Pre-K 4 SA Evaluation Report

Year 4

Authors

Lauren Decker-Woodrow, Ph.D.
Emily Diaz, Ph.D.

Don Barfield
Gay Lamey

August 30, 2017

Report Submitted to:
Early Childhood Education Municipal Development Corporation

Table of Contents

Westat

Executive Summary

Pre-K 4 SA served more than 2,000 children during its fourth year of implementation. The Year 4 evaluation of Pre-K 4 SA sought to address research questions regarding attendance, quality, and kindergarten readiness.

Pre-K 4 SA served slightly more boys (50.6%) than girls (49.4%) during Year 4. The majority of Pre-K 4 SA children were Hispanic (76.3%), with the remaining children identified as Black (9.2%), White (8.0%), and other ethnicities (6.5%). Seventy percent of children attended Pre-K 4 SA free; 9.6 percent on scholarship; and 20.4 percent were tuition-paying children. Of those children who attended Pre-K 4 SA free, 89.8 percent did so based on income eligibility.

Average attendance for Pre-K 4 SA children was 92.4 percent, which increased slightly to 93.6 percent when children who withdrew were excluded. Attendance rates have been stable over the first 4 years of implementation.

Westat conducted classroom observations using the Classroom Assessment Scoring System (CLASS) to assess the quality of teacher-child interactions in Pre-K 4 SA classrooms. Overall, teachers were observed displaying high levels of emotional support and relatively high levels of classroom organization. Instructional support was, on average, nearing the middle of the mid-range and improved the most of the three quality domains since the inception of the program. The level of quality across all three domains was above the most recent available average quality scores of the state's and nation's Head Start grantees in Emotional and Instructional Support while similar in Classroom Organization. Additionally, average scores across all three CLASS domains were above a cited "research threshold" for quality. No significant differences were found for any CLASS domain across centers suggesting similar quality experiences for all Pre-K 4 SA children.

New in Year 4, results from a directly assessed measure of executive function showed scores slightly higher than those found during the spring in a recent evaluation of another large city pre-K initiative for a sample of more than 450 Pre-K 4 SA children.

Finally, Pre-K 4 SA children's kindergarten readiness outcomes (measured by the Teaching Strategies' *GOLD*) were compared to a nationally representative sample of children for six outcomes: cognitive, literacy, mathematics, oral language, physical, and social-emotional. Results indicated that although Pre-K 4 SA children started the school year significantly below the normed

Westat®

sample on five of the six outcomes, they surpassed the normed sample in three outcomes (cognitive, literacy, and mathematics) by the end of the year, and eliminated the initial gap in oral language, physical, and social-emotional.

Limitations of the evaluation include the lack of a control group of children for a comparison to a more similar group of children, as well as lack of direct child assessment measures collected both in the fall and spring to measure growth. Classroom observation data were also based on one observation of each classroom during the spring. Therefore, no inferences can be made about changes in classroom quality over time.

Westat®

Introduction

Improving children's kindergarten readiness and narrowing the achievement gap are twin education goals receiving considerable attention throughout the United States (Barnett, 2011). Public investments in preschool education programs have been promoted on the grounds that they can accomplish these twin goals and produce benefits that lead to a high rate of return over time (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010; Reynolds, Temple, White, Ou, & Robertson, 2011; Rolnick & Grunewald, 2003).

As a result of the evidence for high-quality early education and the recent loss of state-funded seats and slow growth of state programs, new initiatives are emerging. This includes programs at the city level to increase school readiness, decrease achievement gaps, and align early care and education programs with K–12 education systems. San Antonio, Texas, is among several cities that have opted for investing in preschool education, in addition to state mandates, much like the Boston pre-K program (National League of Cities, 2012). San Antonio is unique because the city has funded the program through a voter-approved 1/8 cent increase in local sales tax rates starting April 1, 2013. The program, called Pre-K 4 SA, serves many children who are at risk for falling behind their peers and for lacking in kindergarten readiness, with the goal of increasing early childhood quality and school readiness across the city of San Antonio. Pre-K 4 SA completed a fourth year of implementation as of the completion of the 2016–17 school year.

The purpose of the current report is to present Year 4 evaluation findings of the Pre-K 4 SA program. Investigations included (1) information concerning child attendance, classroom quality information, executive function, and (2) outcome analysis results of the Teaching Strategies GOLD assessment, which is the primary outcome of interest at the end of the pre-K year.

Research Questions

The Year 4 (2016–17) evaluation of Pre-K 4 SA sought to address the following four main research questions:

1. What were the reported levels of child attendance during the pre-K year?

   a. Are attendance rates stable over implementation years?

2. What was the overall observed teacher-child interaction quality in Pre-K 4 SA classrooms in Year 4?

    b. Did the Year 4 interaction quality vary by center?

    c. Has improvement been observed in interaction quality since the inception of the program?

3. How do Pre-K 4 SA children preform on a direct assessment of executive function skills?

4. How do Pre-K 4 SA children compare to a nationally representative normed sample of children on GOLD outcomes?

    a. Do differences in gains in GOLD outcomes vary significantly by center or amount/level of teacher-child interaction quality?

Evaluation Sample and Methods

In this section, demographics characteristics for the sample are provided for children served during the 2016-17 school year (Year 4) as well as a brief discussion of methods used.

Sample

Data were provided for 2,060 children in Year 4. Pre-K 4 SA served slightly more boys (50.6%) than girls (49.4%). Of those more than 2,000 children, the majority represented three districts: Northside Independent School District (ISD), San Antonio ISD, and North East ISD.[1] In addition, 20.4 percent of children paid tuition, and 9.6 percent received scholarships (all other children attended at no cost). Table 1 includes the percentage of children per represented school district.

---

[1] These same three districts were also the majority representation in Years 1, 2, and 3 (2013-14 to 2015-16).

