

Pre-K 4 SA Evaluation Report

Year 5

Authors

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

Don Barfield
Gay Lamey



October 18, 2018

Report Submitted to:
Early Childhood Education Municipal
Development Corporation

Prepared by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

This publication is prepared by Westat under a Professional Services Agreement for the Program Assessment for Pre-K 4 San A Program with the San Antonio Early Childhood Education Municipal Development Corporation, a Texas Municipal Development (City of San Antonio). The content of the publication does not necessarily reflect the views or policies of the San Antonio Early Childhood Education Municipal Development Corporation, a Texas Municipal Development or the City of San Antonio, nor does mention of trade names, commercial products, or organizations imply endorsement by the City of San Antonio.

Table of Contents

<u>Chapter</u>	<u>Page</u>
Executive Summary.....	vi
Introduction	1
Research Questions.....	3
Evaluation Sample and Methods.....	4
Sample: Pre-K Year Analyses (Research Questions 1 – 4).....	4
Sample: Wolf Trap Analyses (Research Question 5).....	5
Sample: First Grade Literacy Analyses (Research Question 6).....	6
Methods	6
Evaluation Results.....	8
Child Attendance in Pre-K 4 SA	8
Attendance Rates Over Time.....	9
Pre-K 4 SA Teacher-Child Interaction Quality.....	9
Interaction Quality by Center.....	12
Interaction Quality Over Time	14
Kindergarten Readiness	14
Executive Function.....	14
Teaching Strategies GOLD	16
Differences in Readiness Outcomes by Center.....	20
Differences in Readiness Outcome Growth by Child and Teacher Characteristics and Classroom Quality.....	22
Relationship between Wolf Trap, Classroom Quality, and GOLD.....	24
First Grade Literacy Outcomes	25
Limitations and Recommendations	27
References.....	R-1
Appendix References	B-4

<u>Appendixes</u>		<u>Page</u>
A	Evaluation Methods	A-1
B	Additional CLASS Results.....	B-1

Tables

1.	Children who attended Pre-K 4 SA by district	4
2.	Children who attended Pre-K 4 SA for free by eligibility criteria	5
3.	Wolf Trap teacher sample by condition and center	5
4.	Pre-K 4 SA attendance over time.....	9
5.	Average Year 5 Pre-K 4 SA CLASS scores	11
6.	Year 5 significant CLASS domain score differences by center.....	13
7.	Year 5 significant CLASS dimension differences by center.....	13
8.	Pre-K 4 SA and normed sample comparison results for six GOLD outcomes across time.....	18
9.	Year 5 significant fall GOLD domain scores by center.....	21
10.	Year 5 significant GOLD growth results based on child and teacher characteristics and classroom quality	23
11.	Elementary school 3rd grade reading performance for Woodcock-Johnson sample	26
A-1.	Descriptions of CLASS dimensions	A-2
B-1.	Average Year 5 CLASS scores by center	A-1

<u>Figures</u>		<u>Page</u>
1.	Average classroom quality scores for Pre-K 4 SA Year 5	10
2.	Pre-K 4 SA and Head Start average classroom quality scores	12
3.	Average CLASS domain scores by program year	14
4.	Pencil Tap correct responses	15
B-1.	Emotional support histograms by center.....	B-2
B-2.	Classroom organization histograms by center.....	B-2
B-3.	Instructional support histograms by center.....	B-3

Executive Summary

Pre-K 4 SA served more than 2,000 children during its fifth year of implementation. The Year 5 evaluation of Pre-K 4 SA sought to address research questions regarding attendance, classroom quality, executive function, and kindergarten readiness during the pre-K year as well as investigate an arts-infused teacher program and a potential longer term relationship between participation in Pre-K 4 SA and first grade literacy outcomes.

Pre-K 4 SA served slightly more boys (51.3%) than girls (48.7%) during Year 5. The majority of Pre-K 4 SA children were Hispanic (76.9%), with the remaining children identified as Black (8.3%), White (7.3%), and other ethnicities (7.6%). More than 75 percent of children attended Pre-K 4 SA for free; 7.9 percent did so on scholarship; and 16.7 percent were tuition-paying children. Of those children who attended Pre-K 4 SA for free, 87.1 percent did so based on income eligibility.

Average attendance for Pre-K 4 SA children was 91 percent, which increased slightly to 92.4 percent when children who withdrew were excluded. Attendance rates have been stable over the first 5 years of implementation.

The Early Childhood Education Municipal Development Corporation contracted with Westat, a large employee-owned, global research firm, to conduct an independent evaluation of the Pre-K 4 SA program. 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 classroom organization. Instructional support was, on average, near the middle of the mid-range. The average 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. Additionally, average scores across all three CLASS domains were above a cited "research threshold" for quality. Significant differences were found for all three CLASS domain between centers and appear to be driven by differences in the dimensions of Teacher Sensitivity, Behavior Management, Instructional Learning Formats, and Quality of Feedback.

Results from a directly assessed measure of executive function showed significant gain scores across the year of executive function skills; these gains were found for all children regardless of several demographic characteristics tested.

Pre-K 4 SA children's kindergarten readiness outcomes (measured by Teaching Strategies' GOLD assessment) 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 gaps in readiness existed for Pre-K 4 SA children at the beginning of the school year on five of the six outcomes, Pre-K 4 SA children reversed those gaps, surpassing the normed sample in two outcomes (cognitive and mathematics) by the end of the year. Initial readiness gaps were also eliminated in literacy, oral language, and physical outcomes and the gap for the social-emotional outcome was closed by more than 50 percent.

Readiness differences were also found within the sample of Pre-K 4 SA children, at the beginning of the pre-K year, as indicated by fall GOLD scores. Specifically, children attending the North and South centers entered the pre-K year ahead of children attending the East and West centers in cognitive, literacy, physical, and social-emotional skills. Children who began the year with higher fall scores also exhibited few gains across the school year on five of the six GOLD outcomes.

Differences in skill gains across the year were also identified for three child characteristics; girls, tuition status, and/or White and Asian children displayed significantly higher gains on several GOLD outcomes compared to boys, children attending for free or on scholarship, and/or Hispanic children. Finally, gains in oral language skills were related to increased levels of Emotional Support; however, gains in social-emotional skills were negatively related to levels of Instructional Support. More information is needed to fully understand these findings.

Results from an investigation into teacher participation in Wolf Trap found no significant relationship between participation, classroom quality, or GOLD outcomes for students. Similarly, despite appearing to attend lower achieving elementary schools, an investigation into first grade literacy for former program children found no significant relationship between Pre-K 4 SA participation and first grade literacy outcomes, as measured by Letter-Word and Passage Comprehension subtests from Woodcock-Johnson Test of Achievement–IV and matching subtests from the Bateria III, Spanish assessment. More specifically, despite appearing to attend lower

achieving elementary schools, former Pre-K 4 SA children were scoring similarly to a comparison group of children attending higher achieving elementary schools.

Looking further into the assessed sample, relationships were identified, within the group of children who previously attended Pre-K 4 SA with respect to pre-K attendance, center, and GOLD outcome gains. Children who previously attended Pre-K 4 SA performed at higher levels of reading comprehension two years later, in first grade, if they had higher attendance rates in Pre-K 4 SA, attended either the North or South centers during the pre-K year, and/or if they had displayed greater gains in any of the six GOLD outcomes during their time with Pre-K 4 SA suggesting some sustained effects of Pre-K 4 SA based on results of the pre-K year.

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 varied direct child assessment measures. Sample sizes for the Wolf Trap and first grade literacy investigations were very small, and no generalizing statements can be made from the results of those analyses. Larger, more representative samples would allow for more confidence in such findings.

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, Savellyev, & Yavitz, 2010; Reynolds, Temple, White, Ou, & Robertson, 2011; Rolnick & Grunewald, 2003).

As a result of the evidence for high-quality early education and the losses 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 fifth year of implementation at the end of the 2017–18 school year.

Over the previous four years, this evaluation has consistently explored who participates in Pre-K 4 SA, attendance in the program, teacher-child instructional quality, and kindergarten readiness outcomes. In addition to these outcome three additional evaluation components are included in the Year 5 evaluation; 1) growth in executive function skills over the pre-k year, 2) relationships between an arts-infused, professional development program, interactions, and readiness scores at the end of the pre-k year, and 3) a longitudinal look at former Pre-K 4 SA children’s literacy and comprehension at the end of 1st grade.

Primary reasons for the additional evaluation components in the current evaluation include interest of Pre-K 4 SA to evaluate program components against stated objectives and goals for program results as well as investigating potential mediators to successful academic outcomes for children. First, executive function skills, in general, have been identified as an important focus for targeting

school readiness and success (Razza & Raymond, 2015). More specifically, “cool” executive functioning tasks (such as the Pencil Tap task, which involves problem solving and cognitive flexibility), have been found to predict children’s later school achievement, including literacy and mathematics outcomes (Brock et al., 2009; Willoughby, Kupersmidt, Voegler-Lee, & Bryant, 2011). Second, Wolf Trap’s arts-integration professional development and classroom residency program brings local artists into the classroom as ‘teaching artists’ to teach alongside the teacher and support the teacher in integrating arts into everyday instruction. Wolf Trap’s program has been previously linked to outcomes for students including mathematics learning (Ludwig & Song, 2015). Lastly, while kindergarten readiness outcomes are explored at the end of the pre-k year, Pre-K 4 SA was interested in understanding potential long term relationships to academic successes of children who previously attended the program as they move through the elementary school grades. To explore potential longer term outcomes, two measures of literacy were explored; word identification and comprehension which are both crucial to literacy development and success (e.g., Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; National Reading Panel, 2000).

