

An Exploration of Pre-K 4 SA Education Centers in Year 10

Authors:

Emily Diaz, PhD

Lauren Decker-Woodrow, PhD

Eishi Adachi, PhD

Eleanore Spies, BS

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 was prepared by Westat under a Professional Services Agreement for the Program Assessment of the Pre-K 4 SA 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, nor does mention of trade names, commercial products, or organizations imply endorsement by the City of San Antonio.

All photos are for illustrative purposes only. Any person depicted in the photos is a model.

Table of Contents

Chapter

Executive Summary	1
Introduction	2
Research Questions	3
Evaluation Sample and Methods	4
Pre-K 4 SA Year 10 Sample	4
Methods	5
Evaluation Results	6
Child Attendance in Pre-K 4 SA	6
Attendance Rates Over Time	7
Pre-K 4 SA Teacher-Child Interaction Quality	7
Interaction Quality by Master Teacher Status	9
Interaction Quality by Center	9
Kindergarten Readiness – GOLD Results	9
Teaching Strategies GOLD	9
Direct Child Assessments	16
Woodcock-Johnson	16
Peabody Picture Vocabulary Test	16
Social-Emotional Assessment	18
Devereux Early Childhood Assessment	18
Conclusions and Looking Ahead	20
Limitations and Recommendations	20
References	21

Table of Contents *Continued*

Appendixes

Appendix A Evaluation Methods	23
Measures	23
Classroom Assessment Scoring System (CLASS)	23
Teaching Strategies GOLD.	24
The Woodcock-Johnson Achievement Test	24
Peabody Picture Vocabulary Test	24
Devereux Early Childhood Assessment	25
Analytic Approach	25
Appendix B Additional CLASS Results	28
Appendix C Additional Woodcock-Johnson and Bateria Results	30
Appendix D Additional Peabody Picture Vocabulary Results	32
Appendix E Additional Devereux Early Childhood Assessment Results	34
Appendix References	36

List of Tables

Table 1 Children who attended Pre-K 4 SA, by eligibility criteria	4
Table 2 Four-year-old children who attended Pre-K 4 SA, by partner district	5
Table 3 Pre-K 4 SA attendance over time (percent)	7
Table 4 Average Year 10 Pre-K 4 SA CLASS scores	8
Table 5 Pre-K 4 SA (3-Year-Old) and normed sample comparison results for six GOLD outcomes across time	11
Table 6 Pre-K 4 SA (4-Year-Old) and normed sample comparison results for six GOLD outcomes across time	14

Table of Contents *Continued*

List of Tables *Continued*

Table A-1	Descriptions of CLASS dimensions	23
Table A-2	Range of age equivalencies by assessment and language	24
Table A-3	Reliability evidence for the DECA by parent and teacher ratings	25
Table B-1	Average Year 10 CLASS scores by center	28
Table B-2	Classroom quality results comparing Master and Non-Master teachers	29
Table C-1	Proportion analysis results meeting age equivalency comparing fall 2022 and spring 2023	30
Table C-2	Age equivalencies results comparing fall 2022 and spring 2023 by subtest	30
Table C-3	Gap analysis age equivalency results comparing fall 2022 and spring 2023 by subtest	31
Table D-1	Descriptives of vocabulary performance levels by timepoint	32
Table D-2	Vocabulary growth results comparing fall 2022 and spring 2023 by score type	32
Table D-3	Descriptives of vocabulary score change comparing fall 2022 and spring 2023	33
Table E-1	Social-emotional results comparing levels in fall 2022 and spring 2023 by outcome	34
Table E-2	Social-emotional equivalencies results comparing fall 2022 and spring 2023 by outcome	35

Table of Contents *Continued*

List of Figures

Figure 1	 Average classroom quality scores for Pre-K 4 SA Year 10	8
Figure 2	 Pre-K 4 SA and Head Start average classroom quality scores.	9
Figure 3	 Proportion of children meeting age equivalency by subtest and time point	16
Figure 4	 Size of achievement gap (in months) between Pre-K 4 SA and normed sample by assessment and time point	16
Figure 5	 Proportion of Pre-K 4 SA children within each receptive vocabulary performance level by time point	17
Figure 6	 Proportion of Pre-K 4 SA children based on receptive vocabulary growth	17
Figure 7	 Proportion of children within each level by social-emotional outcome and time point	19

Executive Summary

Pre-K 4 SA served more than 2,000 children during its 10th year of implementation and was the first year serving 3-year-old children. The Year 10 evaluation of Pre-K 4 SA sought to address research questions regarding attendance, classroom quality, kindergarten readiness, early literacy and numeracy, receptive vocabulary, and social-emotional understanding during the pre-K year.

Pre-K 4 SA served slightly more boys (50.4%) than girls (49.6%) and more 4-year-old children (56.8%) than 3-year-old children (43.2%) during Year 10. Most Pre-K 4 SA children were Hispanic (57.7%), with the remaining children reported as multiracial (13.4%), White (11.2%), Black (9.5%), Asian (7.7%), and other ethnicities (0.5%). Approximately 89 (88.7%) percent of children attended Pre-K 4 SA based on eligibility criteria and 11.3 percent were tuition-paying children.

The average attendance rate for Pre-K 4 SA children was 84.5 percent, which increased slightly to 85.9 percent when children who withdrew were excluded. Average attendance increased compared to last year (82.6%). However, it remains lower compared to pre-pandemic years and is not surprising as similar trends have been evidenced across the nation post-pandemic.

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 and its evaluation partners 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, observed teachers displayed high levels of Emotional Support and Classroom Organization. Instructional Support was, on average, in the midrange. There were no significant between-center differences indicating classroom experiences of similar quality were offered across Pre-K 4 SA centers.

Kindergarten readiness outcomes for Pre-K 4 SA children (measured using Teaching Strategies' GOLD assessment) were compared from fall to spring for six outcomes: cognitive, literacy, mathematics, oral language, physical, and social-emotional. The results showed significant growth for Pre-K 4 SA children on all six outcomes.