Westat®

Table 1.   Children who attended Pre-K 4 SA by district

District name	Number of children	Percentage (%) of total children
Northside	567	27.5
San Antonio	365	17.7
North East	257	12.5
Edgewood	123	6.0
Harlandale	67	3.3
Southwest	51	2.5
Southside	13	0.6
Tuition	420	20.4
Scholarship	197	9.6
Total	2,060	100.0

Note: Children counted by district attend the program at no cost.

The average age of attending children on the first day of school (August 22, 2016) was 4.46 years.[2] The majority of Pre-K 4 SA children were Hispanic (76.3%) with the remaining children reported as Black (9.2%), White (8.0%), and other ethnicities (6.5%). Out of all children enrolled (both tuition and free attending), 85.3 percent were considered economically disadvantaged. Of the children who attended free, this number rose to 89.8 percent. It is important to note, 189 scholarship children may have also met income eligibility criteria (noted as economically advantaged); however, they were not in an attendance zone of a partner school district. Table 2 includes the percentage of children, by eligibility, who attended Pre-K 4 SA at no cost.

Table 2.   Children who attended Pre-K 4 SA free by eligibility criteria

Eligibility criteria	Number of children	Percentage (%) of total eligible children
Economic disadvantage	1,296	89.8
English language learner	209	14.5
Foster care/Conservatorship	13	0.9
Homeless	15	1.0
Military	66	4.6
Eligible total	1,311	100.0

Note: The eligible total is not a sum because children could qualify in more than one category. The percentage of children who attended Pre-K 4 SA free was 70 percent. Children were removed from eligibility criteria counts in this table if they were identified as scholarship or tuition children.

---

[2] This average includes all children in the sample regardless of start date.

Westat

Methods

The first two research questions were addressed through analysis of existing Pre-K 4 SA databases and results from classroom observations. To address the descriptive question pertaining to attendance, data collected by Pre-K 4 SA were submitted to Westat and descriptively analyzed. To address the descriptive and inferential questions, pertaining to classroom quality, data were collected and analyzed from the Classroom Assessment Scoring System (CLASS) (Pianta, La Paro & Hamre, 2008). CLASS is an observational system that assesses classroom practices in preschool by measuring the interactions between children and adults. Observations in the Year 4 evaluation consisted of five 20-minute cycles, followed by 10-minute coding periods.

To address the third research question concerning children's executive function abilities, data were collected and analyzed for a randomly selected group of 471 children, using the Pencil Tap direct assessment (Diamond & Taylor, 1996). Executive function skills, in general, have been identified as an important focus for targeting school readiness and success (Razza & Raymond, 2015). Percentages correct were calculated and used for analyses after summing all correct and incorrect responses, with possible scores ranging from 0 to 16. Once percentages were calculated, these data were descriptively analyzed. Regression analyses were also conducted to explore the relationship between executive function scores and child demographic information. Two-level, multilevel analyses were conducted to examine the relationship between Pencil Tap spring scores and classroom CLASS domain scores.

To address the final research question, inferential tests of differences were conducted between the Pre-K 4 SA children and a nationally representative normed sample of children on the Teaching Strategies GOLD assessment outcomes. GOLD is a teacher-report measure that collects information on children's progress in 36 objectives, three times throughout the year, across six main categories: cognitive, literacy, oral language, mathematics, physical, and social-emotional. In addition, inferential tests were conducted to investigate potential differences in GOLD results by center and whether differences in observed teacher-child interaction quality were related to greater gains in GOLD outcomes for children. See the Appendix for more detailed information on the Year 4 evaluation methodology, including detailed information pertaining to measures used.

Child Attendance in Pre-K 4 SA

Children began attending Pre-K 4 SA at different times. The majority of children (93.2%) began at the start of the academic year (August 22, 2016). The last date children began Pre-K 4 SA was March 22, 2017.[3] Because of these varied dates, some children had the opportunity to attend more days than other children. In fact, the range of possible membership days was 1–177 days, with an average of 166.5 days. Average percentage attendance across all children was 92.4 percent. When considering children who stayed in membership with Pre-K 4 SA through the year (did not withdraw), the attendance percentage increases slightly to 93.6 percent.

Over the course of the year, 193 children withdrew from Pre-K 4 SA. The earliest withdrawal occurred on August 23, 2016, with the last on May 31, 2017. Nearly 41 (40.9%) of the withdrawals occurred before the end of December. No significant differences were found between children who did and did not withdraw in terms of gender (t (1, 2,052.5) = -1.25, p = .21); eligibility to attend Pre-K 4 SA free, on scholarship or tuition (F (2, 2,057) = 1.94, p = .14); or economic disadvantage (t (1, 2,058) = 0.58, p = .56). One difference was found with respect to race/ethnicity (F (3, 2,056) = 7.55, p<.001).[4] Children identified as Black (n = 28; 14.8%) or other ethnicities (n = 23; 17.2%) were more likely to withdraw from Pre-K 4 SA compared to Hispanic children (n = 122; 7.8%). No other statistically significant differences were found with respect to race/ethnicity.

Attendance Rates Over Time

Attendance rates have remained relatively stable over the first 4 years of Pre-K 4 SA implementation. On average, rates have consistently remained between 91–94 percent.

Table 3 displays attendance for all children who attended the program as well as attendance for the subgroup of children who did not withdraw from the program.

---

[3] Although some children did not begin membership in Pre-K 4 SA until late spring, more than 99 percent of all children were in membership by the end of the 2016 calendar year.

[4] Results from Levene's test of homogeneity of variances showed equal variances could not be assumed; therefore, a Welch's ANOVA was conducted.

Westat®

Table 3.   Pre-K 4 SA attendance over time

Enrollment status	Year 1 2013–14	Year 2 2014–15	Year 3 2015–16	Year 4 2016–17
All enrolled children	92.3%	91.3%	92.5%	92.4%
Children who did not withdraw	93.7%	92.5%	93.6%	93.6%

Pre-K 4 SA Teacher-Child Interaction Quality

All 100 Pre-K 4 SA classrooms were observed during Year 4 using the CLASS. Of the 100 classrooms observed, 25 were located at the East center, 25 at the North center, 25 at the South center, and 25 at the West center.

