent scores than all other racial/Ethnic groups in both the fall and spring (see Table 2.3). However, looking at the fall-to-spring growth, both Asian and Hispanic children grew more (8 and 7 months, respectively) than White children (6 months). Black children grew only slightly less than might be expected based on the norming sample. All subgroups of children experienced statistically significant fall-to-spring growth, at the $p < .01$ level.

Table 2.3 Age equivalent-score comparisons by racial and Ethnic group in Early Math

Early Math	n	Age: Years-Months		Fall-to-spring growth	Was growth greater than expected?^
		Fall	Spring		
White	164	4-10	5-4	6 months	Yes: 0.5 months
Asian	113	4-5 (Sig)	5-1 (Sig)	8 months	Yes: 2.5 months
Black	301	4-7 (Sig)	5-0 (Sig)	5 months	No
Hispanic	357	4-2 (Sig)	4-9 (Sig)	7 months	Yes: 1.5 months

NS = not significantly different from the score obtained by White children

Sig = statistically significantly different from the score obtained by White children

^5.5 months is the elapsed time between the fall and spring test administration. Growth that is larger than 5.5 months exceeds expectations based on the Woodcock-Johnson norming sample

Home Language

In the fall, across all tested academic skills, children whose home language was a language other than English¹⁸ had significantly lower age-equivalent scores than children whose home language was English. In the spring, this gap remained statistically significant (see Table 2.4). However, the fall-to-spring growth was greater than might be expected for both groups across all three content areas, with children whose home language was a language other than English growing more than children whose home language was English in two of the three content areas (Letter Recognition and Early Math). Both subgroups of children experienced statistically significant fall-to-spring growth, at the $p < .01$ level.

Table 2.4 Age equivalent-score comparisons by home language in all tested academic skills

Letter Recognition	n	Age: Years–Months		Fall-to-spring growth	Was growth greater than expected?^
		Fall	Spring		
<i>Home Language was English</i>	678	4-10	5-4	6 months	Yes: 0.5 months
Home language was a language other than English	261	4-4 (Sig)	5-0 (Sig)	8 months	Yes: 2.5 months
Pre-Writing					
<i>Home Language was English</i>	693	4-9	5-5	8 months	Yes: 2.5 months
Home language was a language other than English	268	4-6 (Sig)	5-1 (Sig)	7 months	Yes: 1.5 months
Early Math					
<i>Home Language was English</i>	688	4-8	5-2	6 months	Yes: 0.5 months
Home language was a language other than English	264	4-0 (Sig)	4-7 (Sig)	7 months	Yes: 1.5 months

NS = not significantly different from the score obtained by children whose home language was English

Sig = statistically significantly different from the score obtained by children whose home language was English

^5.5 months is the elapsed time between the fall and spring test administration. Growth that is larger than 5.5 months exceeds expectations based on the Woodcock–Johnson norming sample

¹⁸While the home language was a language other than English, we do not know the extent to which the children themselves spoke English.

HRA Eligibility

When we look at the fall performance of children who lived in households that were eligible for HRA services (see Table 2.5 below), we see that they scored significantly behind their peers living in households that were not eligible for HRA services in all tested academic skills. Further, the gap between HRA-eligible and non-HRA-eligible children remained significant in the spring. Again, the fall-to-spring growth was greater than might be expected for both groups, with HRA-eligible children making greater growth than non-HRA-eligible children in Early Math. Both subgroups of children experienced statistically significant fall-to-spring growth, at the $p < .01$ level.

Table 2.5 Age equivalent-score comparisons by HRA eligibility in all tested academic skills

Letter Recognition	n	Age: Years-Months		Fall-to-spring growth	Was growth greater than expected?^
		Fall	Spring		
<i>NOT eligible for HRA</i>	389	4-11	5-5	6 months	Yes: 0.5 months
HRA-eligible	550	4-7 (Sig)	5-1 (Sig)	6 months	Yes: 0.5 months
Pre-Writing					
<i>NOT eligible for HRA</i>	397	4-9	5-4	7 months	Yes: 1.5 months
HRA-eligible	564	4-8 (Sig)	5-3 (Sig)	7 months	Yes: 1.5 months
Early Math					
<i>NOT eligible for HRA</i>	688	4-8	5-2	6 months	Yes: 0.5 months
HRA-eligible	264	4-0 (Sig)	4-7 (Sig)	7 months	Yes: 1.5 months

NS = not significantly different from the score obtained by children who lived in households that were not eligible for HRA services

Sig = statistically significantly different from the score obtained by children who lived in households that were not eligible for HRA services

^5.5 months is the elapsed time between the fall and spring test administration. Growth that is larger than 5.5 months exceeds expectations based on the Woodcock-Johnson norming sample

Executive Functioning Results

Authors: Cybele Raver and Pamela Morris, New York University

A growing empirical research base clearly documents the relevance of executive functioning to short and long term educational outcomes (Hamre & Pianta, 2001; Rimm-Kaufmann et al., 2009), including Kindergarten readiness (Blair & Razza, 2007). “Executive Function” is a term used to describe key higher-order thinking skills (including the ability to delay or inhibit behavior and the ability to retain and manipulate distinct pieces of information) that serve as the cornerstone of children’s self-regulation. For this study, executive functioning skills are measured by two short assessments, the Pencil Tap and Hearts and Flowers.

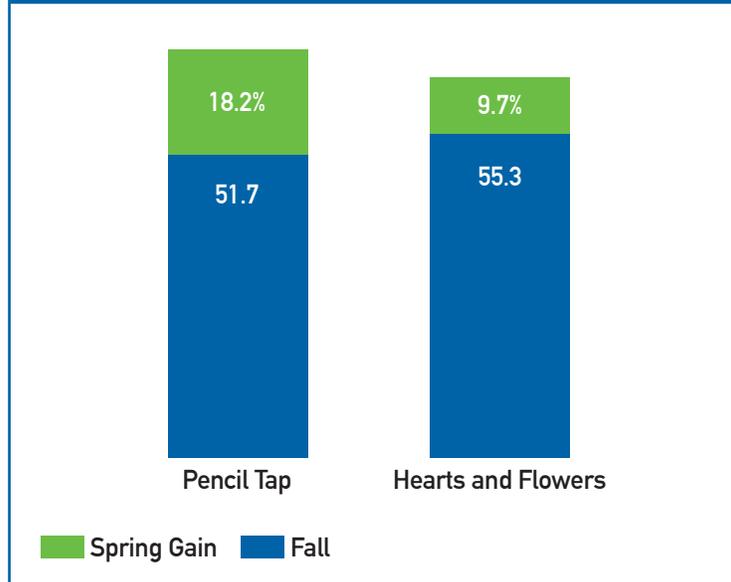
The Pencil Tap measure (Diamond & Taylor, 1996) requires children to delay or inhibit behavior according to a pre-established rule. For example, when the assessor taps her pencil (eraser side down) on the table two times (tap, tap), the child is supposed to tap his pencil on the table one time (tap). When the assessor taps her pencil on the table one time (tap), the child is supposed to tap his pencil two times (tap, tap). Children completed a total of 16 trials in the Pencil Tap task and scores represent the percentage of correct trials.