Early literacy and numeracy results suggested children gained significant understanding across the year. However, most children were not performing at their age level of understanding and in need of additional educational supports. Receptive vocabulary results demonstrated children experienced significant growth and gained additional vocabulary. Moreover, almost one-quarter (23.4%) of children narrowed their achievement gap when comparing their learning to a normative group of children. Social-emotional results suggested children significantly gain skills and demonstrated significant growth in their understanding.

Taken together, results from the Year 10 evaluation suggest children are benefiting from participation in Pre-K 4 SA. Specifically, the Year 10 evaluation results indicate Pre-K 4 SA is providing high-quality instructional environments for more than 2,000 predominantly low-income children from across San Antonio. Limitations of the evaluation include the lack of a control group for comparison to a more similar group of children and reliance on teacher-reported measures of child outcomes in some instances.



Introduction

Improving children's kindergarten readiness and narrowing achievement gaps by providing high-quality early education has and will continue to receive considerable attention throughout the United States (Barnett, 2011; Campbell, Ramey, Pungello, Sparling & Miller-Johnson, 2002; Heckman, Moon, Pinto, Savelyev & Yavitz, 2010; Hill, Gormley & Adelstein, 2015; Reynolds, Temple, White, Ou & Robertson, 2011; Rolnick & Grunewald, 2003). Coupled with the importance of high-quality education experiences is the fact that children who need such experiences the most are often those who do not receive them. Previous research indicated children from racially marginalized communities, economically disadvantaged backgrounds, and children whose primary language is a language other than English are more often exposed to lower quality instruction and learning environments across the United States (Bassok & Galdo, 2016; Valentino, 2018). Moreover, providing quality environments is vital to improving children's social-emotional, behavior, and achievement outcomes (Burchinal et al., 2010; Perlman et al., 2016).

Since federal and state governments provide limited public funding for early learning, municipal governments are increasingly using funding sources in creative ways to provide more equitable access to high-quality early childhood education and care. Pre-K 4 SA, a municipality

funded initiative in San Antonio, Texas, serves approximately 2,000 children who are at risk for falling behind their peers and lacking in kindergarten readiness. Four state-of-the-art early educational centers serve as model sites demonstrating the education potential for young children when they have access to highly skilled and well-compensated teachers, an evidenced-based curriculum, instructional supports, and content specific learning. The 2022-23 school year marks an increase in the outreach of Pre-K 4 SA as this is the first year Pre-K 4 SA has served 3-year-old children.

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. Over the previous 9 years, the evaluation has consistently explored who participates in Pre-K 4 SA, attendance in the program, teacher-child instructional quality, and kindergarten readiness outcomes. The purpose of the current report is to present Year 10 evaluation findings for the program. Investigations included (1) information on child attendance and classroom quality; (2) outcome analysis results from the Teaching Strategies GOLD assessment, which is the primary outcome of interest at the end of the pre-K year; (3) outcome analysis results of children's early literacy, numeracy, and receptive vocabulary; and (4) outcome analysis results of children's social-emotional development.



Research Questions

The Year 10 (2022-23) evaluation of Pre-K 4 SA addressed the following six main research questions:

- 1. A.** What were the reported levels of child attendance during the pre-K year?
 - B.** In what ways have attendance rates changed since the COVID-19 pandemic?
- 2. A.** What was the observed teacher-child interaction quality of Pre-K 4 SA classrooms in Year 10?
 - B.** Did master teachers of Pre-K 4 SA classrooms have higher observed teacher-child interaction quality in Year 10?
 - C.** Were there any significant differences across the education centers?
- 3. A.** How did Pre-K 4 SA children compare to the normed sample on GOLD outcomes over the year?
 - B.** Did Pre-K 4 SA children demonstrate significant growth on GOLD outcomes over the year?
- 4. A.** What proportion of a random sample of Pre-K 4 SA children performed at or above their age level in early literacy and early numeracy over the year? To what extent did the proportion change over the year?
 - B.** Did a random sample of Pre-K 4 SA children demonstrate significant growth in early literacy and early numeracy over the year?
 - C.** Did a random sample of Pre-K 4 SA children experience accelerated learning to help narrow achievement gaps in early literacy and early numeracy?
- 5. A.** What were the performance levels of a random sample of Pre-K 4 SA children over the year?
 - B.** Did a random sample of Pre-K 4 SA children demonstrate significant growth in vocabulary over the year?
 - C.** What types of vocabulary growth did a random sample of Pre-K 4 SA children demonstrate over the year?
- 6. A.** What were the levels of Pre-K 4 SA children's social-emotional understanding over the year? To what extent did the levels change over the year?
 - B.** Did Pre-K 4 SA children demonstrate significant growth in social-emotional learning over the year?



Evaluation Sample and Methods

In this section, demographic characteristics for the sample are provided for children served during the 2022-23 school year as well as a brief discussion of the methods used.



Pre-K 4 SA Year 10 Sample

Data were provided for 2,050 children in Year 10. This is the first year Pre-K 4 SA served children ages 3 and 4. Pre-K 4 SA served slightly more 4-year-old children (56.8%) than 3-year-old children¹ (43.2%), and slightly more boys (50.4%) than girls (49.6%). Most Pre-K 4 SA children were Hispanic (57.7%), with the remaining children reported as multiracial (13.4%), White (11.2%), Black (9.5%), Asian (7.7%), and other ethnicities (0.5%). Furthermore, similar proportions of boys and girls were served based on age ($X^2(1) = 0.523, p = .470$). However, differences were found for race/ethnicity based on age ($X^2(5) = 27.604, p < .001$). Asian child served were more likely to be 4 years old as opposed to 3 years old and multiracial children served were more likely to be 3 years old as opposed to 4 years old.

Table 1 displays the ways in which children attended Pre-K 4 SA based on eligibility criteria. It is important to note there is overlap between criteria as children could qualify in more than one category. The majority attended based on the Texas Education Agency low-income² criteria (52.8%) or the middle-income criteria³ (35.9%). Children also attended if their family had military affiliation⁴ (8.0%) or they were an English language learner (3.8%). An additional 231 (11.3%) children paid tuition to attend.