The Early Childhood Education Municipal Development Corporation contracted with Westat, a large employee-owned, global research firm, to conduct an independent evaluation of the Pre-K 4 SA program. The purpose of the current report is to present Year 5 evaluation findings of the Pre-K 4 SA program. Investigations included (1) information concerning child attendance, classroom quality information, and executive function; (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; (3) an investigation of an arts-infused program known as Wolf Trap; and (4) a potential longer term relationship between participation in Pre-K 4 SA and first grade literacy outcomes.

Research Questions

The Year 5 (2017–18) evaluation of Pre-K 4 SA sought to address the following six 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 5?
 - A. Did the Year 5 interaction quality vary by center?
 - B. Has improvement been observed in interaction quality from the previous year of implementation (Year 4)?
3. How do Pre-K 4 SA children perform on a direct assessment of executive function skills?
 - A. Are gains in executive function significant over the pre-K year?
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?
 - B. Are gains in GOLD outcomes related to gains in executive function skills during the same year?
5. Do teachers who participated in the Wolf Trap program (either in Year 4 or Year 5) display higher classroom quality than a matched group of teachers who did not participate? Do any differences exist in gains in GOLD outcomes for children in Wolf Trap classrooms?
 - A. Is there any difference between teachers who participated in Year 4 (2016-17) and those that participated in Year 5 (2017-18)?
6. Do 1st grade children who attended Pre-K 4 SA in Year 3 (2015-16) perform differently than waitlist children from the same year on two subtests of literacy (letter-word and passage comprehension)?
 - A. Does elementary school quality (as measured by 3rd grade STAAR reading) relate to 1st grade literacy scores?
 - B. Does former Pre-K 4 SA attendance, center affiliation, or gains during the pre-K year (GOLD) predict 1st grade literacy scores?

Evaluation Sample and Methods

In this section, demographics characteristics for the sample are provided for children served during the 2017-18 school year (Year 5) as well as a brief discussion of methods used.

Sample: Pre-K Year Analyses (Research Questions 1 – 4)

Data were provided for 2,053 children in Year 5. Pre-K 4 SA served slightly more boys (51.3%) than girls (48.7%). 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, 16.7 percent of children paid tuition, and 7.9 percent received scholarships (all other children attended at no cost). Table 1 includes the percentage of children per represented school district.

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

District name	Number of children	Percentage (%) of total children
Northside	587	28.6
San Antonio	365	17.8
North East	291	14.2
Edgewood	126	6.1
East Central	72	3.5
Harlandale	53	2.6
Southwest	39	1.9
Southside	15	0.7
Tuition	342	16.7
Scholarship	163	7.9
Total	2,053	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 28, 2017) was 4.47 years.² The majority of Pre-K 4 SA children were Hispanic (76.9%), with the remaining children reported as Black (8.3%), White (7.3%), and other ethnicities (7.6%). Out of all children enrolled (both tuition, scholarship, and free attending), 73.7 percent were considered economically disadvantaged. Of the children who attended free, this number rose to 87.1 percent. It is important to note, an additional 95 percent (155) of the 163 scholarship children also met income eligibility criteria (noted as

¹ These same three districts were also the majority representation in Years 1-4 (2013-14 to 2016-17).

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

economic disadvantage); 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 for free by eligibility criteria

Eligibility criteria	Number of children	Percentage (%) of total eligible children
Economic disadvantage	1,349	87.1
English language learner	268	17.3
Foster care/Conservatorship	33	2.1
Homeless	4	0.3
Military	98	6.3
Eligible total	1,548	–

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 for free was 75 percent. Children were removed from eligibility criteria counts in this table if they were identified as scholarship or tuition children.

Sample: Wolf Trap Analyses (Research Question 5)

Wolf Trap is an arts-based learning partnership that includes training teachers to incorporate arts-integrated learning strategies into early childhood classrooms. Pre-K 4 SA began a partnership with Wolf Trap during the 2016-17 school year. A total of 19 teachers were identified by Pre-K 4 SA as having participating in the Wolf Trap partnership, either in the 2016-17 or the 2017-18 school year. A matched comparison group was formed, as described in the Methods section below, and included 31 comparison teachers who did not participate in Wolf Trap. The participating teachers and matched comparison teachers represented each Pre-K 4 SA center as displayed in Table 3.

Table 3. Wolf Trap teacher sample by condition and center

Enrollment status	East	North	South	West	Total
Wolf Trap					
Year 4	0	4	0	4	8
Year 5	4	1	5	1	11
Total Wolf Trap teachers	4	5	5	5	19
Comparison					
Year 4	0	6	0	8	14
Year 5	4	2	9	2	17
Total Comparison teachers	4	8	9	10	31

Children in the classrooms of the respective sample teachers were also included in analyses to test the relationship between Wolf Trap participation and kindergarten readiness outcomes. Nearly 900

children (887 children; 812 – 886 children depending on GOLD outcome) from 49³ teachers (46 – 49 teachers depending on GOLD outcome) were included in the analyses.

Sample: First Grade Literacy Analyses (Research Question 6)

Direct assessments of literacy were conducted on 92 former Pre-K 4 SA children and 24 former waitlist children who were in the first grade during the 2017-18 school year. Waitlist children were those who registered for the lottery to attend Pre-K 4 SA, but were not randomly selected to attend during their pre-K year. Across all 116 children, the average age was 7.1 years and 52 percent were male. The majority of children were tested in English; however, 9 percent (n=10) were tested in Spanish.

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 5 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 in the fall (pre-test) and spring (post-test) and analyzed for a randomly selected group of 474 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. T-test and one-way ANOVA analyses were conducted to explore the potential relationship between growth in

³ One comparison teacher was excluded from GOLD analyses because no children had complete data on the GOLD outcomes.

pre- and post-test executive function scores and child demographic information. Two-level, multilevel analyses were conducted to examine the relationship between Pencil Tap scores and classroom CLASS domain scores.

The fourth research question was addressed through inferential tests of differences, which 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, child, and teacher demographics, and whether differences in observed teacher-child interaction quality were related to greater gains in GOLD outcomes for children.

To address the fifth research question, teacher demographic information was used to create a matched comparison group of teachers to those who participated in Wolf Trap. The matched groups were formed using three variables: lead⁴ teacher years of experience in Pre-K 4 SA, lead teacher ethnicity, and whether the classroom taught by the teacher was a bilingual/ESL classroom or not.⁵ Teacher prior year CLASS scores were not included in the matching due to the frequency with which teacher pairs change from year to year. As CLASS observations are reflective of both the lead and co-teachers' interactions in the classrooms (and so many teacher pairs changed between the years in question), it was not appropriate to consider prior year CLASS scores as reflective of the current teacher pairs. Teacher demographic information, CLASS data, and child-level GOLD data were compiled from Year 4 and Year 5 data. T-tests and two-level, multilevel models were conducted to explore the potential relationship between Wolf Trap participation, classroom quality scores, and child growth on GOLD outcomes.

⁴ Westat, through Pre-K 4 SA, was only able to obtain years of experience and ethnicity information for lead teachers. This information was not available for teacher assistants.

⁵ Lead teacher certification information was also obtained; however, due to certification requirements to be a lead teacher with Pre-K 4 SA, there was not significant variation in certifications by lead teachers. Therefore, certification was not included in the matching process.

In addressing the final research question, direct assessments (two subtests from the Woodcock-Johnson Test of Achievement–IV⁶ and matching subtests from the Bateria III⁷, Spanish assessment) were conducted on a subsample of former Pre-K 4 SA children as well as former waitlist children from the same year (Year 3; 2015-16). Pre-K 4 SA staff members reached out by phone to former Pre-K 4 SA children’s families as well as former waitlist children’s families during the early spring of the 2017-18 school year, the spring of first grade. A series of ANCOVA analyses were conducted to examine the potential relationship between attendance in Pre-K 4 SA and literacy scores 2 years later. In addition, further t-tests and ANCOVA analyses were conducted to investigate potential differences in elementary school quality as well as potential relationships between attendance, center membership, quality, and outcome gains during the pre-K year with literacy outcomes in first grade. See Appendix A for more detailed information on the Year 5 evaluation methodology, including detailed information pertaining to measures used.

Evaluation Results

Child Attendance in Pre-K 4 SA

Children began attending Pre-K 4 SA at different times. The majority of children (92.7%) began at the start of the academic year (August 28, 2017). The last date a child began Pre-K 4 SA was May 9, 2018.⁸ 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 164.5 days. Average percentage attendance across all children was 91.0 percent. When considering children who stayed in membership with Pre-K 4 SA through the year (did not withdraw), the average number of membership days rose to 174.8 days and the attendance percentage increased slightly to 92.4 percent.

Over the course of the year, 254 children withdrew from Pre-K 4 SA. The earliest withdrawal occurred on August 29, 2017, with the last on June 7, 2018. Nearly 40 percent (37.0%; n=94) of the

⁶ Woodcock-Johnson Test of Achievement–IV (Schrank, McGrew, Mathers, Wendling, & LaForte, 2014).

⁷ Bateria III (Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005).