The Hearts and Flowers task (Diamond et al., 2007) also measures children’s inhibitory control and working memory, using a touch-screen task administered on tablet PCs. Like the Pencil Tap, children have to inhibit behavior according to pre-established rules. However, this task is more complex, as there are two rules that children have to follow (one for Heart stimuli and one for Flower stimuli). As such, children complete three phases: (a) congruent, in which they follow the Hearts rule; (b) incongruent, in which they follow the Flowers rule; and (c) mixed, in which children must follow both the Hearts and the Flowers rules. The Hearts and Flowers task measures two components of children’s performance including speed and accuracy. We report accuracy rather than speed because, for young children, accuracy is the more sensitive measure (Diamond & Kirkham, 2005). The reported scores in the Hearts and Flowers task represent the percentage of correct trials.

The fall-to-spring gain in the percentage of correct responses to the Pencil Tap measure is provided in Figure 6.1. The nearly 20 percent increase in children’s average correct performance is statistically significant. This suggests that, on average, children shifted from limited understanding of and performance on the task to greater mastery of these key self-regulatory skills (inhibiting their impulses and remembering and using the rules of the “game”) over the course of the study period.

On the Hearts and Flowers task, children also demonstrated substantial improvement. For example, in the fall, most students were able to demonstrate basic understanding of the rules of the task, but just over half of students were able to perform at higher levels of proficiency, which required greater impulse control, memory, and higher-order thinking skill. By spring, children had gained substantial levels of proficiency in mastering the more complex levels of the task (See Figure 3.1). For example, over three-quarters of children who were assessed could remember and use more the complex rules of the task (mastering the “incongruent” trials) and the majority of children (64 percent) could use impulse control and memory to perform well on the “mixed” trials of the task.

Figure 3.1 Executive Function, percentage correct



These findings are similar to results reported in other studies, including the Boston pre-K evaluation, with the majority of 3- and 4-year-old children demonstrating similar levels of working memory, inhibitory control, and cognitive flexibility when given the opportunity to enroll in pre-Kindergarten (Raver et al., 2011; Weiland et al., 2013).

Summary

This report describes the growth of a sample of children who participated in the Pre-K for All program, outlining the academic and executive functioning skills they had when assessed in the fall of 2014, as well as their growth in these skills over a five-and-a-half month period of time.

The literature has demonstrated that children who attend formal pre-K programs have higher achievement test scores¹⁹ at kindergarten entry than demographically similar children who did not attend pre-K (Rathbun & Zhang, 2016). And these effects have been found to be larger among children from economically disadvantaged backgrounds (Camilla, Vargas, Ryan, & Barnett, 2010; Magnuson, Ruhm & Waldfogel, 2007). While the current study was not designed to replicate those results, the children participating in Pre-K for All also showed meaningful and important gains in academic and executive functioning skills over the course of a five-and-a-half month study period.

A sample of 1,145 children attending Pre-K for All were assessed in the fall of the program's inaugural year. These assessment scores were within the 'normal' range and were consistent with findings from studies of other pre-K programs across the country (Barnett, 2013; Gormley, Gayer, Phillips, & Dawson, 2005; Hustedt, Barnett, Jung, & Figueras, 2008; Weiland & Yoshikawa, 2013). While comparisons to these other studies are not perfect, they provide a framework for interpreting the findings from this study. For example, as seen in Figure 4.1, compared to other studies that administered the Woodcock-Johnson Early Math (Applied Problems) subtest, the average fall score for Pre-K for All is towards the lower end of the distribution.

The fact that [Pre-K for All] is universal, I believe, lifts all boats. It means children of all backgrounds learn together



and, in many cases, come together in a school setting that brings them together like no other part of their life does. And all children benefit.

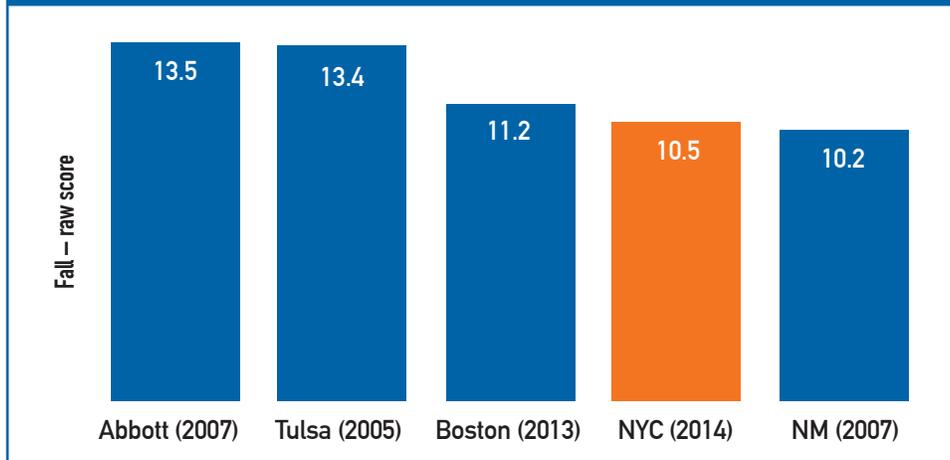
-Mayor de Blasio on National Public Radio

¹⁹Using a wide variety of assessment instruments, including the Woodcock-Johnson

²⁰Defined as a scale score between 85 and 115.