Table 1 | Children who attended Pre-K 4 SA, by eligibility criteria

Eligibility criteria	Number of children	Percentage (%) of total eligible children
Low-Income	1,082	52.8
Middle-Income	736	35.9
Military Affiliated	164	8.0
English Language Learner	78	3.8
Tuition	231	11.3
Eligible total	2,050	100.0

¹There were 32 (1.6%) children who were included as 3-year-olds that were almost 3 on the first day of school with ages ranging from 2.94 to 2.998.

²Children attended based on this category if their parent/guardians' income is defined as 250% of the federal poverty guideline.

³Children attended based of this category if their parent/guardians' income is between the TEA low-income criteria and 85% of the state median income.

⁴Children attended based on this category if their parent/guardian is active duty, retired, or in the reserves. This also includes grandparents who are retired military and caring for children.

Only 4-year-old children are eligible to be affiliated with a partner district. Among all 4-year-old children who attended Pre-K 4 SA, a little over half (52.7%; n = 614) were affiliated with a partner district. The majority represented three districts: Northside Independent School District (ISD), North East ISD, and San Antonio ISD.⁵ Table 2 includes the percentage of 4-year-old children per represented school district.

Methods

All six research questions were addressed by analyzing existing Pre-K 4 SA databases, as well as results from classroom observations and direct child assessments. To address the first question about 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, Westat and its partners collected and analyzed data from the Classroom Assessment Scoring System (CLASS) for Pre-K 4 SA classrooms (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 10 evaluation consisted of five 20-minute cycles, followed by 10-minute coding periods.

The third research question was addressed through inferential tests of differences, which were conducted on the Pre-K 4 SA administered Teaching Strategies GOLD assessment outcomes. GOLD is a teacher-reported measure that collects information on children’s progress on 36 objectives, three times throughout the year, across six main categories: cognitive, literacy, oral language, mathematics, physical, and social-emotional.

To address the fourth set of questions, data collected by Pre-K 4 SA were submitted to Westat and analyzed descriptively and inferentially. The fourth research question was addressed through descriptive and inferential analyses of a random sample of Pre-K 4 SA children in the fall and spring, on the Woodcock-Johnson assessment. Two direct assessments, early literacy (Letter-Word) and numeracy (Applied Problems), were administered. These two assessments are subtests from the Woodcock-Johnson Test of Achievement-IV (Schrank, Mather, & McGrew, 2014) and matching subtests from the Batería III, Spanish assessment (Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005; see Appendix A for more detailed information). They were chosen because they are widely used in early childhood and complement the GOLD findings by providing

Table 2 | Four-year-old children who attended Pre-K 4 SA, by partner district

District name	Number of children	Percentage of total children
Northside	272	44.3
North East	102	16.6
San Antonio	99	16.1
Edgewood	46	7.5
Southwest	29	4.7
East Central	29	4.7
Harlandale	23	3.8
South San	14	2.3
Total	614	100.0

Note: Children counted by district attend the program at no cost. Of all 4-year-old children, 614 (52.7%) were affiliated with a partner district.

⁵ These same three districts were also the majority representation in Years 1–9 (2013-14 to 2021-22).

additional insights from a different perspective: that of a trained assessor as compared to a teacher report (Bloom & Weiland, 2014; McCormick, 2022; Puma, Bell, Cook, & Heid, 2010; Weiland, 2016). The GOLD findings provide an overall perspective and measure multiple aspects of early literacy (e.g., phonological awareness, phonics, and word recognition) and numeracy (e.g., number concepts and operations, spatial relationships and shapes, and knowledge of patterns). Letter-Word findings are more nuanced and measure symbolic learning and identification of isolated letters and words, while Applied Problems measures a child's ability to apply simple number concepts and solve math problems. For each outcome, analyses were conducted for children with data in both time points (fall and spring). To compare to a nationally representative normative sample, raw scores were converted into age equivalents.⁶ To answer research question 3A, descriptive and inferential analyses were conducted. The goal was to determine and compare children's early literacy and early numeracy levels. To answer research question 3B, inferential tests of differences were conducted. To answer question 3C, the difference between a child's early literacy and early numeracy age level and their biological age was calculated and used for analyses. Inferential tests were conducted to determine if the difference was significantly reduced. (See Appendix A for more detailed information on the evaluation methodology, including detailed information pertaining to measures used.)

To address the fifth set of research questions, data collected by Pre-K 4 SA were submitted to Westat and analyzed descriptively and inferentially. A direct assessment of vocabulary was administered to a random sample of children in fall and spring from the Peabody Picture Vocabulary Test-V (PPVT-Dunn & Dunn, 2019). (See Appendix A for more detailed information.) Like Woodcock-Johnson this assessment was chosen because it is widely used in early childhood and complements the GOLD findings by providing additional insights from a different perspective: that of a trained



assessor as compared to a teacher report (Puma, Bell, Cook, & Heid, 2010). The GOLD findings provide an overall perspective and measure multiple aspects of early literacy (e.g., phonological awareness, phonics, and word recognition). The PPVT findings are more nuanced and measure receptive vocabulary knowledge and understanding.

To address the sixth set of research questions, data collected by Pre-K 4 SA were submitted to Westat and analyzed descriptively and inferentially. A teacher report assessment of social-emotional learning, the Devereux Early Childhood Assessment, was administered to children in fall and spring (LeBuffe & Naglieri, 2012). (See Appendix A for more detailed information.)

Evaluation Results

Child Attendance in Pre-K 4 SA

Children began attending Pre-K 4 SA at different times. Most children (80.44%) began at the start of the academic year (August 15, 2022). The last date a child began attending Pre-K 4 SA was April 26, 2023⁷. 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 2 to 175, with an average of 157 days. Average percentage attendance across all children was 84.5 percent. When considering children who attended Pre-K 4 SA through the year (i.e., who did not withdraw), the average number of membership days rose to 167 and the attendance percentage increased to 85.9 percent.

⁶The raw scores were converted into age equivalent values based on norms provided in the technical manual (McGrew, LaForte, & Schrank, 2014).