⁸ Although some children did not begin membership in Pre-K 4 SA until late spring, more than 98 percent of all children were in membership by the end of the 2017 calendar year.

withdrawals occurred before the end of December. No significant differences were found between children who did and did not withdraw in terms of gender ($t(1, 2,051) = -0.48, p = .63$) or economic disadvantage ($t(1, 2,051) = 0.03, p = .98$). One difference was also found with respect to eligibility to attend Pre-K 4 SA for free, on scholarship, or tuition ($F(2, 2,052) = 4.48, p = .01$).⁹ Children identified as attending on scholarship ($n = 163; 19.6\%$) were more likely to withdraw from Pre-K 4 SA than either children attending for free ($n = 1,548; 12.0\%$) or children attending on tuition ($n = 342; 10.8\%$). Although an initial difference was found with respect to race/ethnicity ($F(3, 2,052) = 4.24, p = .005$)¹⁰, no post-hoc analyses were significant for any of the racial/ethnic groups of children.

Attendance Rates Over Time

Attendance rates have remained relatively stable over the first 5 years of Pre-K 4 SA implementation. On average, rates have consistently remained between 91–94 percent. Table 4 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.

Table 4. 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	Year 5 2017-18
All enrolled children	92.3%	91.3%	92.5%	92.4%	91.0%
Children who did not withdraw	93.7%	92.5%	93.6%	93.6%	92.4%

Pre-K 4 SA Teacher-Child Interaction Quality

All 100 Pre-K 4 SA classrooms were observed during Year 5 using the CLASS. The 100 classrooms were evenly distributed across the four Pre-K 4 SA centers (25 classrooms at each center).

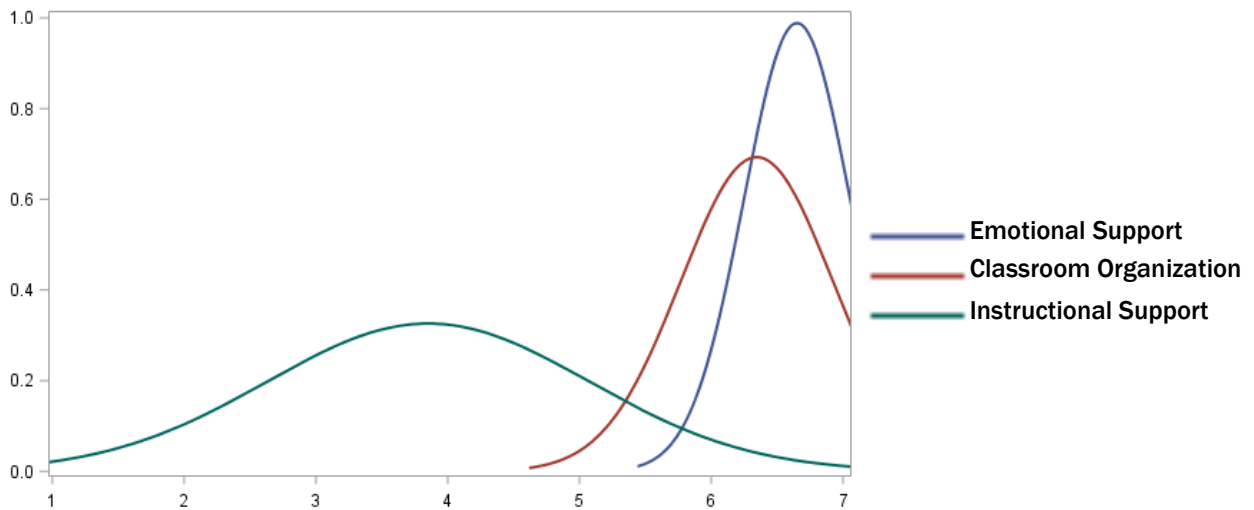
Scores for the Emotional Support domain ranged from 4.75–7.00 (on the 1 to 7 scale) across all five observation cycles, with most scores in the high range of Emotional Support (average score of 6.65), suggesting effective teacher-child interactions were observed most often during the observation period. Similarly, with an overall score in the high range, Classroom Organization domain scores

⁹ Results from Levene’s test of homogeneity of variances showed equal variances could not be assumed; therefore, a Welch’s ANOVA was conducted.

¹⁰ Results from Levene’s test of homogeneity of variances showed equal variances could not be assumed; therefore, a Welch’s ANOVA was conducted.

ranged from 4.80–7.00, which suggests classrooms showed effective interactions with regard to Classroom Organization (average score of 6.35). Finally, Instructional Support domain scores ranged from 1.40–6.60, with an average score near the middle of the middle range at 3.85, 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 5 CLASS domain scores are provided in Figure 1.

Figure 1. Average classroom quality scores for Pre-K 4 SA Year 5



Looking further into the average Emotional Support domain scores, only 6 percent of classrooms ($n = 6$) were observed in the middle range, while 94 percent of classrooms observed provided high levels of Emotional Support in the classroom ($n = 94$). Twenty-two percent of classrooms ($n = 22$) were observed providing middle range Classroom Organization quality, while the remaining 78 percent ($n = 78$) were observed providing high levels of Classroom Organization. Finally, 23 percent of the classrooms ($n = 23$) were observed providing low levels of Instructional Support, while 77 percent ($n = 77$) were observed providing middle or high levels of Instructional Support. Table 5 provides average scores by each of the 10 dimensions and 3 domains.

Table 5. Average Year 5 Pre-K 4 SA CLASS scores

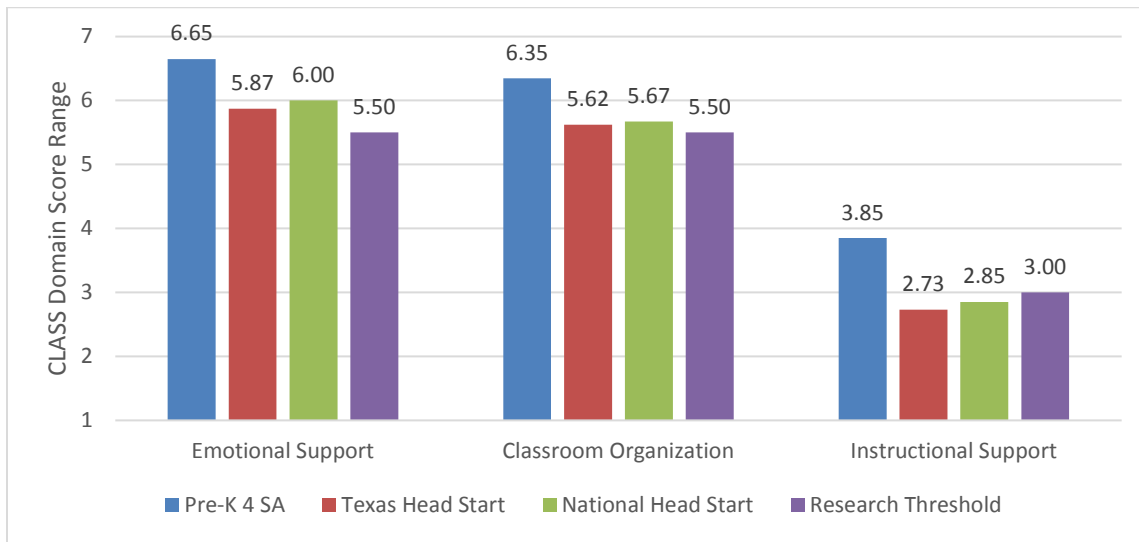
CLASS outcome	Average	Total range observed	Standard deviation (SD)
Emotional Support Domain	6.65	4.75 – 7.00	0.40
Positive Climate	6.64	4.80 – 7.00	0.46
Negative Climate ^a	6.91	6.00 – 7.00	0.22
Teacher Sensitivity	6.48	3.80 – 7.00	0.62
Regard for Student Perspectives	6.58	4.00 – 7.00	0.59
Classroom Organization Domain	6.35	4.80 – 7.00	0.58
Behavior Management	6.37	4.40 – 7.00	0.71
Productivity	6.46	4.00 – 7.00	0.56
Instructional Learning Formats	6.21	4.00 – 7.00	0.69
Instructional Support Domain	3.85	1.40 – 6.60	1.22
Concept Development	3.54	1.20 – 6.40	1.21
Quality of Feedback	3.87	1.40 – 6.60	1.30
Language Modeling	4.14	1.40 – 7.00	1.33

^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 for 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 and 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 Instructional Quality ratings of 3.25 or above score higher on measures of reading, mathematics, and expressive language (Burchinal, Vandergrift, Pianta, & Mashburn, 2010).¹¹

¹¹During 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.

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.

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. One statistically significant comparison result was found for each domain (see Table 6). Two of the significant findings were in favor of the South center compared to the West center, and the other significant finding was in favor of the South center compared to the East center.

Looking further into the center differences, it appears that four dimensions were driving the significant CLASS domain differences (see Table 7). Within Emotional Support, the findings for the dimension of teacher sensitivity were significant in favor of the South center compared to both the East and West centers. Within Classroom Organization, the findings for the dimensions of behavior management and instructional learning formats were both significant in favor of the South center compared to the West center. Within Instructional Support, the findings for the dimension of quality of feedback were significant in favor of the South center compared to both the East and West centers.

No other significant differences were found by center, which indicates similar quality classroom experiences offered across Pre-K 4 SA centers for children for the remaining center comparisons.

Table 6. Year 5 significant CLASS domain score differences by center

CLASS domain	East center group mean	North center group mean	South center group mean	West center group mean	F statistic ^a	df	p-value	Significant center differences	Effect size
Emotional Support	6.49	6.76	6.80	6.55	4.27	96	0.009	East lower than South	0.66
Classroom Organization	6.26	6.45	6.59	6.08	4.03	96	0.010	West lower than South	0.91
Instructional Support	3.48	4.13	4.34	3.46	3.67	96	0.018	West lower than South	0.73

Note. Effect sizes are Hedges' *g*; *df* = degrees of freedom.