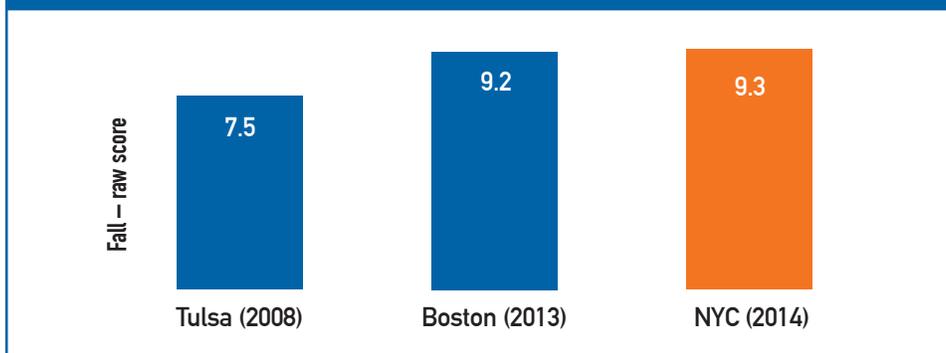
²¹Comparisons are imperfect for a number of reasons: populations may be different demographically, assessments were given in different years, and results are presented in different metrics. For the comparisons presented in Figure 4.1, assessments for all studies were administered to 4 year olds in the fall of the given year.

Figure 4.1 Pre-K for All average fall assessment score in Early Math, compared to other studies

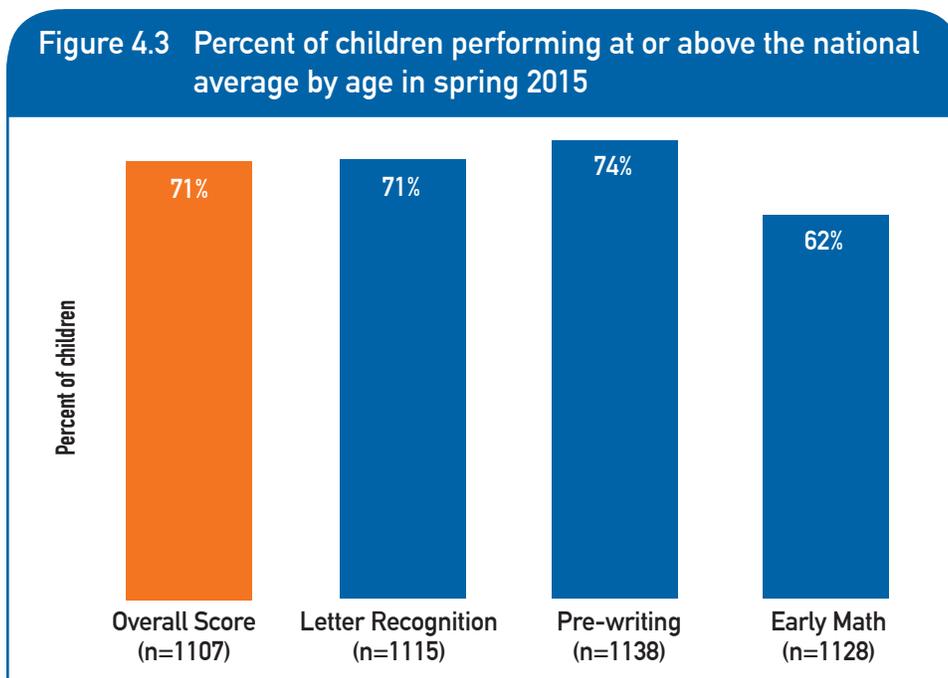


Looking at the studies that administered the Letter Recognition subtest, the fall average raw scores for Pre-K for All children were very similar to those found in Boston (Raver et al., 2011; Weiland & Yoshikawa, 2013), whose sample was demographically similar to that of New York (see Figure 4.2).

Figure 4.2 Pre-K for All fall assessment scores in Letter Recognition, compared to other studies



Pre-K for All children in the study made statistically significant gains across all academic skills (Letter Recognition, Pre-Writing, and Early Math) over the course of a five-and-a-half month testing period. By the spring, children were still classified as being in the ‘normal’ range²² in all of the academic content areas assessed, and were performing at the level of five year-olds nationally. Moreover, as shown in Figure 4.3, **71 percent of children were performing at or above the national average** on the Overall Score (a composite of the three subtests). These are all important findings, given that the population of children in this study (and in New York City as a whole) is more diverse than the norming sample—with higher percentages of Black, Hispanic and Asian children, children whose home language was a language other than English, and children who lived in households that were eligible for HRA services.



Among the different subgroups of children, Hispanic children, children whose home language was a language other than English²³, and children living in households eligible for HRA services scored significantly behind their peers when tested in the fall. While these subgroups of children made statistically significant progress by the spring and were still performing within the “average” range nationally, these disparities remained statistically significant.

²²Defined as a scale score between 85 and 115.

²³It is important to remember that of the children whose home language is a language other than English, 60 percent are Hispanic, which may explain why these two groups score similarly.

Growth was also observed in the two assessments of executive functioning. The gains in the percentage of correct responses to the Pencil Tap and Hearts and Flowers tasks were 10 percent and 18 percent, respectively. While these gains were also confounded with maturation effects, other studies have typically found executive functioning gains to be non-significant or very small (Gormley et al., 2011; Magnuson et al., 2007; Weiland & Yoshikawa, 2013).

It is important to remind the reader that the performance of study children on these assessments of academic and executive functioning skills cannot be solely attributed to their participation in Pre-K for All. Because the study design did not include a comparison group, we cannot know how much children might have grown in these skills had they not participated in the program. Moreover, comparisons based on the Woodcock-Johnson norming sample should be viewed descriptively and interpreted with caution.

As one of a series of seven reports describing the first year of Pre-K for All, this report focused on how children participating in Pre-K for All grew during a five-and-a-half month testing period in the 2014-15 school year. We found that children performed very well when initially tested in the fall and made statistically significant and meaningful gains over time. As a snapshot of student learning, these findings serve as an initial guidepost into the City's efforts and will inform the City's commitment to continuously improve program quality.