⁷Although some children did not begin attendance at Pre-K 4 SA until late spring, more than 95% of all children were in attendance by the end of the 2022 calendar year.

Over the course of the year, 224 children (10.9%) withdrew from Pre-K 4 SA. The earliest withdrawal occurred on August 17, 2022, and the latest on May 25, 2023. Approximately 40 percent (41.5%; n=93) of the withdrawals occurred before the end of December. We found no significant differences between children who did and did not withdraw in terms of gender ($t(2,043.8) = 1.14, p = .26$), whether they were 3 or 4 years old on the first day of school ($t(282.48) = -1.42, p = .16$) attending based on the middle-income criteria ($t(285.60) = 1.74, p = .08$), attending based on English language learner status ($t(255.41) = -1.64, p = .10$), attending based on military affiliation ($t(261.43) = -1.81, p = .07$). We did find significant differences for children who attended based on race/ethnicity ($F(5, 2,044) = 2.62, p = .02$), the TEA low-income criteria ($t(285.62) = -3.78, p < .001$), children dually enrolled with a partner school district ($t(272.69) = -3.19, p = .002$), and children attending on tuition status ($t(360.43) = 4.36, p < .001$). Asian children were more likely to withdraw from Pre-K 4 SA when compared to all other children. Children attending based on the TEA low-income criteria were more likely to withdraw from Pre-K 4 SA than children not attending based on this criterion. Children who were dually enrolled with a partner district were more likely to withdraw from Pre-K 4 SA than children not enrolled with a partner district. Children whose family was paying tuition were less likely to withdraw from Pre-K 4 SA than children whose family was not paying tuition.

Attendance Rates Over Time

Prior to the pandemic, attendance rates had remained relatively stable and consistently remained between 91-94 percent. Table 3 displays attendance for all children who attended the program as well as attendance for the subgroup of children who did not withdraw from the program. It is not surprising that attendance dropped below this range in recent years given the COVID-19 pandemic.⁸ Furthermore, despite these lower attendance rates, there was an increase from last year.

Pre-K 4 SA Teacher-Child Interaction Quality

Across the four education centers, Pre-K 4 SA classrooms (n=90)⁹ were observed using the CLASS (Pianta, La Paro, & Hamre, 2008) during Year 10. Scores for the Emotional Support domain ranged from 5.10 to 7.00 (on a 1-7 scale), with most scores in the high range (average score of 6.44), suggesting observed teacher-child interactions, in this domain, were most often rated as high quality. Similarly, with an overall score in the high range, Classroom Organization domain scores ranged from 3.73 to 7.00, which suggests classrooms showed effective interactions regarding Classroom Organization (average score of 6.04). Finally, Instructional Support domain scores ranged from 1.60 to 6.73, with an average

Table 3 | Pre-K 4 SA attendance over time (percent)

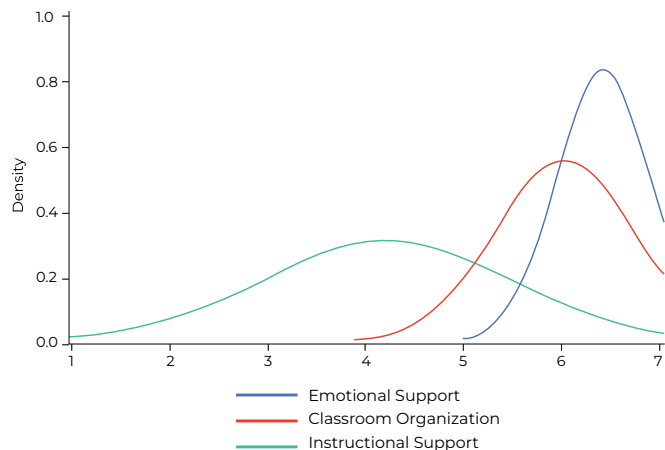
Enrollment Status	Pre-Pandemic						Pandemic			Post-Pandemic
	Year 1 2013-14	Year 2 2014-15	Year 3 2015-16	Year 4 2016-17	Year 5 2017-18	Year 6 2018-19	Year 7 2019-20	Year 8 2020-21	Year 9 2021-22	Year 10 2022-23
All enrolled children	92.3	91.3	92.5	92.4	91.0	91.5	91.0	88.2	82.6	84.5
Children who did not withdraw	93.7	92.5	93.6	93.6	92.4	92.6	92.2	90.4	85.3	85.9

⁸ Similar trends have been found based on the *National Survey of Public Education's Response to COVID-19* (Carminucci, Hodgman, Rickles, and Garet, 2021). Furthermore, recent rates of chronic absenteeism across the nation have increased compared to typical school years prior to the pandemic (National Center for Education Statistics, 2022).

⁹ Ten classrooms (10%) were found to have invalid data and were removed from analyses.

score in the middle range (4.24), which suggests in some observed interactions teachers provided support that extended children’s thinking or asked questions that encouraged children to analyze and reason. This is a noteworthy finding as it demonstrates an increase of one point from last year’s results (3.24), suggesting teachers gained skills in this domain. Each of the Year 10 CLASS domain scores is represented visually in Figure 1.

Figure 1. | Average classroom quality scores for Pre-K 4 SA Year 10



Looking further into the average Emotional Support domain scores, approximately 19 percent of classrooms (n=17) were observed in the middle range, while 81 percent of classrooms observed provided high levels (n=73). Forty-one percent of classrooms (n=37) were observed providing middle-range Classroom Organization quality, while the remaining 59 percent (n=53) provided high levels. Finally, 19 percent of the classrooms (n=17) were observed providing low levels of Instructional Support, 73 percent (n=66) provided midrange levels, and 8 percent (n = 7) provide high levels. Table 4 provides average scores by each of the 10 dimensions and 3 domains.