Scores for the Emotional Support domain ranged from 3.65–7.00 (on the 1 to 7 scale) across all five observation cycles, with most scores in the high range of Emotional Support (average score of 6.24), suggesting effective teacher-child interactions were observed most often during the observation period. Slightly lower, with an overall score at the upper end of the middle range, Classroom Organization domain scores ranged from 2.47–7.00, which suggests classrooms showed a mix of effective interactions with periods when interactions were not as effective with regard to Classroom Organization (average score of 5.60). Finally, Instructional Support domain scores ranged from 1.53–6.07, with an average score at the low end of the middle range at 3.55, which suggests only some observed interactions included support from teachers that extended children's thinking or asked questions that encouraged children to analyze and reason throughout the observation period. Visual representations of each of the Year 4 CLASS domain scores are provided in Figure 1.

Figure 1.   Average classroom quality scores for Pre-K 4 SA Year 4
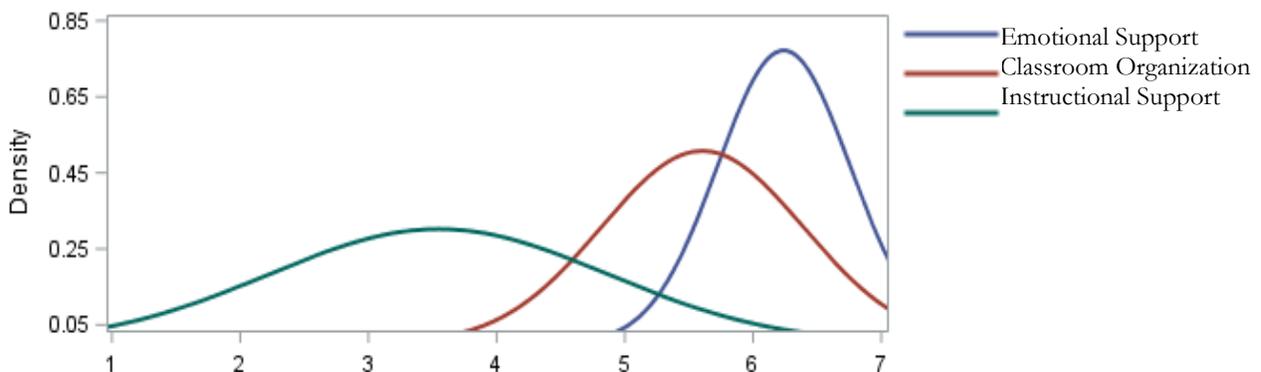
Looking further into the average Emotional Support domain scores, only 24 percent of classrooms (*n* = 24) were observed in the middle range, while 76 percent of classrooms observed provided high levels of Emotional Support in the classroom (*n* = 76). Sixty-five percent of classrooms (*n* = 65) were observed providing low or middle range Classroom Organization quality, while the remaining 35 percent (*n* = 35) were observed providing high levels of Classroom Organization. Finally, 49 percent of the classrooms (*n* = 49) were observed providing low levels of Instructional Support, 51 percent (*n* = 51) were observed providing middle or high levels of Instructional Support. Table 4 provides average scores by each of the 10 dimensions and 3 domains.

Table 4.   Average Year 4 Pre-K 4 SA CLASS scores

CLASS outcome	Average	Total range observed	Standard deviation (SD)
Emotional Support Domain	6.24	3.65—7.00	0.52
Positive Climate	6.27	3.40—7.00	0.65
Negative Climate[a]	6.90	4.80—7.00	0.29
Teacher Sensitivity	5.89	3.00—7.00	0.79
Regard for Student Perspectives	5.89	3.40—7.00	0.67
Classroom Organization Domain	5.60	2.47—7.00	0.79
Behavior Management	5.75	1.20—7.00	0.98
Productivity	5.89	4.20—7.00	0.58
Instructional Learning Formats	5.17	1.40—7.00	1.12
Instructional Support Domain	3.55	1.53—6.07	1.32
Concept Development	3.42	1.20—6.00	1.26
Quality of Feedback	3.52	1.80—6.20	1.31
Language Modeling	3.70	1.00—6.40	1.52

[a] Negative Climate is initially scored with lower values representing no or low negative climate. These scores are then reverse-coded to reflect the same direction (higher values are positive) as the other dimensions.

Past research using the CLASS has often noted the low scores that are commonly seen with respect to the Instructional Support domain (LaParo, Pianta, & Shuhlman, 2004; Locasale-Crouch et al., 2007; Mashburn et al., 2008). To place Pre-K 4 SA CLASS scores in context, the National Institute of Early Education Research (NIEER) (Barnett & Friedman-Krauss, 2016) recently found average scores across Texas and the United States to be lower than those found in the current study. Additionally, Barnett & Friedman-Krauss (2016) compared state and national findings by research thresholds; Pre-K 4 SA scores are visually depicted along with Texas and national Head Start average scores as well as stated research thresholds (see Figure 2). Previous research has found that children in classrooms with Emotional Support scores over 5 also have higher teacher ratings of social competence and lower ratings of behavior problems, while children from classrooms with

Westat®

Instructional Quality ratings of 3.25 or above score higher on measures of reading, mathematics, and expressive language (Burchinal, Vandergrift, Pianta, & Mashburn, 2010).[5]