References

- Blair, C. & Razza, R.P. (2007) Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, 78(2), 647-663.
- Bradley-Johnson, S., (2004). Review of the Woodcock-Johnson Test of Achievement: Third Edition. *Journal of Psychoeducational Assessment*, 22(3).
- Camilla, G., Vargas, S., Ryan, S., & Barnett, W.S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, 112, 579-620.
- Campbell, F.A., Ramey, C. T., Pungella, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science*, 6, 42-57.
- Cizek, G. J., & Sandoval, J. (2003). Review of the Woodcock-Johnson III. In B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), *The fifteenth mental measurements yearbook* (pp. 1019-1028). Lincoln, NE: Buros Institute of Mental Measurements.
- Clark, C.A. C., Pritchard, V. E., Woodward, L.J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology*, Vol 46(5), 1176-1191.
- Diamond A., Barnett S., Thomas J., Munro S. (2007). Preschool program improves cognitive control. *Science* 318, 1387-1388.
- Diamond, A., & Kirkham, N. (2005). Not Quite as Grown-Up as We Like to Think: Parallels Between Cognition in Childhood and Adulthood. *Psychological Science*, 16:291
- Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the ability to remember what I said and to “do what I say, not what I do”. *Developmental Psychology*, 29, 315-334.
- Duncan, G.J., Claessens, A., Huston, A.C., Pagani, L.S., Engel, M., Sexton, H., et al. (2007). School readiness and later achievement. *Development Psychology*, 43, 1428-1446.
- Edwards, O. and Oakland, T. (2006). Factorial Invariance of Woodcock-Johnson III Scores for African Americans and Caucasian Americans. *Journal of Psychoeducational Assessment* Volume 24 Number 4 December 2006 358-366.
- Flynn, James R. (2009). *What Is Intelligence: Beyond the Flynn Effect* (expanded paperback ed.). Cambridge: Cambridge University Press.
- Gengoux, G. (2013). *Age Equivalents*. Springer: NY. Accessed online on February 12, 2016: http://rd.springer.com/referenceworkentry/10.1007%2F978-1-4419-1698-3_282.
- Gormley, W.T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal preK on cognitive development. *Developmental Psychology*, 41(6), 872-884
- Gormley, W.T., Phillips, D., Newmark, K., Perper, K., & Adelstein, S. (2011). *Social-emotional effects of early childhood education programs in Tulsa*. Center for Research on Children in the United States.
- Graham, J.W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549-576.
- Hamre, B.K. & Pianta, R.C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development* 72(2):625-38.
- Heckman, J.J., Moon, S.H., Pinto, R., Savellyev, P.A., & Yavitz, A. (2010). The rate of return to the High Scope Perry Preschool Program. *Journal of Public Economics*, 94, 114-128.

-
- Hustedt, J.T., Barnett, W.S., Jung, K., & Figueras, A. (2008). Impacts of New Mexico PreK on Children's School Readiness at Kindergarten Entry: Results from the Second Year of a Growing Initiative. The National Institute for Early Education Research Rutgers University
- Jaffe, L.E. (2009) Woodcock-Johnson Publisher documentation. Itasca IL: Riverside Publishing.
- Magnuson, K.A., Ruhm, C., & Waldfogel, J. (2007b). Does Pre-Kindergarten improve school preparation and performance? *Economics of Education Review*, 26, 33-51.
- McClelland, M.M., Acock, A.C., & Morrison, F.J., (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, 21 471-490.
- McGrew, K.S., & Woodcock, R.W. (2001). *Technical Manual: Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.
- National Early Literacy Panel (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington DC: National Institute for Literacy.
- Rathbun, A., & Zhang, A. (2016). *Primary Early Care and Education Arrangements and Achievement at Kindergarten Entry* (NCES 2016-070). National Center for Education Statistics, U.S. Department of Education. Washington, DC.
- Raver, C. C., Jones, S. M., Li-Grining, C. P., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's impact on low-income preschoolers' pre-academic skills: Self-regulation and teacher-student relationships as two mediating mechanisms. *Child Development*, 82(1), 362–378.
- Reynolds, A.J., Temple, J.A., White, B., Ou, S., & Robertson, D.L. (2011). Age-26 cost benefit analysis of the Child-Parent Center Early Education Program. *Child Development*, 82, 379-404.
- Reynolds A.J., Richardson B.A., Hayakawa M., et al. (2014). Association of a full-day vs part-day preschool intervention with school readiness, attendance, and parent involvement. *Journal of the American Medical Association*; 312(20):2126-2134.
- Rimm-Kaufman, S.E.; Curby, T.W.; Grimm, K. J.; Nathanson, L.; Brock, Laura L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology*, 45(4), 958-972.
- Silverstein, M. L., & Nelson, L. D. (2000). Clinical and Research Applications of Revising Psychological Tests. *Psychological Assessment*, 298-303.
- Welsh, J.A., Nix, R.L., Blair, C., Bierman, K.L., Nelson, K.E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology*, Vol 102(1), 43-53.
- Weiland, C. & Yoshikawa, H. (2013). Impacts of a Pre-Kindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, Vol 84, No 6, pp. 2112-2130.
- Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public Pre-Kindergarten program, *Early Childhood Research Quarterly*, 28: 199-209.
- Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson test of Achievement*. Ithaca NY: Riverside.

Assessment Details

The Woodcock-Johnson provides a variety of score options for interpreting performance. These include the standard score (SS), age equivalents (AE), raw score, and the W score. The raw score is converted into a W score, which is the foundational index on which the SS and AE scores are based. The W score is a measure of an individual's level of proficiency on the test (W ability) compared to the level of difficulty of the test, and is a transformation of the Rasch analytic model using item response theory. In calculating the W score, the median ability level for each age group on a test is calculated. This corresponds to the difficulty level at which 50 percent responded correctly and 50 percent responded incorrectly. Thus the median W ability represents the average difficulty level that each group can manage—this is defined as the reference W. This reference W is the score against which a W ability is compared. The difference between an individual's W ability and their peer group reference W is termed the W difference. The W difference is the value from which the age equivalent (AE) scores are derived. Most importantly, the W score puts the raw score on an equal interval scale. This characteristic allows the differences between two sets of scores situated anywhere along the scale to be compared. Equal interval scores are the most appropriate for statistical calculations. Because the W score is an equal-interval scale, it is particularly useful for reporting an individual's growth. An increase in a person's W ability represents actual growth in the measured skill. The W scale is constructed so that an increase in 10 W units represents the individual's ability to perform with 75 percent success tasks that he or she had previously performed with 50 percent success. This is true for any 10 point increase on the W scale regardless of what is being measured or the difficulty level of the task (Woodcock, 1999). Accordingly, if a person's ability increases from one time to the next, the W score will increase also. This is not the case with AE scores. **Age equivalent scores describe the general level of development of a skill compared with others of the same age in the norming sample. If a person improves in a trait at the same rate as his or her peers, the AE will not change from one testing occasion to the next.** However, according to the assessment author (Jaffe, 2009), these scores can be used for forming recommendations regarding instructional level and materials, and can also be used in placement decisions based on a criterion of significantly advanced or delayed performance.