Past research using the CLASS has often noted the lower scores commonly seen in the Instructional Support domain (Early Childhood Learning & Knowledge Center, 2020; La Paro, Pianta, & Shuhlman, 2004; Locasale-Crouch et al., 2007; Mashburn et al., 2008). To place Pre-K 4 SA CLASS scores in context, the Office of Head Start found in their 2019-20 annual review¹⁰ average scores across the United States and the top 10 percent of Head Start grantees were lower than those found in the current study with one exception (Early Childhood Learning & Knowledge Center, 2020). The Classroom Organization domain for Pre-K 4

Table 4 | Average Year 10 Pre-K 4 SA CLASS scores

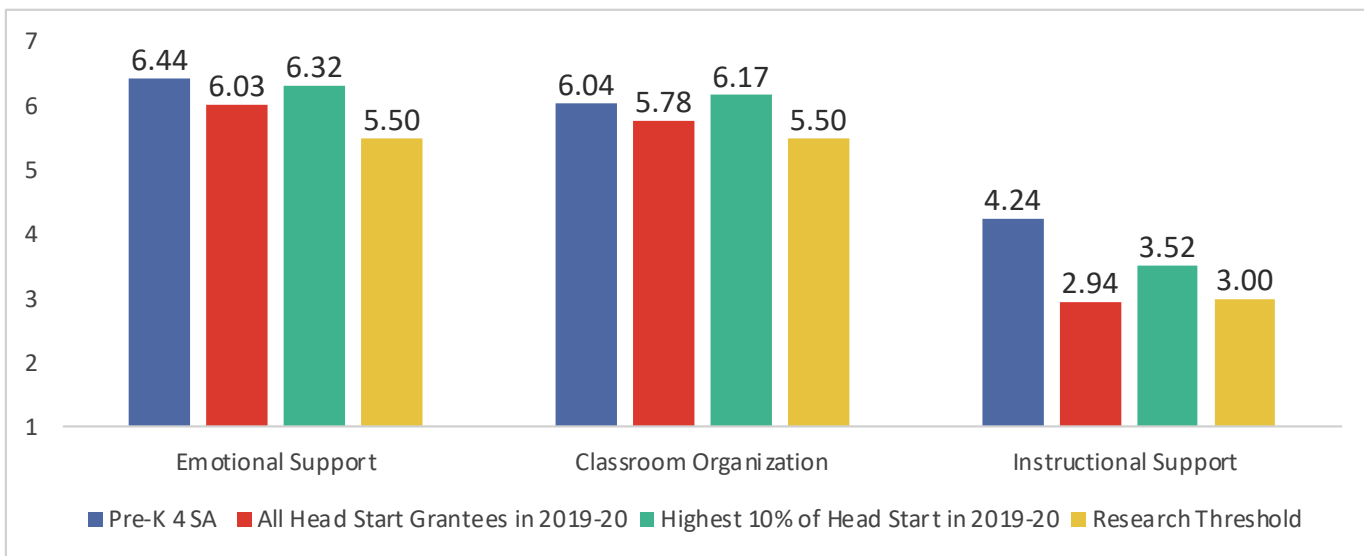
CLASS outcome	Average	Total range observed	Standard deviation
Emotional Support domain	6.44	5.10-7.00	0.48
Positive climate	6.39	4.60-7.00	0.67
Negative climate ^a	6.95	6.20-7.00	0.15
Teacher sensitivity	6.17	4.20-7.00	0.68
Regard for student perspectives	6.26	4.20-7.00	0.63
Classroom Organization domain	6.04	3.73-7.00	0.72
Behavior management	5.89	3.60-7.00	0.83
Productivity	6.28	3.80-7.00	0.70
Instructional learning formats	5.95	3.40-7.00	0.78
Instructional Support domain	4.24	1.60-6.73	1.29
Concept development	4.04	1.60-6.60	1.31
Quality of feedback	4.33	1.20-7.00	1.47
Language modeling	4.36	1.80-7.00	1.25

^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.

¹⁰This is the most recent year of publicly available data for comparison; most likely due to the pandemic.

SA was slightly lower (0.13 points) than the top 10 percent of Head Start grantees. In Figure 2, Pre-K 4 SA scores are visually depicted with the national Head Start and top 10 percent of Head Start average scores, as well as the research thresholds.

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 to compare Pre-K 4 SA and Head Start classrooms for this evaluation.

Source: Early Childhood Learning & Knowledge Center (2020). *A national overview of grantee CLASS scores in 2020*.

Available at: <https://eclkc.ohs.acf.hhs.gov/data-ongoing-monitoring/article/national-overview-grantee-class-scores-2020>.

Interaction Quality by Master Teacher Status

The three CLASS domains and 10 dimensions were analyzed to determine if there were significant differences in classroom teacher-child interactions based on whether a Master teacher led the classroom. Significant differences were found for the Classroom Organization domain and all three associated dimensions (Behavior Management, Productivity, and Instructional Learning Formats) in favor of Master teachers (for more detailed information, see Appendix B Table B-2). Classrooms led by Master teachers were observed to have higher quality, ranging from 0.33 to 0.41 points. There were no significant differences found for the Emotional Support and Instruction Support domains and their associate dimensions. These findings indicate classroom experiences were of similar quality across all Pre-K 4 SA teachers for these two domains and dimensions.

Interaction Quality by Center

The three CLASS domains and 10 dimensions were analyzed to determine if there were significant differences in classroom teacher-child interactions based on center location. No significant differences¹¹ were found by center, which indicates classroom experiences of similar quality were offered across all Pre-K 4 SA centers.

Kindergarten Readiness – GOLD Results

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. Three-year-old children (77.7%; n = 688) and 4-year-old children (84.6%; n = 985) were included

¹¹ A significant difference was found for Negative Climate by center ($F(3,86) = 5.36, p = 0.002$). However, the results are not practically significant as all the center average scores are in the high range.

in analyses¹² if they had outcome data for all three time points¹³ in at least one of the following six outcomes: cognitive, literacy, mathematics, oral language, physical, and social-emotional. As data were not collected on a comparison or control group, comparisons were conducted using the nationally representative normed data¹⁴ for the GOLD assessment (Lambert, 2020). The results are presented separately for children aged 3 and 4 years old as the norms vary depending on age level.