^a Although Levene's test for homogeneity of variance was violated for emotional and instructional support, parametric tests were maintained due to the robustness of ANOVA when sample sizes are equal across groups (Field, Miles, & Field, 2012). Therefore, while ANOVA tests were conducted for all three CLASS domains, the adjusted Welch's F statistic is reported for Emotional Support and Instructional Support while the traditional ANOVA F statistic is reported for Classroom Organization.

Table 7. Year 5 significant CLASS dimension differences by center

CLASS domain/ dimension	East center group mean	North center group mean	South center group mean	West center group mean	F statistic ^a	df	p-value	Significant center differences	Effect size
Emotional Support Teacher Sensitivity	6.28	6.58	6.77	6.30	5.28	96	0.003	East and West lower than South	0.75 0.95
Classroom Organization Behavior Management	6.18	6.54	6.66	6.10	4.02	96	0.010	West lower than South	0.82
Instructional Learning Formats	6.19	6.30	6.47	5.86	3.67	96	0.015	West lower than South	0.84
Instructional Support Quality of Feedback	3.44	4.15	4.43	3.45	4.07	96	0.009	East and West lower than South	0.68 0.77

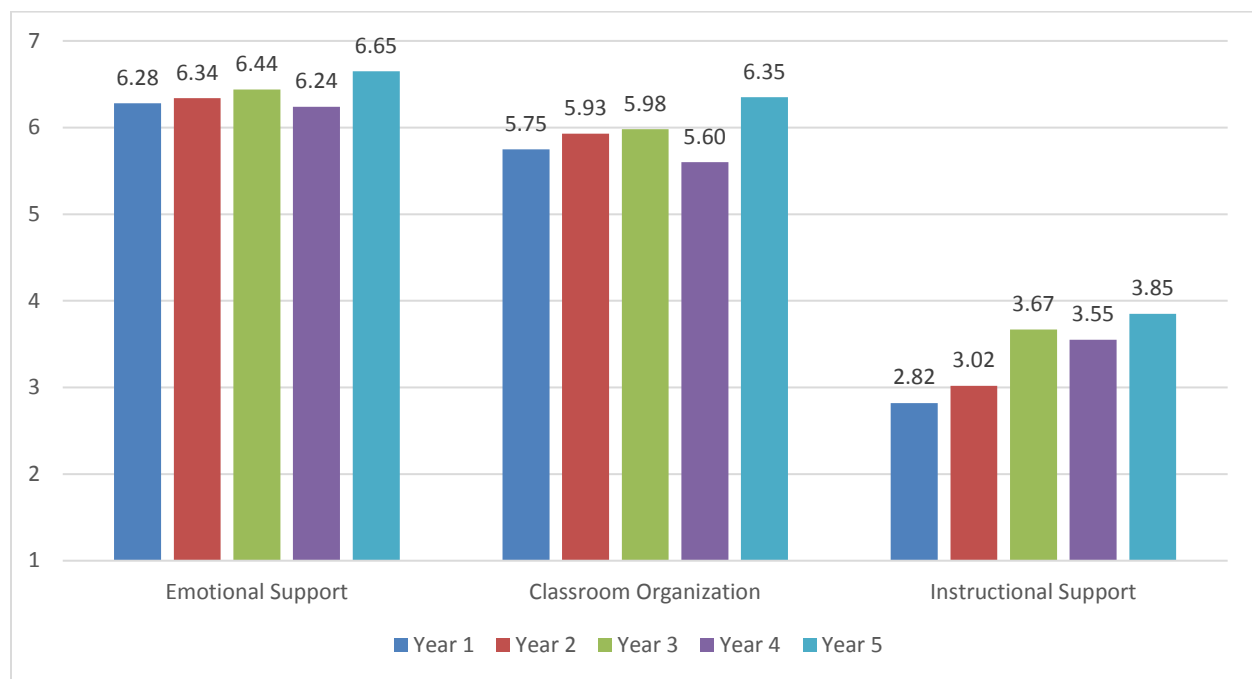
Note. Effect sizes are Hedges' *g*; *df* = degrees of freedom.

^a Although Levene's test for homogeneity of variance was violated for Teacher Sensitivity, parametric tests were maintained due to the robustness of ANOVA when sample sizes are equal across groups (Field, Miles, & Field, 2012). Therefore, while ANOVA tests were conducted for all 10 CLASS dimensions, the adjusted Welch's F statistic is reported for Teacher Sensitivity while the traditional ANOVA F statistic is reported for Behavior Management, Instructional Learning Formats, and Quality of Feedback.

Interaction Quality Over Time

During the first 5 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 over time; however, scores did drop slightly between Years 3 and 4 before rising again in Year 5. When compared to Year 4, Year 5 CLASS domain scores for Emotional Support ($t=6.28$; $p<0.000$) and Classroom Organization ($t=7.61$; $p<0.000$) were significantly higher. Instructional Support was not, although Instructional Support between Years 4 and 5 did appear to increase. Figure 3 depicts the change in average interaction quality for the program over time.

Figure 3. Average CLASS domain scores by program year



Kindergarten Readiness

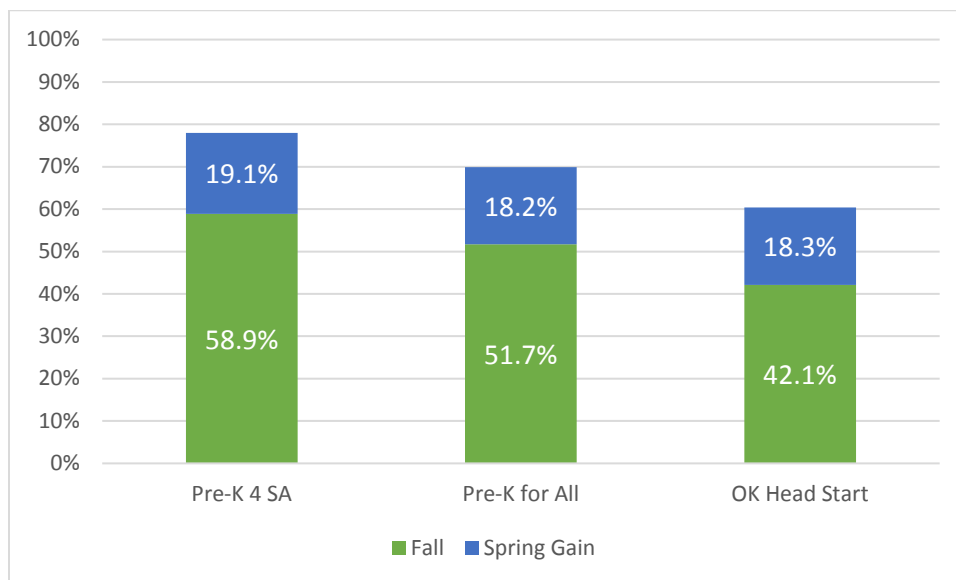
Executive Function

A randomly selected group of 558 Pre-K 4 SA children were selected to participate in the Pencil Tap assessment; 474 children (84.9% of the selected group) were assessed using the Pencil Tap during both the fall and spring of the pre-K year. In terms of demographic characteristics, there were no significant differences between children who were assessed at both time points and those who were

not with respect to child gender ($t(556) = -0.70, p = .49$); eligibility to attend Pre-K 4 SA at no cost or on tuition ($F(2, 87.71) = 2.83, p = .06$); economic disadvantage ($t(556) = 1.27, p = .21$); or race/ethnicity ($F(5, 552) = 1.30, p = .26$). This suggests the sample, and findings, can be considered representative of children who attended Pre-K 4 SA during the 2017-18 school year.

The fall-to-spring gain in the percentage of correct responses to the Pencil Tap measure is provided in Figure 4. The nearly 20 percent (19.1%) increase in children’s average correct performance is statistically significant. This suggests, on average, that children shifted from a limited understanding of, and performance on, the task to greater mastery of these key self-regulatory skills (inhibiting their impulses and remembering to use the rules of the “game”) over the course of the pre-K year.

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 and/or Oklahoma samples for this evaluation.

Sources: Westat, Metis Associates, Branch Associates (2016). Pre-K for all: Snapshot of student learning. Retrieved from Research Gate: https://www.researchgate.net/publication/308712510_Pre-K_for_All_Snapshot_of_Student_Learning/figures and Choi, J. Y., Castle, S., Williamson, A. C., Young, E., Worley, L., Long, M., & Horm, D. M. (2016). Teacher-child interactions and the development of executive function in preschool-age children attending Head Start. *Early Education and Development*, 27(6), 751-769.

T-test and one-way ANOVA analyses were conducted to investigate whether child demographics or CLASS scores explained any of the variance in growth in executive function as measured by the change in Pencil Tap scores over the course of the pre-K year. No significant differences were

found in relation to gender, race/ethnicity, economic disadvantage, status (free, scholarship, or tuition), English language learner status, or CLASS scores.¹²

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 (86.6%; $n = 1,778$) 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 ($X^2(1) = 0.024, p = .878$), free lunch status ($X^2(1) = 0.118, p = .731$), and race ($X^2(6) = 10.556, p = .103$); however, differences were found for tuition status ($X^2(2) = 9.433, p = .009$). More specifically, children able to be included in at least one outcome analysis were more likely to be tuition children ($Z = 2.75, p = 0.006$).

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 started the year significantly higher than the normed sample in mathematics. Furthermore, for all three time points, Pre-K 4 SA children were significantly higher than the norm group and the difference between Pre-K 4 SA children and the normed group increased across all three time points in mathematics. Over the course of our series of evaluations, the Pre-K 4 SA sample has appeared to increase in initial mathematics scores compared to the normative sample. More information is needed to understand what mechanisms might be behind this apparent continuing increase in mathematics readiness prior to the pre-k year.