Figure 2.   Pre-K 4 SA and head start average classroom quality scores
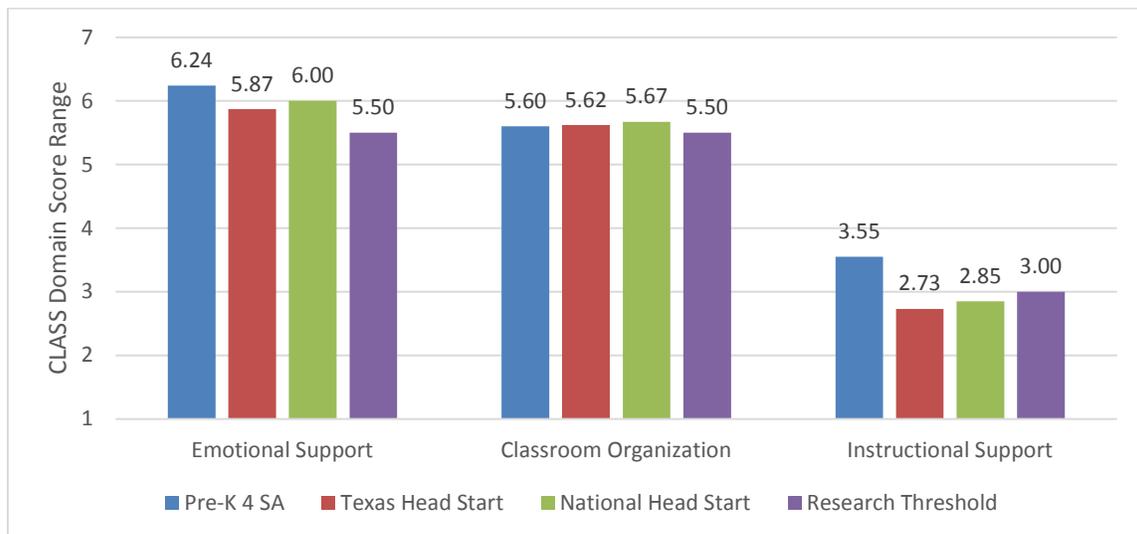
Note: This visual representation is for descriptive purposes only; no statistical tests have been conducted between Pre-K 4 SA and Head Start classrooms for this evaluation.

Source: Barnett, W. S. & Friedman-Krauss, A. (2016). *State(s) of Head Start*. National Institute for Early Education Research. Retrieved from http://nieer.org/wp-content/uploads/2016/12/HS_Digest_States_of_Head_Start.pdf.

---

[5] During the time the study data were collected, the CLASS was broken into two rather than three domains—Emotional Support and Instructional Quality. Direct comparisons of Burchinal et al., 2010 study findings to those presented in the current report should not be made as the dimensions within each domain are not consistent.

Interaction Quality by Center

The three CLASS domains were analyzed to determine if there were significant differences in classroom teacher-child interactions across Pre-K 4 SA centers. No significant differences were found, by center, on any of the three CLASS domain scores, thus, suggesting that statistically equivalent quality was observed, on average, across the four Pre-K 4 SA centers. More specifically, this indicates similar quality classroom experiences offered across Pre-K 4 SA centers for students.

Interaction Quality Over Time

During the first 4 years of implementation, 100 percent of Pre-K 4 SA classrooms were observed. As seen in Figure 3, the overall Emotional Support and Classroom Organization scores have been relatively stable overtime; however, scores did drop slightly between Years 3 and 4. A relative increase has been seen in the Instructional Support domain since the inception of the program (0.73). When compared to Year 1, Year 4 CLASS domain scores for Emotional Support and Classroom Organization were not significantly different; Instructional Support was (Z = -2.00, p = .045).[6] Figure 2 depicts the change in average interaction quality for the program over time.

**Figure 3.   Average CLASS domain scores by program year**

---
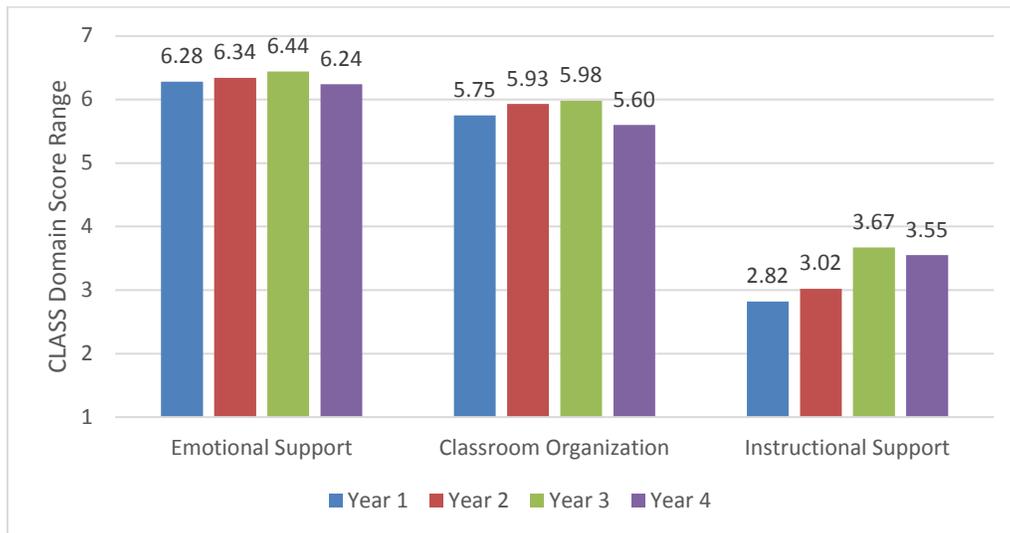
[6] Scores were not normally distributed in a least one year for all three CLASS domains. Therefore, the Wilcoxon Mann-Whiney test was conducted for each.

Westat®

Kindergarten Readiness

Executive Function

A randomly selected group of 500 Pre-K 4 SA children were selected to participate in the Pencil Tap; 471 students (94.2% of the selected group) were assessed using the Pencil Tap during the spring of the pre-K year. On average, Pre-K 4 SA children's percentage of correct responses was nearly 78 percent (77.6%). This average percentage of correct responses is nearly 8 percent (7.7%) higher than a recent city-level pre-K evaluation from New York (Westat, Metis Associates, & Branch Associates, 2016). Figure 4 presents the city-level pre-K program results.