Three-Year-Old Results

When starting Pre-K 4 SA, 3-year-old children began the fall significantly below the normed sample on three of the six GOLD outcomes (oral language, physical, and social-emotional) and on par¹⁵ with the normed sample on two outcomes (cognitive and literacy). By spring, Pre-K 4 SA children scored statistically significantly ($p < .001$) higher than the normed sample on three outcomes (cognitive, literacy, and physical¹⁶). Effect sizes (measured by Hedges' g) for the significant results were all small (0.18 for cognitive, 0.20 for literacy, and 0.07 for physical). Over the course of the pre-K year, Pre-K 4 SA children gained an additional 12.38 scale score points (17.7% more) in cognitive, 10.82 scale score points (20.5% more) in literacy, and 4.93 scale score points (6.9% more) in physical.

Spring results for the oral language and social-emotional outcomes indicated the initial gaps between Pre-K 4 SA children and the normed sample were eliminated by the end of the school year. By spring, no significant differences were found between Pre-K 4 SA children and the normed sample for these two outcomes. To achieve the elimination of the initial gaps, Pre-K 4 SA children gained an additional 1.43 scale score points (2.0% more) in oral language and 1.26 scale score points (2.0% more) in social-emotional.

Across all three timepoints, Pre-K 4 SA children were significantly higher than the normed sample in mathematics. Furthermore, over the course of the pre-K year, Pre-K 4 SA children gained an additional 17.13 scale score points (27.6% more) in mathematics than the normative group of children. This trend agrees with previous findings. Over the course of our series of evaluations, the 4-year-old 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. See Table 5 for more information.



¹²As children were not randomly sampled, demographic tests of differences were conducted to determine if the sample of children included and excluded from analyses were similar (see Appendix A Analytic Approach for more detailed information).

¹³Out of 175 3-year-old children not able to be included in Gold analyses, 53 (30.3%) appeared to have invalid data for at least one time point. In cognitive, four children were excluded, in literacy 38 children were excluded, in mathematics three children were excluded, in oral language 37 children were excluded, in physical two children were excluded, and in social-emotional three children were excluded. Out of 141 4-year-old children not able to be included in Gold analyses, 56 (39.7%) appeared to have invalid data for at least one time point. In cognitive, four children were excluded, in literacy 37 children were excluded, in mathematics two children were excluded, in oral language 45 children were excluded, in physical one child was excluded, and in social-emotional two children were excluded.

¹⁴Pre-K 4 SA children were compared to the updated normed sample based on age bands (Lambert, 2020).

¹⁵While Pre-K 4 SA children were technically higher or lower in their score than the normed sample, this difference was not statistically significant.

¹⁶The p -value for the physical outcome was 0.0167.

Table 5 | Pre-K 4 SA (3-Year-Old) 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; Red line = normed sample)
Cognitive	Fall	372.41	374.62	-2.21	-1.077	1112.55	0.2817	Non-Significant	No difference	
	Winter	422.27	415.58	6.69	3.512	1158.27	0.0005	Significant	Pre-K 4 SA	
	Spring	456.80	444.42	12.38	5.995	1153.58	0.0000	Significant	Pre-K 4 SA	
Literacy	Fall	428.52	428.30	0.22	0.11	1065.02	0.9110	Non-Significant	No difference	
	Winter	469.46	461.74	7.72	4.37	970.39	0.0000	Significant	Pre-K 4 SA	
	Spring	491.90	481.08	10.82	6.67	1065.84	0.0000	Significant	Pre-K 4 SA	
Mathematics	Fall	295.24	290.08	5.16	2.32	1000.36	0.0206	Significant	Pre-K 4 SA	
	Winter	343.04	327.62	15.42	8.55	1075.48	0.0000	Significant	Pre-K 4 SA	
	Spring	369.18	352.05	17.13	9.33	1058.69	0.0000	Significant	Pre-K 4 SA	

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).

Table 5 | Pre-K 4 SA (3-Year-Old) 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; Red line = normed sample)
Oral Language	Fall	388.89	402.14	-13.25	-4.288	882.04	0.0000	Significant	Normed	
	Winter	438.89	442.82	-3.93	-1.237	853.62	0.2164	Non-Significant	No difference	
	Spring	475.50	474.07	1.43	0.428	869.78	0.6686	Non-Significant	No difference	
Physical	Fall	477.39	496.07	-18.68	-7.670	1046.04	0.0000	Significant	Normed	
	Winter	532.48	537.37	-4.89	-2.385	1133.17	0.0173	Significant	Normed	
	Spring	572.52	567.59	4.93	2.397	1231.55	0.0167	Significant	Pre-K 4 SA	
Social-emotional	Fall	365.04	379.88	-14.84	-7.547	1050.62	0.0000	Significant	Normed	
	Winter	414.11	416.85	-2.74	-1.499	1079.04	0.1341	Non-Significant	No difference	
	Spring	444.07	442.81	1.26	0.675	1180.38	0.4999	Non-Significant	No difference	

df = degrees of freedom.

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

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

Four-Year-Old Results

When starting Pre-K 4 SA in the fall, 4-year-old children began significantly below the normed sample on all six GOLD outcomes (cognitive, literacy, mathematics, oral language, physical, and social-emotional). By spring, Pre-K 4 SA children scored statistically significantly ($p < .001$) higher than the normed sample on one outcome (mathematics). The effect size (Hedges' g) for the significant result was small (0.12 for mathematics). Over the course of the pre-K year, Pre-K 4 SA children gained an additional 6.08 scale score points (9.5% more) in mathematics.

Spring results for the cognitive outcome indicated the initial gap between Pre-K 4 SA children and the normed sample were eliminated by the end of the school year. By spring, no significant difference was found between Pre-K 4 SA children and the normed sample for this outcome. To achieve the elimination of the initial gap, Pre-K 4 SA children gained an additional 2.08 scale score points (2.6% more) in cognitive.