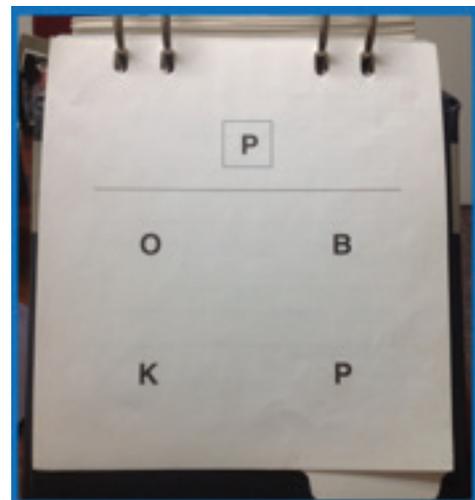
Next, examples of each of the subtests in the assessment battery are provided. First, we present the core subtests and then the supplemental subtest.

ASSESSMENT BATTERY

Letter Recognition (Letter-Word Identification)

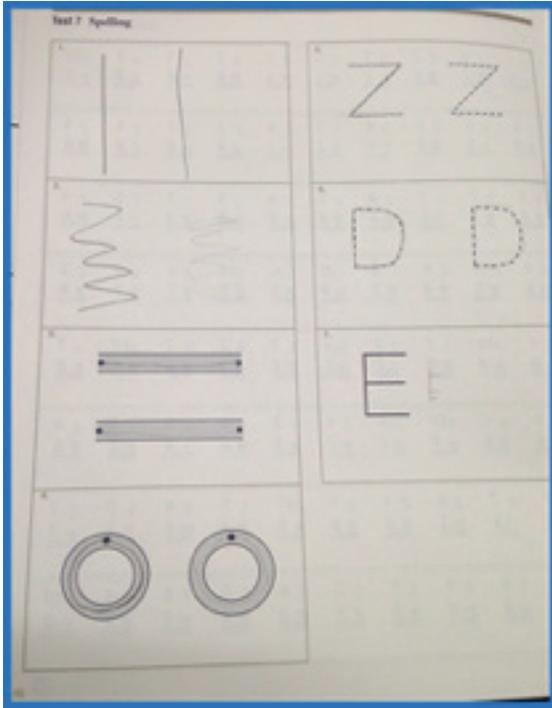
Letter-Word Identification assesses children's letter and word identification ability. Items include identifying and pronouncing letters and words presented to the child.

Sample instructions to assessor: Point to the letter at top of subject's page and say "This is letter P" Run your hand across the four letters and say "Find the 'P' down here"



Pre-Writing (Spelling)

Spelling assesses children's prewriting skills, such as drawing lines and tracing, writing letters, and spelling orally presented words.



Sample instructions to assessor: Say "Watch me." make single vertical pencil mark 1 to 2 inches long in left side of box for item 1. Hand pencil to subject and say "Now you do it right here" (point to the right side of the box for item 1). Collect pencil when subject has finished.

Early Math (Applied Problems)

Applied Problems assesses children's ability to solve numerical and spatial problems presented verbally with accompanying pictures of objects.



Sample instructions to assessor: Point to top picture and say "How many dogs are there in this picture?" Point to middle picture and say "How many crayons are there in this picture?"

SUBTEST CORRELATIONS

Pearson Correlation Sig (2-tailed) N	Letter Recognition	Pre-Writing	Early Math	Conceptual Matching
Letter Recognition	1 1,115			
Pre-Writing	.595** .000 1,111	1 1,138		
Early Math	.542** .000 1,111	.534** .000 1,124	1 1,128	
Conceptual Matching	.323** .000 1,106	.278** .000 1,131	.257** .000 1,120	1 1,132

** Correlation is significant at the $p < .01$ level ** Correlation is significant at the $p < .01$ level

Findings by Degree of Proficiency (W Score)

W scores represent a common scale of measurement that represents a child's ability to accurately complete the given task, given the level of difficulty of the task. An increase in a child's W score represents improved proficiency in the trait measured²⁴, making it particularly useful for reporting growth over time. **If a child's ability increases from one time to the next, the W score will also increase** (Woodcock & Dahl, 1971; Woodcock, 1999). This differentiates the W score from Peer Comparisons (Scale scores)—reported in appendix C, and the Age Equivalent (AE) scores—reported in Chapter 2.

In this appendix we report the proficiency scores of Pre-K for All children in the tested academic skills (Letter Recognition, Pre-Writing, and Early Math). The appendix concludes with the findings broken out by subgroups (race/Ethnicity, home language, and HRA eligibility status).

Example: Michael was tested in the Fall and again in the spring. Although his W score increased by 2 points, his scale score and AE score remained the same. Michael made progress, but he did so at the same rate as his peers of the same age. Thus, the scale and AE scores did not change. It is the W score that demonstrates the absolute change in proficiency.

B.1: Overall findings reported in proficiency (W) scores

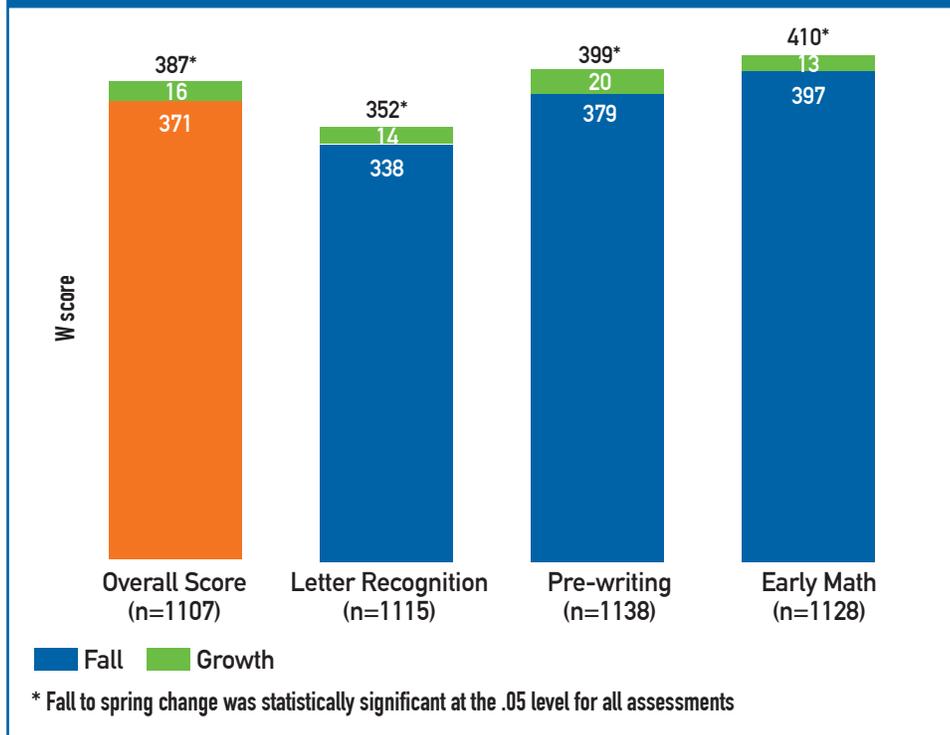
Children attending Pre-K for All made statistically significant²⁵ fall-to-spring gains in all tested academic skills (Letter Recognition, Pre-Writing, and Early Math). While children started with the highest proficiency levels in Early Math, the largest gains were in Pre-Writing (see Figure 1.1). The Overall Score (which is a composite of Pre-Writing, Letter Recognition, and Early Math scores) demonstrates that by the end of the 2014-2015 school year, Pre-K for All children made meaningful gains in proficiency.