Figure 4. Pencil tap correct responses
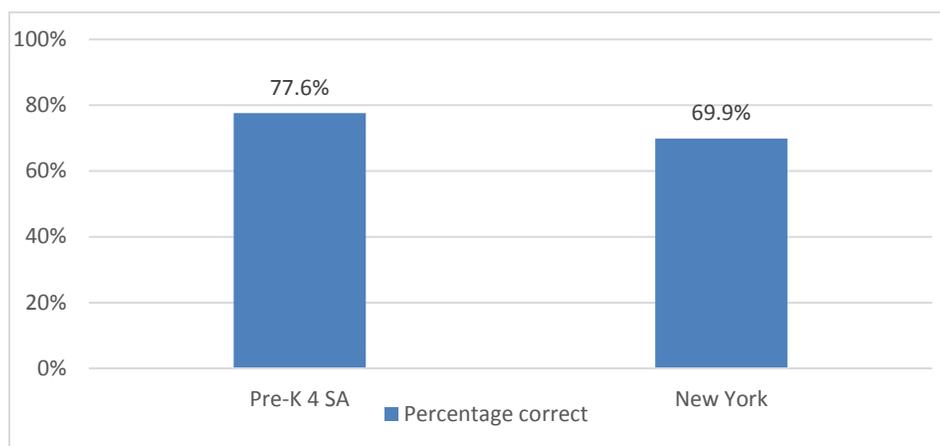
Note: This visual representation is for descriptive purposes only; no statistical tests have been conducted between the Pre-K 4 SA and New York samples for this evaluation.

A stepwise, linear regression analysis was conducted to investigate whether student demographics explained any of the variance in executive function as measured by the Pencil Tap at the end of the pre-K year.[7] Two factors, gender and enrollment status, were found to significantly predict executive function. More specifically, girls were found to display significantly higher executive function skills as compared to boys, ($\beta = .10$, $t(464) = 2.25$, $p = .025$) and tuition children were found to display significantly (although marginally) higher executive function skills as compared to children attending for free ($\beta = .10$, $t(464) = 2.01$, $p = .045$). It is important to note that these differences may have existing upon entry into the pre-K year. No significant differences were found in relation to race/ethnicity, economic disadvantage, English language learner status, or military family status.

---

[7] As CLASS scores were not predictive of executive function scores in the spring, a more parsimonious model that only included child demographics was run; those results for the two demographic factors are reported here.

Westat®

No significant relationship was identified between executive function and CLASS scores.[8] This may be due to existing executive function differences at the beginning of the pre-K. The Year 5 evaluation will address this limitation by collecting executive function data at the beginning and end of the school year so that gains in executive function scores can be explored in relation to the classroom quality provided by Pre-K 4 SA.

Teaching Strategies GOLD

Pre-K 4 SA used the GOLD assessment to collect information on children at three time points throughout the academic year: fall, winter, and spring. Children (89.37%; n = 1,841) were included in analyses if they had outcome data for all three time points in at least one of the following six outcomes: cognitive, language, literacy, mathematics, physical, and social-emotional. No significant differences were found between children included and not included in analyses for gender ($X^2(1) = 0.092, p = .762$) or free lunch status ($X^2(1) = 0.001, p = .983$); however, differences were found for tuition status ($X^2(1) = 4.500, p = .034$) and race ($X^2(6) = 16.085, p = .013$). More specifically, children able to be included in at least one outcome analysis were more likely to be tuition children (Z = 2.12, p = 0.034). For race, children who were included in at least one outcome analysis were more likely to be Hispanic (Z = -3.21, $p = .001$) and less likely to be African American (Z = -2.453, $p = 0.014$) compared to all other racial categories.

As data were not collected on a comparison or control group, comparisons were conducted using the nationally representative normed data for the GOLD assessment (Lambert, Kim, & Burts, 2013). When starting Pre-K 4 SA, children began the fall significantly below the normed sample on five of the six GOLD outcomes; Pre-K 4 SA children began the year similar to the normed sample in mathematics. By spring, the Pre-K 4 SA children scored statistically significantly ($p<.001$) higher than the normed sample on three outcomes (cognitive, literacy, and mathematics). In two of those (literacy and mathematics), Pre-K 4 SA children already significantly exceeding the normed group in by the winter time point. Effect sizes (Hedges' *g*) for the significant results ranged from small (0.31 for cognitive) to medium (0.49 literacy and 0.66 for mathematics). Over the course of the pre-K year, Pre-K 4 SA children gained an additional 31.92 scale score points (27.8% more) in cognitive,

---

[8] A multilevel modeling approach was used as individual child assessments (Pencil Tap) were clustered within classrooms (Raudenbush & Bryk, 2002). A two-level model was used, with children at level 1 and classrooms at level 2.

Westat®

30.32 scale score points (35.4% more) in literacy, and 31.33 scale score points (38.7% more) in mathematics than the normative group of children.

Spring results for the oral language, physical, and social-emotional outcomes indicated the initial gaps between Pre-K 4 SA children and the normed sample for all three outcomes were eliminated by the end of the school year. By spring, no significant differences were found between Pre-K 4 SA children and the normed sample for oral language, physical, or social-emotional. To achieve the elimination of the initial gaps, Pre-K 4 SA children gained an additional 18.46 scale score points (16.5% more) in oral language, 16.18 scale score points (15.2% more) in physical, and 21.70 scale score points (19.4% more) in social-emotional than the normative group of children. See Table 5.