A gap still remained for four out of the six outcomes (literacy, oral language, physical, and social-emotional). Literacy was reduced by 70.3 percent at spring (a reduction from an initial gap of more than 12 scale score points to approximately 3 scale score points). Oral language was reduced by 66.7 percent at spring (a reduction from an initial gap of more than 29 scale score points to approximately 10 scale score points). Physical was reduced by 88.7 percent at spring (a reduction from an initial gap of more than 31 scale score points to approximately 3 scale score points). Social-emotional was reduced by 61.0 percent at spring (a reduction from an initial gap of more than 30 scale score points to approximately 12 scale score points). See Table 6 for more information.



Table 6 | Pre-K 4 SA (4-Year-Old) 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; Red line = normed sample)
Cognitive	Fall	408.84	427.94	-19.10	-11.508	1870.77	0.0000	Significant	Normed	
	Winter	464.98	472.29	-7.31	-4.539	1748.00	0.0000	Significant	Normed	
	Spring	508.54	506.46	2.08	1.127	1683.32	0.2601	Non-Significant	No difference	
Literacy	Fall	462.83	474.47	-11.64	-8.05	1582.36	0.0000	Significant	Normed	
	Winter	500.60	506.52	-5.92	-4.56	1357.29	0.0000	Significant	Normed	
	Spring	523.56	527.01	-3.45	-2.60	1442.16	0.0093	Significant	Normed	
Mathematics	Fall	334.33	341.41	-7.08	-4.79	1746.58	0.0000	Significant	Normed	
	Winter	380.95	379.14	1.81	1.38	1729.88	0.1681	Non-Significant	No difference	
	Spring	411.33	405.25	6.08	4.28	1745.01	0.0000	Significant	Pre-K 4 SA	

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).

Table 6 | Pre-K 4 SA (4-Year-Old) 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; Red line = normed sample)
Oral Language	Fall	432.51	461.29	-28.78	-11.905	1350.4	0.0000	Significant	Normed	
	Winter	489.41	507.14	-17.73	-6.708	1253.64	0.0000	Significant	Normed	
	Spring	535.37	544.96	-9.59	-3.216	1234.65	0.0013	Significant	Normed	
Physical	Fall	517.38	547.91	-30.53	-16.717	1768.04	0.0000	Significant	Normed	
	Winter	579.24	593.33	-14.09	-8.263	1712.19	0.0000	Significant	Normed	
	Spring	625.05	628.50	-3.45	-1.981	1824.23	0.0477	Significant	Normed	
Social-Emotional	Fall	396.09	426.44	-30.35	-18.326	1591.52	0.0000	Significant	Normed	
	Winter	449.19	466.51	-17.32	-11.532	1728.17	0.0000	Significant	Normed	
	Spring	485.50	497.35	-11.85	-6.878	1747.25	0.0000	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.”

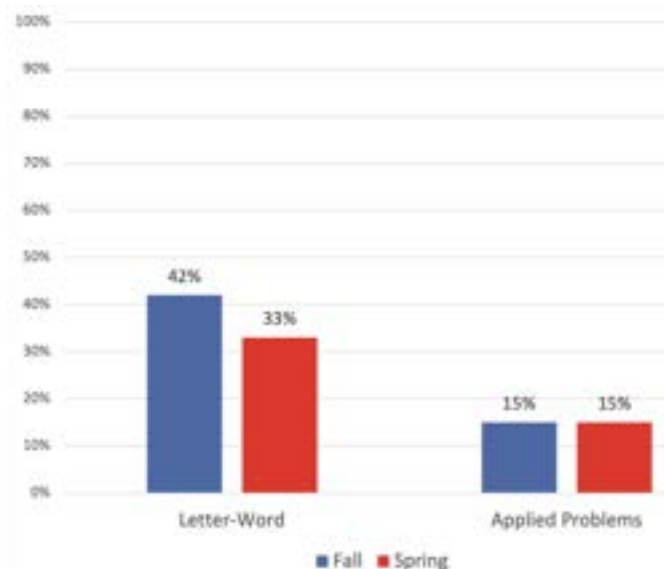
Note: Group mean information is presented in scaled scores. The Adjusted Significance column indicates significance levels (p-values) after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg technique (1995).

Direct Child Assessments

Woodcock-Johnson

Westat analyzed data from a random sample (n = 60) collected by Pre-K 4 SA on two subtests of a direct child achievement assessment: *Letter-Word and Applied Problems* from the Woodcock-Johnson Psycho-Educational Battery-Fourth Edition and Bateria III. For early literacy, 42 percent of Pre-K 4 SA children were at or above their age level in fall and, at the end of the year, 33 percent of children were at or above their age level, as displayed in Figure 3. For early numeracy, in fall, 15 percent of children were at or above their age level, and at the end of the year, 15 percent of children were at or above their age level. Taken together, these findings imply most children are performing below their age level at both time points and are behind what would be considered “ready” from a nationally representative lens. As this is the first year conducting this analysis, there are no prior trends in which to compare. The evaluation of the subsequent school year (2023-24) will conduct the same analysis and can be used to determine if similar trends are observed. There was no difference in children performing at or above their age level in early literacy (e.g., identifying isolated letters and words) and numeracy (e.g., analyze and solve math problems by applying simple number concepts) in spring compared to fall (for more detailed information, see Appendix C Table C-1).

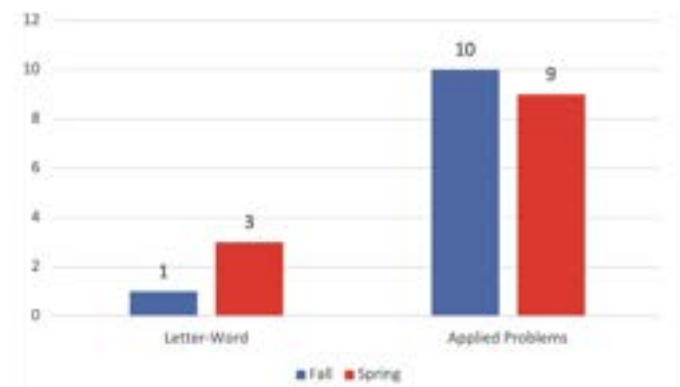
Figure 3. | Proportion of children meeting age equivalency by subtest and time point



When considering growth across the year, the findings indicated there was significant growth for both outcomes. For early literacy, there was approximately 4 months of growth in learning and for early numeracy approximately 6 months of growth in learning during 6 months of time. (For more detailed information, see Appendix C, Table C-2.) Therefore, these findings suggest children gained significant knowledge of early literacy and numeracy from fall to spring.