By spring, the Pre-K 4 SA children scored statistically significantly ($p < .001$) higher than the normed sample on two outcomes (cognitive and mathematics). Effect sizes (Hedges' g) for the significant results ranged from small (0.11 for cognitive) to medium (0.52 mathematics). Over the course of the pre-K year, Pre-K 4 SA children gained an additional 23.84 scale score points (20.7% more) in

¹²A 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.

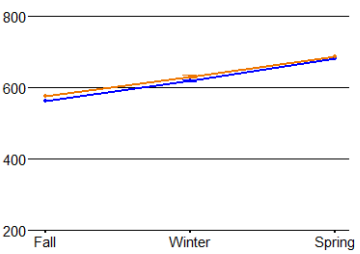
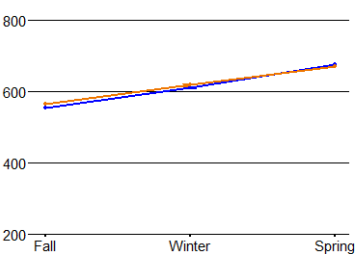
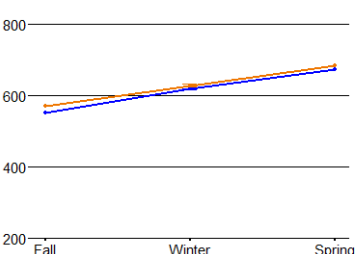
cognitive and 15.97 scale score points (19.7% more) in mathematics than the normative group of children.

Spring results for the literacy, oral language, and physical 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 (the gap still persisted for social-emotional). By spring, no significant differences were found between Pre-K 4 SA children and the normed sample for literacy, oral language, or physical. To achieve the elimination of the initial gaps, Pre-K 4 SA children gained an additional 16.60 scale score points (19.4% more) in literacy, 7.18 scale score points (6.4% more) in oral language, and 14.66 scale score points (13.8% more) in physical than the normative group of children. Although a gap still remained for the social-emotional outcome, it was reduced by 53 percent at spring (a reduction from an initial gap of more than 18 scale score points to approximately 9 scale score points). See Table 8.

Table 8. 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	558.42	575.72	-17.30	-7.069	1301.20	0.000	Significant	Normed	
	Winter	626.35	636.00	-9.65	-4.063	1318.65	0.000	Significant	Normed	
	Spring	697.26	690.71	6.55	2.501	1447.70	0.013	Significant	Pre-K	
Literacy	Fall	556.47	576.00	-19.53	-9.792	1476.50	0.000	Significant	Normed	
	Winter	612.39	623.10	-10.71	-5.600	1468.25	0.000	Significant	Normed	
	Spring	658.72	661.65	-2.93	-1.397	1306.10	0.163	Non-Significant	No difference	
Mathematics	Fall	586.06	578.93	7.13	3.807	1359.71	0.000	Significant	Pre-K	
	Winter	638.18	622.33	15.85	8.967	1406.05	0.000	Significant	Pre-K	
	Spring	683.01	659.91	23.10	11.646	1361.56	0.000	Significant	Pre-K	

Table 8. 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	562.71	574.43	-11.72	-4.506	1181.55	0.000	Significant	Normed	
	Winter	618.69	630.80	-12.11	-4.480	1232.43	0.000	Significant	Normed	
	Spring	681.63	686.17	-4.54	-1.550	1320.62	0.121	Non-Significant	No difference	
Physical	Fall	554.72	564.82	-10.10	-4.256	1234.86	0.000	Significant	Normed	
	Winter	610.74	618.47	-7.73	-3.484	1372.08	0.001	Significant	Normed	
	Spring	675.84	671.27	4.57	1.723	1512.81	0.085	Non-Significant	No difference	
Social-Emotional	Fall	552.13	570.67	-18.54	-7.504	1273.38	0.000	Significant	Normed	
	Winter	617.64	628.05	-10.41	-4.512	1261.36	0.000	Significant	Normed	
	Spring	673.74	682.47	-8.73	-3.211	1267.46	0.001	Significant	Normed	

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 Benjamini-Hochberg technique (1995).

Differences in Readiness Outcomes by Center

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 with regard to center differences and 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.

Each fall time point for the six GOLD domains was also analyzed to determine if there were significant differences across Pre-K 4 SA centers for children's assessed skills at entry into Pre-K 4 SA. There were statistically significant comparisons for four (Social-Emotional, Physical, Cognitive, and Literacy) of the six GOLD domains (see Table 9). All of the significant findings favored children who entered into the North or South centers compared to children who entered into the East or West centers with medium and large effect sizes. These findings suggest that children enrolled in the North and South centers came to Pre-K 4 SA with more skills in these areas as compared to children enrolled in the East and West centers when assessed with the GOLD.

Table 9. Year 5 significant fall GOLD domain scores by center

GOLD outcome	East center group mean	North center group mean	South center group mean	West center group mean	F statistic ^a	df	p-value	Effect Size	Significant center differences
Cognitive	542.04	566.32	577.81	547.06	4.22	(3,96)	0.008	0.84 0.84	East lower than South West lower than South
Literacy	551.54	563.67	569.96	539.93	5.84	(3,94)	0.001	1.02 1.18	West lower than North West lower than South
Physical	535.13	567.87	568.71	546.42	4.49	(3,96)	0.005	0.82 0.81 0.71 0.70	East lower than North East lower than South West lower than North West lower than South
Social-Emotional	537.44	564.06	573.42	533.58	6.73	(3,97)	0.000	0.73 0.94 0.89 1.10	East lower than North East lower than South West lower than North West lower than South

df = degrees of freedom.

Note: There were no significant differences for Oral language and Mathematics. Due to violations of the independence assumption a cluster regression analysis was conducted. Effect sizes between 0.5 and 0.8 are medium and effect sizes greater than 0.8 are large.

Differences in Readiness Outcome Growth by Child and Teacher Characteristics and Classroom Quality

Analyses were also conducted within the Pre-K 4 SA sample to determine if variance in growth in GOLD outcomes was accounted for by child and lead teacher demographics¹³ and the three CLASS domains. There were significant results for all six GOLD outcomes (see Table 10) based on child demographic, and two of the six GOLD outcomes were significant in relation to classroom quality (as measured by the CLASS). There were no significant findings based on lead teacher demographics.

Child Characteristics

There were significant differences in the GOLD outcomes based on child gender, free or tuition status, and race/ethnicity. With respect to gender, girls were assessed as having higher growth than boys across all six outcomes. In the cognitive domain, girls grew a little over seven (7.13) additional scale score points compared to boys. In the literacy, mathematics, oral language, and physical domains, girls grew over three (3.67, 3.43, 3.90, and 3.82, respectively) additional scale score points compared to boys. In the social-emotional domain, girls grew almost 10 (9.85) additional scale score points compared to boys. With respect to free or tuition status, children who paid tuition had higher growth compared to those children who attended for free or based on scholarship in cognitive, literacy, and oral language. In the cognitive and literacy domains, tuition children grew over four (4.90 and 4.40, respectively) additional scale score points compared to free and scholarship children. In the oral language domain, tuition children grew over six (6.33) additional scale score points compared to free and scholarship children.

¹³Child characteristics included gender, free or tuition status (scholarship children were categorized as free), and race/ethnicity. Lead teacher demographics included race/ethnicity and years of Pre-K 4 SA experience. Assistant teacher demographics were not included as this data was not available.

Table 10. Year 5 significant GOLD growth results based on child and teacher characteristics and classroom quality

Child Characteristics	Significant GOLD outcomes	Coefficient	Standard Error	Z	p-value	Group Favored ^b	
Gender	Cognitive	7.134	1.577	4.52	0.000	Girls	
	Literacy	3.671	1.175	3.12	0.002		
	Mathematics	3.433	1.174	2.92	0.003		
	Oral language	3.902	1.620	2.41	0.016		
	Physical	3.824	1.556	2.46	0.014		
Free or tuition status ^a	Social-Emotional	9.851	1.464	6.73	0.000	Tuition	
	Cognitive	-4.903	2.145	-2.29	0.022		
	Literacy	-4.402	1.579	-2.79	0.005		
Race	Hispanic and White/Asian	Oral language	-6.327	2.173	-2.91	0.004	White/Asian
		Cognitive	-6.839	2.865	-2.39	0.017	
		Literacy	-5.850	2.111	-2.77	0.006	
Fall GOLD score	Mathematics	-7.910	2.148	-3.68	0.000	N/A	
	Cognitive	-0.100	0.026	-3.82	0.000		
	Literacy	-0.310	0.017	-18.38	0.000		
	Mathematics	-0.245	0.019	-13.03	0.000		
	Physical	-0.206	0.030	-6.96	0.000		
Classroom Quality	Instructional Support	Social-Emotional	-0.237	0.023	-10.12	0.000	N/A
		Emotional Support	50.399	19.583	2.57	0.010	
	Instructional Support	-9.514	4.385	-2.17	0.030		

Note: There were no differences based on lead teacher demographics when controlling for child characteristics and classroom quality.

^a Scholarship students were included in the free category

^b If a statically significant difference was found, the group whose score was greater (the “favored” group) is listed in this column.

The fall GOLD score and classroom quality analyses are marked as N/A since no groups are compared.