Table 5. Pre-K 4 SA and normed sample comparison results for six GOLD outcomes across time

Outcome	Time point	Pre-K 4 SA mean	Normed mean	Gap (Pre-K – normed)	*t*-test statistic	*df*	Initial *p*-value	Adjusted significance	Group favored[a]	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normed sample)
Cognitive	Fall	562.88	575.72	-12.84	-5.206	1334.49	0.000	Significant	Normed	
	Winter	638.20	636.00	2.20	0.928	1305.31	0.354	Non-Significant	No difference	
	Spring	709.79	690.71	19.08	7.195	1506.10	0.000	Significant	Pre-K	
Literacy	Fall	571.88	576.00	-4.12	-2.093	1419.61	0.037	Significant	Normed	
	Winter	635.99	623.10	12.89	6.632	1535.61	0.000	Significant	Pre-K	
	Spring	687.85	661.65	26.20	11.663	1603.51	0.000	Significant	Pre-K	
Mathematics	Fall	581.60	578.93	2.67	1.415	1391.65	0.157	Non-Significant	No difference	
	Winter	641.15	622.33	18.82	10.428	1499.35	0.000	Significant	Pre-K	
	Spring	693.91	659.91	34.00	16.088	1640.23	0.000	Significant	Pre-K	

Table 5. Pre-K 4 SA and normed sample comparison results for six GOLD outcomes across time (continued)

Outcome	Time point	Pre-K 4 SA mean	Normed mean	Gap (Pre-K – named)	t-test statistic	df	Initial p-value	Adjusted significance	Group favored[a]	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normed sample)
Oral Language	Fall	558.72	574.43	-15.71	-5.824	1334.78	0.000	Significant	Normed	
	Winter	622.50	630.80	-8.30	-3.008	1321.89	0.003	Significant	Normed	
	Spring	688.92	686.17	2.75	0.913	1442.61	0.361	Non-Significant	No difference	
Physical	Fall	543.99	564.82	-20.83	-8.327	1467.30	0.000	Significant	Normed	
	Winter	606.97	618.47	-11.50	-5.243	1325.23	0.000	Significant	Normed	
	Spring	666.62	671.27	-4.65	-1.865	1252.96	0.062	Non-Significant	No difference	
Social-Emotional	Fall	551.91	570.67	-18.76	-7.369	1403.62	0.000	Significant	Normed	
	Winter	622.64	628.05	-5.41	-2.317	1318.54	0.021	Significant	Normed	
	Spring	685.41	682.47	2.94	1.058	1365.11	0.290	Non-Significant	No difference	
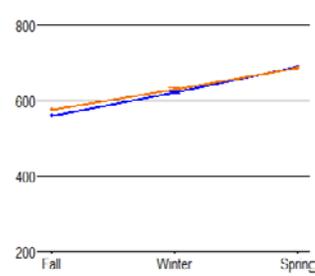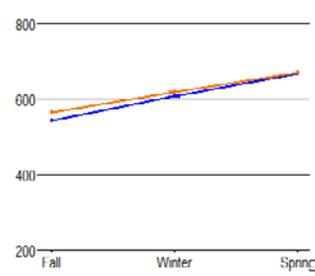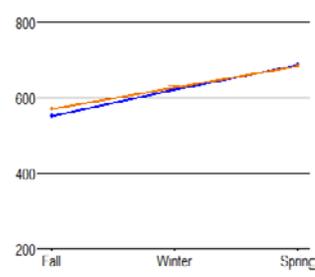
df = degrees of freedom.

[a] If a statically significant difference was found, the group whose score was greater (the "favored" group) is listed in this column. If there was no statistically significant difference, this column states that there was "no difference."

Note: Group mean information is presented in scaled scores. The Adjusted Significance column indicates significance levels (p-values) after adjustment to correct for multiple hypothesis testing using the BenjaminiHochberg technique (1995).

Differences in Readiness Outcomes

Analyses were also conducted within the Pre-K 4 SA sample to explore potential differences related to GOLD outcomes for children. These analyses were conducted between centers and to explore the variance in GOLD outcomes accounted for by the three CLASS domains: Emotional Support, Classroom Organization, and Instructional Support. Results showed there was no significant variation in growth across centers for all six GOLD outcomes; meaning the average growth was the same for children in all four centers. Additionally, across all six GOLD outcomes, results for the three CLASS domains were non-significant. This indicates teacher and child interactions in the classroom were not related to child outcomes on the GOLD assessment, after taking into account demographic characteristics of the children.

Limitations and Recommendations

Three important limitations of the Year 4 evaluation require mention. First, the current evaluation ultimately rests on a primary outcome that is a teacher report rather than a direct child measure conducted by unbiased data collectors. Because a teacher-report measure is the primary outcome of interest, variance in the results related to teacher bias or other teacher factors cannot be excluded. Pre-K 4 SA did add the Pencil Tap in the spring as a direct assessment of children's executive function skills; however, no information was collected in the fall. During Year 5, Pre-K 4 SA has included a pre-, post- assessment of executive function using the Pencil Tap, which will allow for the measurement of growth across the pre-K year in executive function skills. We continue to encourage the consideration of adding an additional directly assessed outcome measure such as oral language or another outcome of primary interest to Pre-K 4 SA.

Second, due to resource constraints, Westat was not able to collect information on a control or comparison group of children with which to compare the Pre-K 4 SA children with respect to kindergarten readiness outcomes. This is important because the normed sample that was used for comparison purposes is most likely very different from the Pre-K 4 SA children. Normed samples are created to be reflective of the demographic proportions similar to those found in the U.S Census. When a comparison or control group can be formed with children who are most like the Pre-K 4 SA children, more confidence can be had with respect to resulting differences on outcomes, meaning there can be more confidence that differences are the result of the program in question and

Westat®

not a result of other factors.[9] This is particularly true when using a control group formed from random assignment into the program. A recommendation related to this limitation is the consideration of additional funding to form a control group of children based from the lottery selection process for admittance to Pre-K 4 SA. Data can be collected from this group and compared between children who attend Pre-K 4 SA and children who do not.

Third, classroom observation data are based on one observation of each classroom during the spring. As such, no inferences can be made about changes in classroom quality over time. Although this was primarily due to resource and time constraints. Consideration should be given to conducting multiple observations across a year (in a random selection of classrooms) to begin to understand potential changes or consistencies in classroom interactional quality.

---

[9] One way to form such a group of children, similar in nature to Pre-K 4 SA children, would be to work with Teaching Strategies to create a matched comparison group from the normed sample of children in the future.