Children attending Pre-K for All made statistically significant fall-to-spring gains in all tested academic skills

²⁴The W-J publisher uses item response theory to convert the raw score into an ability score (W score) (Hambleton and Swaminathan, 1991).

²⁵Statistically greater than zero, or what would have been expected due to chance alone.

Figure B.1 Change in W score from fall to spring in tested academic skills



²⁶Passage Comprehension assesses children's ability to match conceptually similar pictures with appropriate words that maintain the semantic properties of the stimulus.

B.2: Subgroup findings reported in proficiency scores

Race and Ethnicity

Hispanic children had **Letter Recognition** proficiency scores statistically significantly below that of White children in the fall and this gap remained significant in the spring (see Table B.1). There were no other statistically significant differences for Asian or Black children when compared to White children. Children of all races made statistically significant learning gains from fall to spring in Letter Recognition²⁷. Asian, Black, White, and Hispanic children gained 13, 12, 14 and 15 points, respectively.

Table B.1 W score comparisons by racial and Ethnic group in Letter Recognition

Letter Recognition	n	W Scores		Fall-to-spring growth	Was growth statistically significant?*
		Fall	Spring		
White	162	342	356	14	Yes
Asian	114	349 (NS)	361 (NS)	12	Yes
Black	296	347 (NS)	359 (NS)	12	Yes
Hispanic	350	330 (Sig)	345 (Sig)	15	Yes

NS = not significantly different from the score obtained by White children

Sig = statistically significantly different from the score obtained by White children

*Growth was statistically significantly greater than zero, measured at the .05 level

Looking at **Pre-Writing skills**, Black, Asian, and White children had statistically equivalent fall scores. Hispanic children entered Pre-K for All with Pre-Writing proficiency levels statistically significantly below that of White children and also ended the year with significantly lower proficiency in the subtest (see Table B.2). Children of all races made statistically significant learning gains from fall to spring²⁸. Asian children made the largest gains with 26 points, followed by White, Black, and Hispanic children with 23, 21, and 19 points, respectively.

²⁷The fall-to-spring gain was significantly greater than zero, and therefore unlikely to have occurred by chance alone.

²⁸The fall-to-spring gain was significantly greater than zero, and therefore unlikely to have occurred by chance alone.

Table B.2 W score comparisons by racial and Ethnic group in Pre-Writing

Pre-Writing	n	W Scores		Fall-to-spring growth	Was growth statistically significant?*
		Fall	Spring		
White	165	380	403	23	Yes
Asian	114	382 (NS)	408 (NS)	26	Yes
Black	304	384 (NS)	405 (NS)	21	Yes
Hispanic	360	374 (Sig)	393 (Sig)	19	Yes

NS = not significantly different from the score obtained by White children
 Sig = statistically significantly different from the score obtained by White children
 *Growth was statistically significantly greater than zero, measured at the .05 level

In **Early Math**, all non-White racial and Ethnic groups had proficiency levels statistically significantly below that of White children, and this gap remained in the spring (see Table B.3). Children of all races and Ethnicities made statistically significant learning gains from fall to spring²⁹. Asian children made the largest gains, increasing 15 points from fall to spring. Hispanic, White, and Black children followed with gains of 14, 12, and 10 points respectively.

Table B.3 W score comparisons by racial and Ethnic group in Early Math

Early Math	n	W Scores		Fall-to-spring growth	Was growth statistically significant?*
		Fall	Spring		
White	164	406	418	12	Yes
Asian	113	397 (Sig)	412 (Sig)	15	Yes
Black	301	401 (Sig)	411 (Sig)	10	Yes
Hispanic	357	392 (Sig)	406 (Sig)	14	Yes

NS = not significantly different from the score obtained by White children
 Sig = statistically significantly different from the score obtained by White children
 *Growth was statistically significantly greater than zero, measured at the .05 level

²⁹The fall-to-spring gain was significantly greater than zero, and therefore unlikely to have occurred by chance alone.

Home Language

In the fall, children whose home language was a language other than English³⁰ exhibited significantly lower scores than children in English-speaking homes across all tested academic skills. In the spring this gap remained statistically significant (see table B.4). Both groups made statistically significant learning gains from fall to spring.³¹

Table B.4 W score comparisons by home language in all tested academic skills

Letter Recognition	n	W score		Fall-to-spring growth	Was growth statistically significant?*
		Fall	Spring		
<i>Home Language was English</i>	678	344	356	12	Yes
Home language was a language other than English	261	4-4 (Sig)	346 (Sig)	17	Yes
Pre-Writing					
<i>Home Language was English</i>	693	382	403	21	Yes
Home language was a language other than English	268	373 (Sig)	393 (Sig)	20	Yes
Early Math					
<i>Home Language was English</i>	688	403	414	11	Yes.
Home language was a language other than English	264	386 (Sig)	402 (Sig)	16	Yes

NS = not significantly different from the score obtained by children whose home language was English

Sig = statistically significantly different from the score obtained by children whose home language was English

*Growth was statistically significantly greater than zero, measured at the .05 level

³⁰While the home language is not English, we do not know the extent to which the children themselves are fluent in English.