The fall GOLD score was significantly related to growth in five of the GOLD domains. As children entered Pre-K 4 SA with higher scores in fall, the potential for growth over time was limited, resulting in negative findings. Results ranged from -0.1 in cognitive to -0.3 in literacy. For the cognitive domain, this finding implies that for every one scale score point increase in the fall, the growth from fall to spring was decreased by 0.1 scale score points. Similarly, for the literacy domain, this finding implies that for every one scale score point increase in the fall, the growth from fall to spring was decreased by 0.3 scale score points.

Classroom Quality

There were two significant findings based on classroom quality. In the oral language domain, there was a positive finding with respect to emotional support. For every one point increase in emotional support, children grew a little over 50 (50.40) scale score points in oral language. In the social-emotional domain, there was a negative finding with respect to instructional support. For every one-point increase in instructional support, children's social emotional growth decreased by almost 10 (9.51) points. This is a counterintuitive finding but may be related to focus on certain aspects of instruction that could be taking away from other important areas. One explanation may be that High Scope was implemented in this evaluation year, with an increased focus on instruction, however, more information is needed to understand this finding.

Relationship between Wolf Trap, Classroom Quality, and GOLD

T-test analyses were conducted within Pre-K 4 SA between the Wolf Trap and matched comparison teachers in relation to classroom quality, as measured by the CLASS, during the spring of the respective participation year (either Year 4 or Year 5, depending on the year of participation). No significant differences were found for classroom quality in relation to Wolf Trap participation.

Two-level multilevel analyses were also conducted to investigate whether a significant relationship exists between Wolf Trap participation and student GOLD outcomes. No significant differences were found for gains in any of the six GOLD outcomes for children in relation to teacher Wolf Trap participation.

First Grade Literacy Outcomes

Pre-K 4 SA and Waitlist Comparison

ANCOVA analyses were conducted between former Pre-K 4 SA children and children who had been on the waitlist for the same pre-K year (Year 3; 2015-16). After controlling for child age and assessment language, no significant differences were found for either the Letter-Word or Passage Comprehension subtests from Woodcock-Johnson Test of Achievement–IV and matching subtests from the Bateria III, Spanish assessment. Over the entire sample ($n=114$),¹⁴ the average Letter-Word subtest score was at an equivalent of achievement expected from an 8-year-old child or a child who was more than halfway through the second grade, suggesting that, on average, both former Pre-K 4 SA children and children who had been on the waitlist for the same pre-K year were performing slightly above expectations for their age and grade level on word identification.

With regard to Passage comprehension, the average subtest score was at the equivalent of achievement expected from a child who was not yet 7 years old or a child who was still in the beginning of first grade, suggesting that, on average, both former Pre-K 4 SA children and children who had been on the waitlist for the same pre-K year were performing slightly below expectations for their age and grade level on reading comprehension.

A t-test was also conducted to investigate whether the small sample of former Pre-K 4 SA and waitlist children matriculated into elementary campuses with similar campus-level literacy achievement (as measured by third grade STAAR). Although no results were statistically significant at the 0.05 level, two results were marginally significant from the children's kindergarten year (2016-17); 1) percentage of 3rd grade students within the same campuses who did not meet grade-level performance ($t=-1.750$; $p<0.10$) and 2) average third grade reading scale scores ($t=1.684$; $p<0.10$) (see Table 11). More specifically, for the sample assessed using the two Woodcock-Johnson subtests of literacy, during their kindergarten year, former Pre-K 4 SA children came from campuses with an average of 5 percent more children failing to meet grade level reading performance in third grade compared to campuses attended by former waitlist children. Additionally, former waitlist children attended elementary schools with average third grade reading scale scores nearly 20 points higher

¹⁴Although 116 children were assessed, two assessments in each the Letter-Word and Passage Comprehension subtests were invalid and were, therefore, excluded from analyses.

than in schools attended by former Pre-K 4 SA children. No significant differences were identified for the first grade year, although trends followed the same direction.

Table 11. Elementary school 3rd grade reading performance for Woodcock-Johnson sample

Group membership (2015-16)	Kindergarten year (2016-17)		First grade year (2017-18)	
	Did not meet grade-level	Average scale score	Did not meet grade-level	Average scale score
Former waitlist children	26.45%	1439.30	23.58%	1437.21
Former Pre-K 4 SA children	31.48%	1420.91	27.72%	1426.11

Note: Families of children were asked to provide the current elementary school attended in first grade and all STAAR information was collected for those campuses. It is possible that some children attended a different campus for kindergarten and then moved for the first grade year.

Within Pre-K 4 SA Comparison

Further ANCOVA analyses were conducted using data from the former Pre-K 4 SA children included in the first grade analyses. Three findings are of note within children who previously attended Pre-K 4 SA: significant relationships were identified between Pre-K 4 SA attendance, center attended, and GOLD scores from the pre-K year in relation to first grade passage comprehension. More specifically, as pre-K attendance increased, children’s scores on passage comprehension significantly increased ($t=2.11$; $p=0.038$), meaning that for every 2 percent increase in Pre-K 4 SA attendance, children’s passage comprehension scores increased by nearly one month of achievement (0.9 months). Additionally, children who attended either the North or South centers demonstrated significantly higher passage comprehension scores two years later in first grade compared to children who attended either the East or West centers ($t=7.18$; $p<0.000$). This finding somewhat mirrors the fact that, classroom quality and GOLD gain differences were previously found between centers, for this cohort of students (who attended Pre-K 4 SA during the 2015-16 school year). More specifically, the Year 3 evaluation previously found significant differences across all three classroom quality domains between the East and South centers, as well as for Instructional Support between the East and North center when this cohort of children attended Pre-K 4 SA. Additionally, fewer gains were observed in five of the six GOLD outcomes for this cohort of children with the East and West centers reporting smaller average gains across outcomes compared to the North and South centers.

Finally, significant relationships were found between growth in all six GOLD outcomes during the pre-K year and passage comprehension scores two years later. More specifically, as total gain between fall and spring pre-K year scores in each respective GOLD outcome increased, significant increases were found in passage comprehension scores at first grade (significant results ranging from $t=2.42$; $p=0.018$ – $t=3.67$; $p<0.000$). For example, for every 25 additional points gained during the pre-K year in the social-emotional, physical, or cognitive domains, passage comprehension scores increased by one month of achievement. (The same is true for every 20 additional points gained in literacy and for every 16.5 points in oral language or mathematics.)

Limitations and Recommendations

Four important limitations of the Year 5 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 for the kindergarten readiness outcomes at the end of the pre-K year. 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. 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 whom to compare the Pre-K 4 SA children with respect to kindergarten readiness outcomes at the end of the pre-K year. 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 not a result of other factors.¹⁵

¹⁵One 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.

Third, the sample on which analyses were conducted concerning participation in the Wolf Trap program was quite small. In addition, the classroom quality observations likely occurred while teachers were still participating, not giving time for a full completion of the program prior to data collection (as program participation occurred in the spring). The spring time point of participation may also not allow for enough time for participation to have significant impact on child growth. Together, these factors may have made it less likely to identify potential relationships between participation and outcomes for teachers and children. Pre-K 4 SA may want to consider conducting CLASS observations at two time points for teachers new to the Wolf Trap program; once in the fall prior to participation and once in the spring, with participation occurring shortly after the fall time point to allow for some time to pass before assessing any potential changes.

Finally, although some waitlist and former program children did participate in the data collection efforts related to first grade literacy; such a small sample raises concern as to the validity of the results. Only 5.2 percent (n=92) of the former program children and only 1.9 percent (n=24) of the former waitlist children attended, not only leading to an underpowered sample but also, likely, a non-representative sample. There are likely specific reasons or differences that led these particular families to participate, such as parents with a higher interest in their child's academic performance might participate in these sorts of activities at a higher rate. To alleviate some of these issues in the future with longitudinal direct assessments, Pre-K 4 SA should attempt to work with partner school districts to conduct short, developmentally appropriate direct assessments within schools whereby larger numbers of children are quickly assessable and parents/guardians do not have to find travel time out of their schedules to participate, which would likely lead to much greater response rates/sample sizes from which to draw meaningful conclusions.

References

- 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.
- 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.
- 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.
- Choi, J. Y., Castle, S., Williamson, A. C., Young, E., Worley, L., Long, M., & Horm, D. M. (2016). Teacher–child interactions and the development of executive function in preschool-age children attending Head Start. *Early Education and Development*, 27(6), 751-769.
- 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.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Thousand Oaks, CA: Sage.
- Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the High/Scope 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.
- Ludwig, M., & Song, D. M. (2015). Evaluation of Professional Development in the Use of Arts-Integrated Activities With Mathematics Content: Findings From the Evaluation of the Wolf Trap Arts in Education Model Development and Dissemination Grant. Final Grant Report. Available from <https://www.air.org/system/files/downloads/report/Wolf-Trap-Report-Arts-Integration-Jan15-Feb16-update.pdf>.
- 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.
- Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *The Batería III Woodcock- Muñoz: Pruebas de aprovechamiento*. Itasca, IL: Riverside.
- 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-children-case-studies-april-2012.pdf>.
- National Reading Panel (US). (2000). *National reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development, National Institutes of Health.
- Pianta, R., LaParo, K., & Hamre, B. (2008). *Classroom assessment scoring system*. Baltimore, MD: Brookes Publishing.
- 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.
- 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.
- Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, 96(2), 265.

Schrank, F. A., McGrew, K. S., Mather, N., Wendling, B. J., & LaForte, E. M. (2014). *Woodcock-Johnson IV Tests of Achievement: Form A*. Riverside Publishing Company.

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.

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.

Appendix A
Evaluation Methods

This page is intentionally blank

Appendix A

Evaluation Methods

Here we provide information on measures used in the Pre-K 4 SA Year 5 evaluation, as well as details on the analytic approach to analyses reported.