Westat

Barnett, W. S. & Friedman-Krauss, A. (2016). *State(s) of Head Start*. National Institute for Early Education Research. Retrieved from http://nieer.org/wp-content/uploads/2016/12/HS_Digest_States_of_Head_Start.pdf.

Barnett, W. S. (2011). Four reasons the United States should offer every child a preschool education. In E. Zigler, W. Gilliam, & W. S. Barnett (Eds.), *The pre-K debates: Current controversies and issues* (pp. 34–39). Baltimore: Brookes Publishing.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B Methodological, 57*(1), 289–300.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25*(2), 166–176.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science*, *6*, 42–57.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "do as I say, not as I do". *Developmental Psychobiology, 29*, 315–334.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, *94*, 114–128.

La Paro, K. M., Pianta, R. C., & Shuhlman, M. (2004). Classroom Assessment Scoring System (CLASS): Findings from the pre-k year. *Elementary School Journal*, *104*(5), 409–426.

Lambert, R., Kim, D., & Burts, D. (2013). *Technical manual for the Teaching Strategies GOLD assessment system* (CEMETR-2013-05). University of North Carolina Charlotte, Center for Educational Measurement and Evaluation. Retrieved from https://education.uncc.edu/ceme/ceme-technical-reports.

Locasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly, 22*(1), 3–17.

Mann, T. D., Hund, A. M., Hesson-McInnis, M. S., & Roman, Z. J. (2017). Pathways to school readiness: Executive functioning predicts academic and social–emotional aspects of school readiness. *Mind, Brain, and Education, 11*(1), 21–31.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D. M., & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*(3), 732-749.

National League of Cities. (2012). *Educational alignment for young children*. Washington, DC: Institute for Youth, Education and Families. Available from http://www.nlc.org/documents/Find%20City%20Solutions/IYEF/Early%20Childhood/educational-alignment-for-young-chldren-case-studies-april-2012.pdf.

Pianta, R., LaParo, K., & Hamre, B. (2008). *Classroom assessment scoring system*. Baltimore, MD: Brooks Publishing Company.

Razza, R., & Raymond, K. (2015). *Executive functions and school readiness: Identifying multiple pathways for school success*. In S. Robson & S. F. Quinn (Eds.), Routledge International Handbook of Young Children's Thinking and Understanding. New York, NY: Routledge.

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.

Rolnick, A., & Grunewald, R. (2003). Early childhood development: Economic development with a high public return. *The Region, 17*(4), 6–12.

Westat, Metis Associates, Branch Associates (2016). *Pre-K for all: Snapshot of student learning*. Retrieved from New York City Department of Education: http://schools.nyc.gov/NR/rdonlyres/688449CA-8003-46F0-BE1E-E2AB5F649CE2/0/Westat_Metis_Branch PreK_Study_Snapshot_of_Student_Learning_Finalrm.pdf.

Westat®

Appendix

Evaluation Methods

This page is intentionally blank

Appendix
Evaluation Methods

Here we provide information on measures used in the Pre-K 4 SA Year 4 evaluation, as well as detail on the analytic approach to analyses reported.

Measures

Classroom Assessment Scoring System (CLASS)

The CLASS (Pianta, La Paro & Hamre, 2008) is an observational system that assesses classroom practices in preschool by measuring the interactions between children and adults. Observations in the Year 4 evaluation consisted of five, 20-minute cycles, followed by 10-minute coding periods. Scores were assigned during various classroom activities, and then averaged across all cycles for an overall quality score.

Interactions were measured through 10 different dimensions (see Table A-1 for descriptions of each CLASS dimension) that are divided into three larger domains. The Emotional Support domain is measured through the use of four dimensions: positive climate, negative climate, teacher sensitivity, and regard for student perspectives. The CLASS also measures Classroom Organization through three dimensions: productivity, behavior management, and instructional learning formats and Instructional Support through three dimensions: concept development, quality of feedback, and language modeling.

The CLASS uses a 7-point Likert-type scale, for which a score of 1 or 2 indicates low-range quality and a score of 6 or 7 indicates high-range quality. Each dimension and domain is assigned a score during each 20-minute cycle (or, observation period). The number of children and adults in the classroom was also recorded during each 20-minute cycle.

Westat®

Table A-1.   Descriptions of CLASS dimensions

Domain	Dimension	Description
Emotional Support	Positive Climate	Reflects the emotional connection between teachers and children and among children, and the warmth, respect, and enjoyment communicated by verbal and nonverbal interactions.
	Negative Climate	Reflects the overall level of expressed negativity in the classroom. The frequency, quality, and intensity of teacher and peer negativity are key to this dimension.
	Teacher Sensitivity	Encompasses the teacher's awareness of and responsiveness to students' academic and emotional needs.
	Regard for Student Perspectives	Captures the degree to which the teacher's interactions with students and classroom activities emphasize students' interests, motivations, and points of view and encourage student responsibility and autonomy.
Classroom Organization	Behavior Management	Encompasses the teacher's ability to provide clear behavior expectations and use effective methods to prevent and redirect misbehavior.
	Productivity	Considers how well the teacher manages instructional time and routines and provides activities for students so that they have the opportunity to be involved in learning activities.
	Instructional Learning Formats	Focuses on the ways in which teachers maximize students' interest, engagement, and abilities to learn from lessons and activities.
Instructional Support	Concept Development	Measures the teacher's use of instructional discussions and activities to promote students' higher-order thinking skills and cognition and the teacher's focus on understanding rather than on rote instruction.
	Quality of Feedback	Assesses the degree to which the teacher provides feedback that expands learning and understanding and encourages continued participation.
	Language Modeling	Captures the effectiveness and amount of teacher's use of language-stimulation and language-facilitation techniques.