³¹The fall-to-spring gain was significantly greater than zero, and therefore unlikely to have occurred by chance alone.

HRA Eligibility

When we look at children who live in households that are eligible for HRA assistance (see Table 1.5 below), we see that they are significantly behind their non-HRA-eligible peers in all tested academic skills in the fall. Further, the gap between HRA-eligible and non-HRA-eligible children remained significant in the spring. Again, the fall to spring growth was greater than expected for both groups, across all three content areas, with HRA-eligible children making the same or greater growth than non-eligible children.

Table B.5 W score comparisons by HRA eligibility in all tested academic skills

Letter Recognition	n	Age: Years-Months		Fall-to-spring growth	Was growth greater statistically significant?*
		Fall	Spring		
<i>NOT eligible for HRA</i>	389	346	5-5	13	Yes
HRA-eligible	550	335 (Sig)	5-1 (Sig)	22	Yes
Pre-Writing					
<i>NOT eligible for HRA</i>	397	382	5-4	22	Yes
HRA-eligible	564	378 (Sig)	5-3 (Sig)	20	Yes
Early Math					
<i>NOT eligible for HRA</i>	393	404	416	12	Yes
HRA-eligible	559	394 (Sig)	407 (Sig)	13	Yes

NS = not significantly different from the score obtained by children who lived in households that were not eligible for HRA services

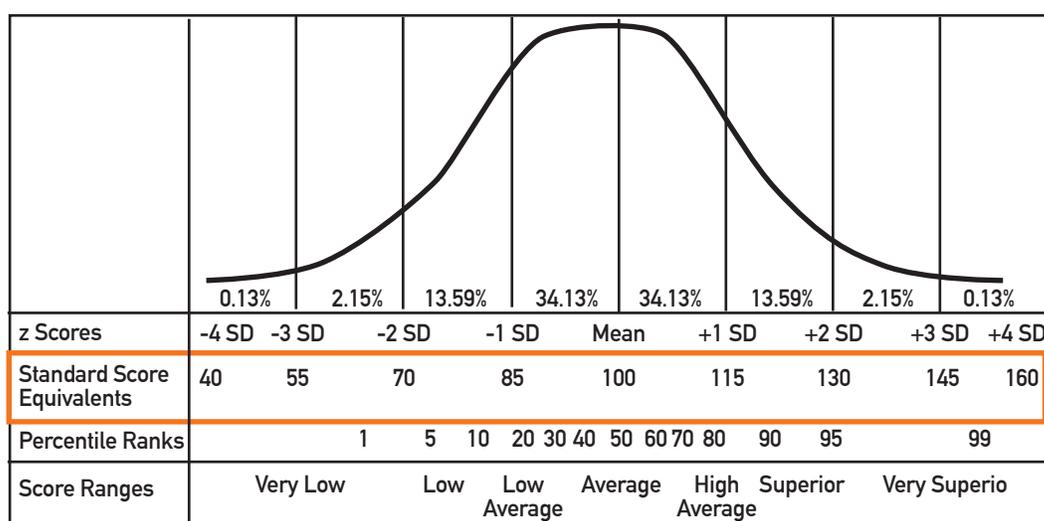
Sig = statistically significantly different from the score obtained by children who lived in households that were not eligible for HRA services

*Growth was statistically significantly greater than zero, measured at the .05 level

Findings by Comparison with Peers (Scale Score)

In this appendix we report the growth of Pre-K for All children in the tested academic skills (Letter Recognition, Pre-Writing, and Early Math) in comparison to the Woodcock-Johnson national norming sample. The scale score describes a child's performance relative to the average performance of the norming sample. The average is a scale score of 100. A scale score of 100 means that the child scored in the middle of the distribution – the average of the norming sample. The “Score Ranges” row in Figure C.1 indicates that

Figure C.1 Distribution of scale scores

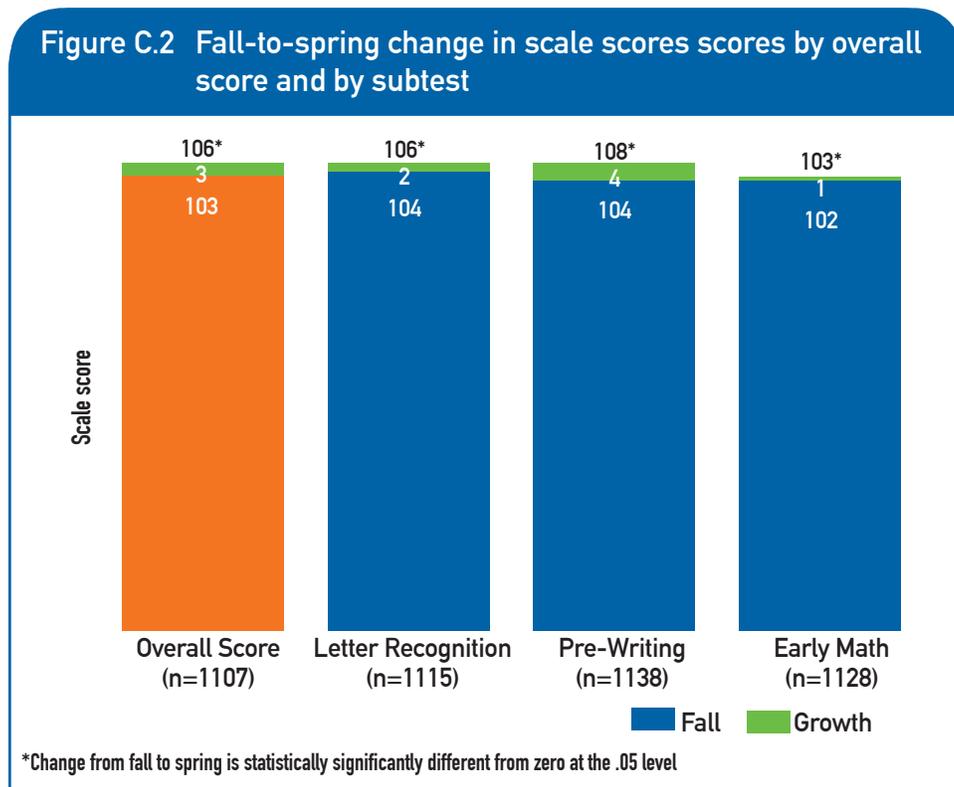


scores within a standard deviation³² on either side of the mean are considered “average.” A standard score of 115 indicates that the child scored one standard deviation above the mean of the norming sample (high average), while a score of 130 is two standard deviations above the mean (superior). Given the different method of calculating these scores and their statistical properties, the significance test results may differ slightly from those presented in Chapter 2.