Measures

Classroom Assessment Scoring System (CLASS)

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; as well as 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.

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 involves 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; 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.

The Woodcock-Johnson Achievement Test

The Woodcock-Johnson Test of Achievement–IV (Schrank, McGrew, Mathers, Wendling, & LaForte, 2014) is an individually administered norm-referenced test to assess reading, oral language, mathematics, written language, and academic knowledge. We used two subtests in the current evaluation: the Letter-Word Identification subtest and the Passage Comprehension subtest. For children requiring assessment in Spanish, matching subtests from the Bateria III were used (Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005). In both versions, the Letter-Word subtest is a test of basic literacy skills involving symbolic learning and identifying isolated letters and words. The child identifies letters that are in large type and reads the words correctly. In the passage comprehension subtest, reading comprehension is assessed using a cloze procedure with children reading short passages and identifying missing key words that make sense in the context of that passage. In both subtests, items are set in difficulty order, with the easiest first and the most difficult last. Testing stops when the student scores zero on six successive items.

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 two 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 5?* and *Did the Year 5 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 from the previous year of implementation (Year 4)*, t-tests were conducted. To address the direct assessment question, *How do Pre-K 4 SA children perform on a direct assessment of executive function skills?*, the percentages of correct responses were calculated and descriptively analyzed. To explore the relationship between executive function scores and child demographic information, t-tests and ANOVA analyses were conducted. Additionally, two-level, multilevel analyses were conducted to examine the relationship between Pencil Tap spring scores and CLASS domain scores.

The primary pre-K year 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. As children were nested in classrooms, a cluster regression was conducted to investigate whether there were significant differences in the fall GOLD scores and growth in the GOLD scores by center. For GOLD outcomes that showed significant findings, follow-up pairwise center comparisons were conducted and the Benjamini-Hochberg technique (1995) was used to adjust for multiple hypothesis testing. To determine if there was variation in GOLD growth, a multilevel modeling approach was used as individual child assessments (GOLD) were clustered within classrooms (Raudenbush & Bryk, 2002). A two-level model was used, with children at level 1 and classrooms at level 2. Child gender, free or tuition status, race/ethnicity, fall GOLD score, the three CLASS domains, lead teacher years of experience in Pre-K 4 SA, and lead teacher race/ethnicity were included in the model to see which characteristics would account for variance in GOLD growth.

The full model for GOLD growth is denoted as:

$$\begin{aligned}
 GOLD_{ij} = & \gamma_{00} + \gamma_{10} * Child\ Gender_i + \gamma_{20} * Child\ FreeTuition_i + \\
 & \gamma_{30} Child\ RaceEthnicity\ (Hispanic\ versus\ White)_i + \\
 & \gamma_{40} Child\ RaceEthnicity\ (Hispanic\ versus\ African\ American)_i + \\
 & \gamma_{50} Child\ RaceEthnicity\ (White\ versus\ African\ American)_i + \gamma_{60} Child\ Fall\ GOLD_i + \\
 & \gamma_{01} Emotional\ Support_j + \gamma_{02} Classroom\ Management_j + \gamma_{03} Instructional\ Support_j + \\
 & \gamma_{04} Lead\ Teacher\ Years\ Experience_j + \\
 & \gamma_{05} Lead\ Teacher\ RaceEthnicity\ (Hispanic\ versus\ White)_j + \\
 & \gamma_{06} Lead\ Teacher\ RaceEthnicity\ (Hispanic\ versus\ African\ American)_j + \\
 & \gamma_{07} Lead\ Teacher\ RaceEthnicity\ (White\ versus\ African\ American)_j + u_{0j} + e_{ij}
 \end{aligned}$$

Where:

$GOLD_{ij}$ is the individual growth for child i in classroom j ,

γ_{00} is the overall grand mean growth score,

γ_{10} is the child gender effect (girls were coded as 1 and boys as 0),

γ_{20} is the child free or tuition status effect (Free and scholarship were coded as 1 and tuition as 0),

γ_{30} is the child Hispanic versus White race/ethnicity effect,

γ_{40} is the child Hispanic versus African American race/ethnicity effect,

γ_{50} is the child White versus African American race/ethnicity effect,

γ_{60} is the child Fall GOLD score effect,

γ_{01} is the classroom emotional support effect,

γ_{02} is the classroom management effect,

γ_{03} is the classroom instructional support effect,

γ_{04} is the lead teacher's years of Pre-K 4 SA experience effect,

γ_{05} is the lead teacher Hispanic versus White race/ethnicity effect,

γ_{06} is the lead teacher Hispanic versus African American race/ethnicity effect,

γ_{07} is the lead teacher White versus African American race/ethnicity effect,

u_{0j} is the deviation of teacher j , and

e_{ij} is the deviation of child i in classroom j .

To address the fifth question, *Do teachers who participated in the Wolf Trap program (either in Year 4 or Year 5) display higher classroom quality than a matched group of teachers who did not participate? Do any differences exist in gains in GOLD outcomes for children in Wolf Trap classrooms?*, a multilevel modeling approach was used as individual child assessments (GOLD) were clustered within classrooms (Raudenbush & Bryk, 2002). A two-level model was conducted, with children at level 1 and classrooms/teachers at level 2. Child gender, free or tuition status, race/ethnicity, fall GOLD score, the three CLASS domains, lead

teacher years of experience in Pre-K 4 SA, a flag to indicate which year of Wolf Trap participation (Year 4 or Year 5), a flag to indicate participating in Wolf Trap (or comparison group), and the interaction between year and Wolf Trap participation were included in the model.

The full model for GOLD growth is denoted as:

$$\begin{aligned}
 GOLD_{ij} = & \gamma_{00} + \gamma_{10} * Child\ Gender_i + \gamma_{20} * Child\ FreeTuition_i + \\
 & \gamma_{30} Child\ RaceEthnicity\ (Hispanic\ versus\ White)_i + \\
 & \gamma_{40} Child\ RaceEthnicity\ (Hispanic\ versus\ African\ American)_i + \\
 & \gamma_{50} Child\ RaceEthnicity\ (White\ versus\ African\ American)_i + \gamma_{60} Child\ Fall\ GOLD_i + \\
 & \gamma_{01} Emotional\ Support_j + \gamma_{02} Classroom\ Management_j + \gamma_{03} Instructional\ Support_j + \\
 & \gamma_{04} Lead\ Teacher\ Years\ Experience_j + \gamma_{05} Year_j + \gamma_{06} Wolf\ Trap\ participation_j + \\
 & \gamma_{07} Year\ and\ Wolf\ Trap\ Interaction)_j + u_{0j} + e_{ij}
 \end{aligned}$$

Where:

$GOLD_{ij}$ is the individual growth for child i in classroom j ,

γ_{00} is the overall grand mean growth score,

γ_{10} is the child gender effect (girls were coded as 1 and boys as 0),

γ_{20} is the child free or tuition status effect (Free and scholarship were coded as 1 and tuition as 0),

γ_{30} is the child Hispanic versus White race/ethnicity effect,

γ_{40} is the child Hispanic versus African American race/ethnicity effect,

γ_{50} is the child White versus African American race/ethnicity effect,

γ_{60} is the child Fall GOLD score effect,

γ_{01} is the classroom emotional support effect,

γ_{02} is the classroom management effect,

γ_{03} is the classroom instructional support effect,

γ_{04} is the lead teacher's years of Pre-K 4 SA experience effect,

γ_{05} is the year effect (year 4 versus year 5),

γ_{06} is the Wolf Trap effect (Wolf Trap teachers were coded as 1 and comparison as 0),

γ_{07} is the year and Wolf Trap interaction effect,

u_{0j} is the deviation of teacher j , and

e_{ij} is the deviation of child i in classroom j .

To address the final research question, *Do 1st grade children who attended Pre-K 4 SA in Year 3 (2015-16) perform differently than waitlist children from the same year on two subtests of literacy; letter-word and passage comprehension?*, a series of ANCOVA analyses were conducted. Primary covariates included in the models were child age and testing language (English or Spanish). For all necessary post-hoc analyses, the Benjamini-Hochberg technique (1995) was employed to adjust for multiple hypothesis testing.