Pencil Tap

The Pencil Tap is a brief direct measurement of children's executive function skills. It is one of seven age-appropriate tasks that make up the Preschool Self-Regulation Assessment (PSRA). Prior to the PSRA, the Pencil Tap emerged from the peg-tapping task (Blair, 2002; Diamond & Taylor, 1996). Executive function skills, in general, have been identified as an important focus for targeting school readiness and success (Razza & Raymond, 2015). Previous research has also found differences between "cool" executive functioning tasks (such as the Pencil Tap task, which involve problem solving and cognitive flexibility), and "hot" executive functioning tasks, which require the regulation of emotions during problem solving (Bassett, Denham, Wyatt, & Warren-Khot, 2012;

Westat®

Brock, Rimm-Kaufman, Nathanson, & Grimm, 2009; Mann, Hund, Hesson-McInnis, & Roman, 2017). Additionally, children's "cool" executive functioning has been found to predict children's later school achievement, including literacy and mathematics outcomes (Brock, et al., 2009; Willoughby, Kupersmidt, Voegler-Lee, & Bryant, 2011).

The Pencil Tap task is a "cool" executive function task that requires children to inhibit a natural tendency to mimic the action of the experimenter while remembering the rule for the correct response, and is thought to assess inhibitory control, attention skills, and working memory. When the test administrator tapped a pencil once, the child was directed to tap their pencil twice. When the administrator tapped twice, the child was directed to tap once.

Teaching Strategies GOLD

The GOLD is a teacher-report measure selected and used by Pre-K 4 SA that collects information on children's progress in 36 objectives across six main categories: Cognitive, Literacy, Oral Language, Mathematics, Physical, and Social-Emotional. (Other categories are available to be tailored to specific programs.) The GOLD assessment is conducted at three time points throughout the year: fall, winter, and spring.

Analytic Approach

Research questions were addressed through analysis of study-collected data as well as existing Pre-K 4 SA databases. To address the first questions, *What were the reported levels of child attendance during the pre-K year?* and *Are attendance rates stable over implementation years?,* data collected by Pre-K 4 SA were submitted to Westat and descriptively analyzed. To address the questions, *What was the overall observed teacher-child interaction quality in Pre-K 4 SA classrooms in Year 4?* and *Did the Year 4 interaction quality vary by Center?,* data were analyzed from the CLASS observations both descriptively and inferentially using analysis of variance (ANOVA). To assess whether *improvement had been observed in interaction quality since the inception of the program?* The Wilcoxon Mann-Whiney test were conducted by domain as scores were not normally distributed. To address the direct assessment question, *How do Pre-K 4 SA children preform on a direct assessment of executive function skills?*, the percentage of correct responses were calculated and descriptively analyzed. Regression analyses were then conducted to explore the relationship between executive function scores and child demographic information.

Westat®

Additionally, two-level, multilevel analyses were conducted to examine the relationship between Pencil Tap spring scores and CLASS domain scores.

The primary outcome research question, *How do Pre-K 4 SA children compare to a nationally representative normed sample of children on GOLD outcomes?* was addressed through independent samples *t*-tests between the Pre-K 4 SA children and a nationally representative normed sample of children on the GOLD assessment outcomes. In addition, inferential tests were conducted to investigate potential differences in GOLD results by center and whether differences in CLASS domain scores were related to higher GOLD outcomes for children. More specifically, three-level, multilevel models were used to investigate center differences, and the potential relationships between CLASS domain scores and GOLD outcomes. A multilevel modeling approach was used as individual child observations were clustered within classrooms and centers (Raudenbush & Bryk, 2002). The three-level model used included children at level 1, classrooms at level 2, and centers at level 3.

The full model for GOLD growth is denoted as:

$$GOLD_{ijk} = \gamma_{000} + r_{0jk} + u_{00k} + e_{ijk}$$

where $GOLD_{ijk}$ is the individual growth for child *i* in classroom *j* in center *k,* γ_{000} is the overall grand mean growth score, r_{0jk} is the deviation of teacher *j* in center *k*, and u_{00k} is the deviation of center *k*, and e_{ijk} is the deviation of child *i* in classroom *j* in center *k*. No covariates were added to the model for two reasons. There was a small sample size of four centers at level 3 and because most of the available child covariates were dichotomous. The combination of these reasons would likely result in the model failing to converge (West, Welch, & Galecki, 2007).

Appendix References

Bassett, H. H., Denham, S., Wyatt, T. M., & Warren-Khot, H. K. (2012). Refining the preschool self-regulation assessment for use in preschool classrooms. *Infant and Child Development, 21*(6). doi: 10.1002/icd.1763.

Blair C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist, 57*, 111–127.

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of 'hot' and 'cool' executive function to children's academic achievement, learning-related behaviours, and engagement in kindergarten. *Early Childhood Research Quarterly, 24*(3), 337-349. doi: 10.1016/j.ecresq.2009.06.001.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "do as I say, not as I do". *Developmental Psychobiology, 29*, 315–334.

Mann, T. D., Hund, A. M., Hesson-McInnis, M. S., & Roman, Z. J. (2017). Pathways to school readiness: Executive functioning predicts academic and social–emotional aspects of school readiness. *Mind, Brain, and Education, 11*(1), 21–31.

Pianta, R., LaParo, K., & Hamre, B. (2008). *Classroom assessment scoring system*. Baltimore, MD: Brooks Publishing Company.

Raudenbush, S. W., & Bryk, A. S. (2002). *hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Razza, R., & Raymond, K. (2015). *Executive functions and school readiness: Identifying multiple pathways for school success*. In S. Robson & S. F. Quinn (Eds.), Routledge International Handbook of Young Children's Thinking and Understanding. New York, NY: Routledge.

West, B. T., Welch, K. B., & Galecki, A. T. (2007). *Linear mixed models: A practical guide using statistical software*. Boca Raton, FL: CRC Press.

Willoughby, M., Kupersmidt, J., Voegler-Lee, M., & Bryant, D. (2011). Contributions of hot and cool self-regulation to preschool disruptive behavior and academic achievement. *Developmental neuropsychology*, *36*(2), 162-180.

Westat®