³²The standard deviation is an average of individual differences from the mean of a group of scores. One standard deviation away from the mean in either direction accounts for approximately 68 percent of the people in the sample. Two standard deviations away from the mean account for roughly 95 percent of the people. And three standard deviations account for about 99 percent of the people.

C.1: Overall findings reported in scale-score metric

Figure C.2 presents the scale scores for each of the tested academic skills in the fall of 2014 (blue bars). These scores are all slightly higher than the national average of 100, ranging from 102 to 104--all of which are still considered to be in the “average” range. Over the course of the school year, children experienced statistically significant growth--between 2 and 4 points (the green bars), with Spring scale scores between 103 and 108--still in the “average” range based on the norming sample.



C.2: Subgroup findings reported in scale-scores

Race and Ethnicity

Asian children tested significantly higher than White children in the fall with **Letter Recognition**, but by the spring, this gap had been eliminated (see Table C.1). Hispanic children had scale scores statistically significantly lower than those of White children in the fall and this gap persisted into the spring. There were no statistically significant differences for Black children when compared to White children. The fall to spring growth was statistically significant for White and Hispanic children, but not for Asian and Black children.

Table C.1 Scale score comparisons by racial and Ethnic group in Letter Recognition

Letter Recognition	n	Scale Scores		Fall-to-spring growth	Was growth statistically significant?*
		Fall	Spring		
<i>White</i>	162	106	107	1	Yes
Asian	114	109 (Sig)	110 (NS)	1	No
Black	296	108 (NS)	108 (NS)	0	No
Hispanic	350	100 (Sig)	102 (Sig)	2	Yes

NS = not significantly different from the score obtained by White children

Sig = statistically significantly different from the score obtained by White children

*Growth was statistically significantly greater than zero, measured at the .05 level

Looking at **Pre-Writing** skills, Black children scored statistically significantly higher than White children in the fall (see Table C.2). There were no statistically significant differences in the fall. By the spring, this difference between Black and White children was no longer statistically significant. Asian children scored statistically significantly higher than White children in the spring. Conversely, Hispanic children scored statistically significantly lower than White children in the spring. The fall to spring growth was statistically significant for all racial/Ethnic subgroups.

Table C.2 Scale score comparisons by racial and Ethnic group in Pre-Writing

Pre-Writing	n	Scale Scores		Fall-to-spring growth	Was growth statistically significant?*
		Fall	Spring		
<i>White</i>	165	104	109	5	Yes
Asian	114	106 (NS)	113 (Sig)	7	Yes
Black	304	106 (Sig)	110 (NS)	4	Yes
Hispanic	360	101 (NS)	104 (Sig)	3	Yes

NS = not significantly different from the score obtained by White children

Sig = statistically significantly different from the score obtained by White children

*Growth was statistically significantly greater than zero, measured at the .05 level

In **Early Math** the fall scale scores for Asian, Black and Hispanic children were all significantly lower than White children (see Table C.3). However, by the spring this gap was eliminated for Asian children, while scale scores for Black and Hispanic children remained significantly lower than those of White children. The fall to spring growth was statistically significant for Hispanic and Asian children, but not for White and Black children.

Table C.3 Scale score comparisons by racial and Ethnic group in Early Math

Early Math	n	Scale Scores		Fall-to-spring growth	Was growth statistically significant?*
		Fall	Spring		
<i>White</i>	164	107	108	1	No
Asian	113	106 (Sig)	106 (Sig)	3	Yes
Black	301	106 (Sig)	104 (Sig)	1	No
Hispanic	357	101 (Sig)	98 (Sig)	3	Yes

NS = not significantly different from the score obtained by White children

Sig = statistically significantly different from the score obtained by White children

*Growth was statistically significantly greater than zero, measured at the .05 level

Home Language

In the fall, children whose home language was a language other than English scored significantly lower than children in English-speaking homes across all tested academic skills. In the Spring this difference remained statistically significant. Children from both groups made statistically significant learning gains from fall to spring.

Table C.4 Scale score comparisons by home language in all tested academic skills

Letter Recognition	n	Scale Scores		Fall-to-spring growth	Was growth statistically significant?*
		Fall	Spring		
<i>Home Language was English</i>	678	107	108	1	Yes
Home language was a language other than English	261	99 (Sig)	103 (Sig)	4	Yes
Pre-Writing					
<i>Home Language was English</i>	693	105	110	5	Yes
Home language was a language other than English	268	100 (Sig)	104 (Sig)	4	Yes
Early Math					
<i>Home Language was English</i>	688	105	106	1	Yes
Home language was a language other than English	264	95 (Sig)	98 (Sig)	3	Yes

NS = not significantly different from the score obtained by children whose home language was English

Sig = statistically significantly different from the score obtained by children whose home language was English

*Growth was statistically significantly greater than zero, measured at the .05 level

HRA Eligibility

When we look at children who live in households that are eligible for HRA assistance (see Table C.5 below), we see that they are significantly behind their non-HRA-eligible peers in all tested academic skills in the fall. Further, the gap between HRA-eligible and non-HRA-eligible children remained significant in the spring. Children from both groups statistically significant learning gains from fall to spring.

Table C.5 Scale score comparisons by HRA eligibility in all tested academic skills

Letter Recognition	n	Scale score		Fall-to-spring growth	Was growth statistically significant?*
		Fall	Spring		
<i>NOT eligible for HRA</i>	389	108	109	1	Yes
HRA-eligible	550	102 (Sig)	104 (Sig)	2	Yes
Pre-Writing					
<i>NOT eligible for HRA</i>	397	106	111	5	Yes
HRA-eligible	564	103 (Sig)	106 (Sig)	3	Yes
Early Math					
<i>NOT eligible for HRA</i>	393	106	108	2	Yes
HRA-eligible	559	99 (Sig)	101 (Sig)	2	Yes

NS = not significantly different from the score obtained by children who lived in households that were not eligible for HRA services

Sig = statistically significantly different from the score obtained by children who lived in households that were not eligible for HRA services

*Growth was statistically significantly greater than zero, measured at the .05 level

³⁵The fall-to-spring gain was significantly greater than zero, and therefore unlikely to have occurred by chance alone.