Appendix B

Additional CLASS Results

Appendix B

Additional CLASS Results

Table B-1. Average Year 5 CLASS scores by center

CLASS outcome	East		North		South		West	
	M (SD)	Total range observed	M (SD)	Total range observed	M (SD)	Total range observed	M (SD)	Total range observed
Emotional Support Domain	6.49 (0.59)	(4.75 – 7.00)	6.76 (0.23)	(6.10 – 7.00)	6.80 (0.28)	(5.95 – 7.00)	6.55 (0.34)	(5.70 – 7.00)
Positive Climate	6.49 (0.63)	(4.80 – 7.00)	6.74 (0.28)	(6.00 – 7.00)	6.80 (0.37)	(5.80 – 7.00)	6.54 (0.46)	(5.60 – 7.00)
Negative Climate ^a	6.83 (0.32)	(6.00 – 7.00)	6.97 (0.07)	(6.80 – 7.00)	6.94 (0.21)	(6.00 – 7.00)	6.89 (0.20)	(6.20 – 7.00)
Teacher Sensitivity	6.28 (0.85)	(3.80 – 7.00)	6.58 (0.46)	(5.40 – 7.00)	6.77 (0.34)	(6.00 – 7.00)	6.30 (0.59)	(5.00 – 7.00)
Regard for Student Perspectives	6.35 (0.85)	(4.00 – 7.00)	6.76 (0.35)	(5.60 – 7.00)	6.71 (0.42)	(5.40 – 7.00)	6.48 (0.56)	(5.40 – 7.00)
Classroom Organization Domain	6.26 (0.62)	(4.80 – 7.00)	6.45 (0.47)	(5.47 – 7.00)	6.59 (0.46)	(5.27 – 7.00)	6.08 (0.64)	(4.80 – 7.00)
Behavior Management	6.18 (0.81)	(4.60 – 7.00)	6.54 (0.52)	(5.00 – 7.00)	6.66 (0.53)	(5.00 – 7.00)	6.10 (0.79)	(4.40 – 7.00)
Productivity	6.42 (0.60)	(5.20 – 7.00)	6.50 (0.53)	(5.20 – 7.00)	6.64 (0.46)	(5.40 – 7.00)	6.28 (0.60)	(4.40 – 7.00)
Instructional Learning Formats	6.19 (0.64)	(4.60 – 7.00)	6.30 (0.59)	(5.00 – 7.00)	6.47 (0.61)	(5.20 – 7.00)	5.86 (0.81)	(4.00 – 7.00)
Instructional Support Domain	3.48 (1.39)	(1.40 – 6.60)	4.13 (0.91)	(3.07 – 6.47)	4.34 (1.35)	(1.73 – 6.07)	3.46 (0.98)	(1.60 – 5.67)
Concept Development	3.24 (1.28)	(1.20 – 6.40)	3.74 (0.92)	(2.80 – 6.20)	3.96 (1.35)	(1.60 – 5.80)	3.24 (1.13)	(1.40 – 6.00)
Quality of Feedback	3.44 (1.45)	(1.40 – 6.40)	4.15 (1.00)	(2.80 – 6.60)	4.43 (1.40)	(2.00 – 6.60)	3.45 (1.07)	(1.80 – 6.20)
Language Modeling	3.77 (1.56)	(1.60 – 7.00)	4.50 (1.06)	(2.80 – 6.60)	4.62 (1.43)	(1.60 – 6.60)	3.69 (1.01)	(1.40 – 5.80)

M=mean; SD=standard deviation

Figure B-1. Emotional support histograms by center

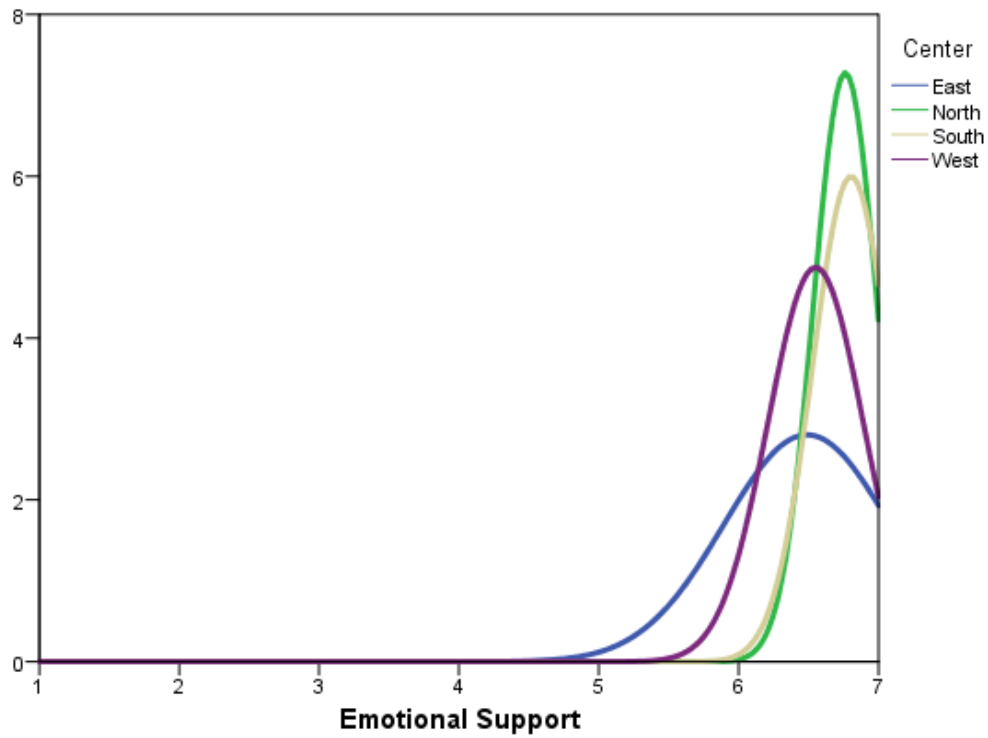


Figure B-2. Classroom organization histograms by center

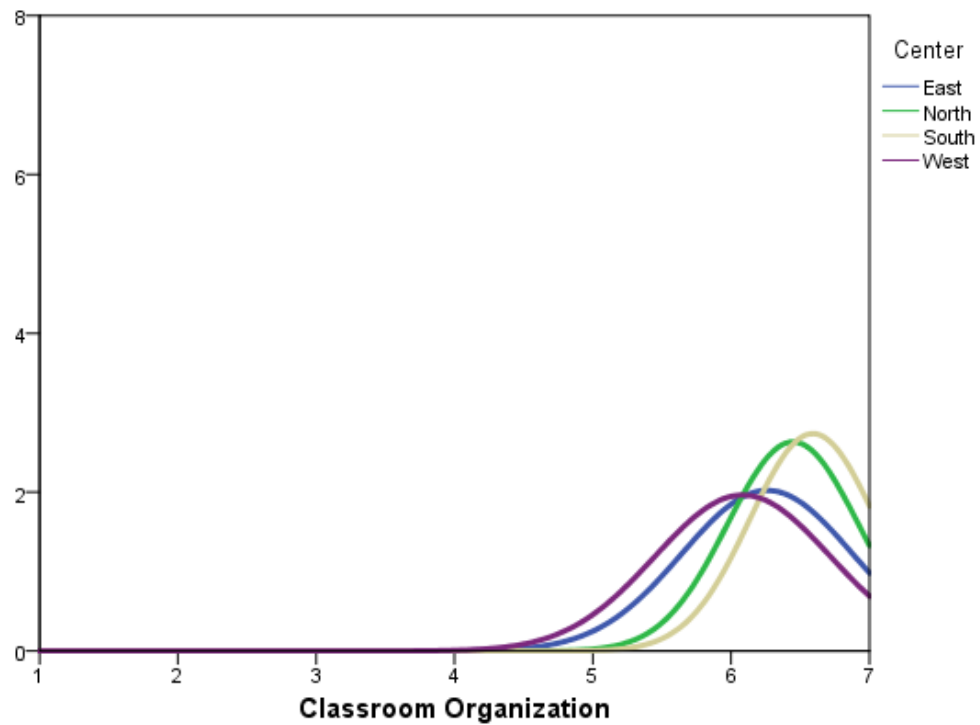
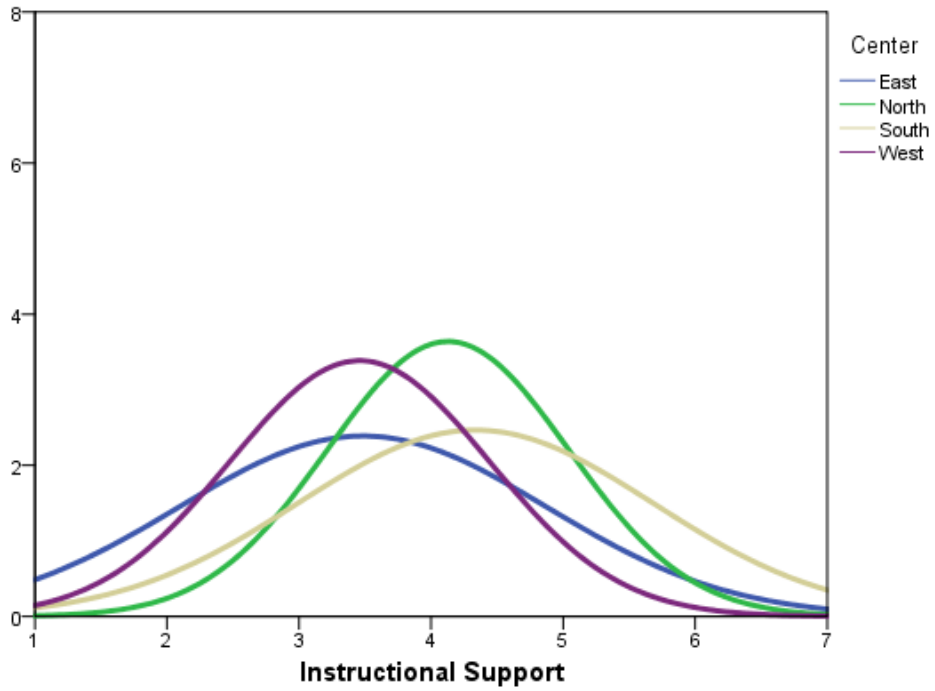


Figure B-3. Instructional support histograms by center



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.
- 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.
- 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.
- Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *The Bateria III Woodcock-Muñoz: Pruebas de aprovechamiento*. Itasca, IL: Riverside.
- Pianta, R., LaParo, K., & Hamre, B. (2008). *Classroom assessment scoring system*. Baltimore, MD: Brookes Publishing.
- 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.
- Schrank, F. A., McGrew, K. S., Mather, N., Wendling, B. J., & LaForte, E. M. (2014). *Woodcock-Johnson IV Tests of Achievement: Form A*. Riverside Publishing Company.
- 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.