When analyzing if Pre-K 4 SA children experienced accelerated learning to help narrow achievement gaps in early literacy and early numeracy, the results indicated that accelerated learning did not occur. (For more detailed information, see Appendix C, Table C-3.) As displayed in Figure 4, there was a gap between children’s understanding and their age equivalent norms at both time points. For early literacy, in fall, on average, children were 1 month below the norms, and in spring, children were 3 months below the norms. For early numeracy, in fall, on average, children were 10 months below the norms, and in spring, children were 9 months below the norms. Therefore, the gap between Pre-K 4 SA children and the national norm did not significantly decrease for either outcome.

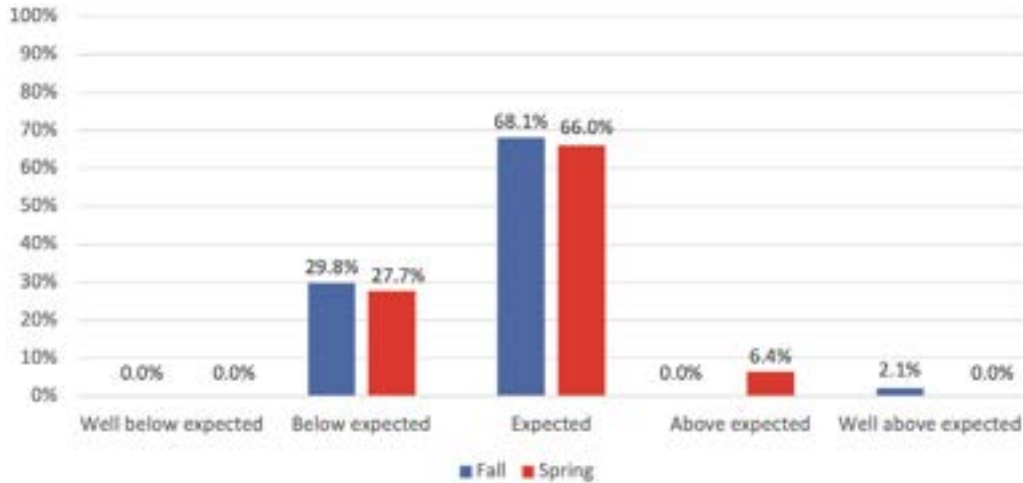
Figure 4. | Size of achievement gap (in months) between Pre-K 4 SA and normed sample by assessment and time point



Peabody Picture Vocabulary Test

To measure children’s receptive vocabulary, Westat analyzed data from a random sample (n=47) collected by Pre-K 4 SA on the Peabody Picture Vocabulary Test. For this assessment, children are presented pictorial images of words and asked to select the correct one.

Figure 5. | Proportion of Pre-K 4 SA children within each receptive vocabulary performance level by time point

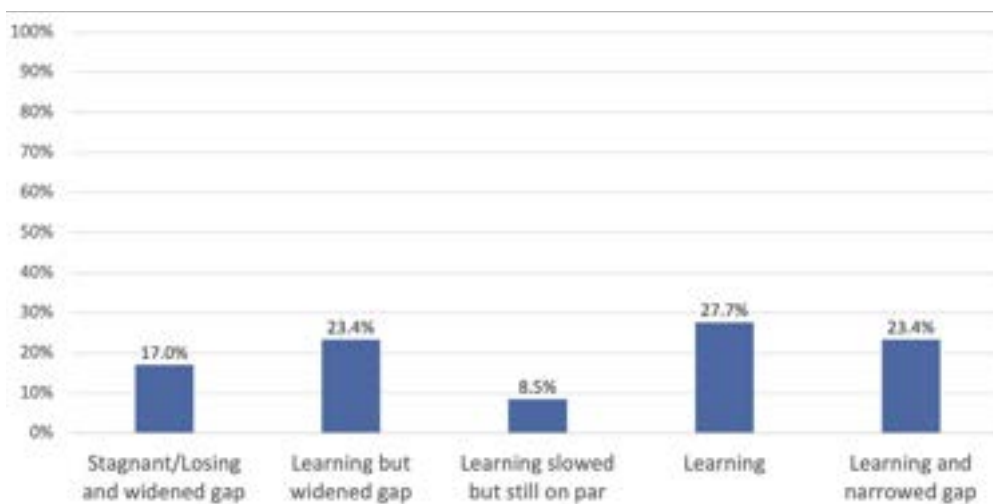


To evaluate children’s understanding, their scores are converted into five performance levels: (1) well below expected, (2) below expected, (3) expected, (4) above expected, and (5) well above expected. These levels are based on a normative sample and represents the developmental trajectory of children based on their age. When considering the performance levels of Pre-K 4 SA children over the year, the majority of children were performing in the expected range in both fall and spring (see Figure 5).

To better understand how children were progressing throughout the year, analyses of vocabulary growth were conducted to assess changes over time and comparisons to a normative sample. Together, these two findings provide a holistic perspective of children’s learning across the year. Results indicated overall children

experienced significant growth and gained additional vocabulary (approximately 5 growth scale points) when comparing their fall and spring scores across the year (for more detailed information, see Appendix D, Table D-2). When comparing children’s growth across the year to the normative sample, there was no significant difference between the two groups. This indicates Pre-K 4 SA children were progressing at a rate that is typical of children of their same age. As indicated in Figure 6, children exhibited five different types of growth. Most children learned new vocabulary or improved their understanding of current vocabulary. Moreover, for 11 children (23.4%), the gap between the Pre-K 4 SA child’s performance and the performance of typically developing children of the same age has narrowed (for more detailed information, see Appendix D, Table D-3).

Figure 6. | Proportion of Pre-K 4 SA children based on receptive vocabulary growth





Social-Emotional Assessment

Devereux Early Childhood Assessment

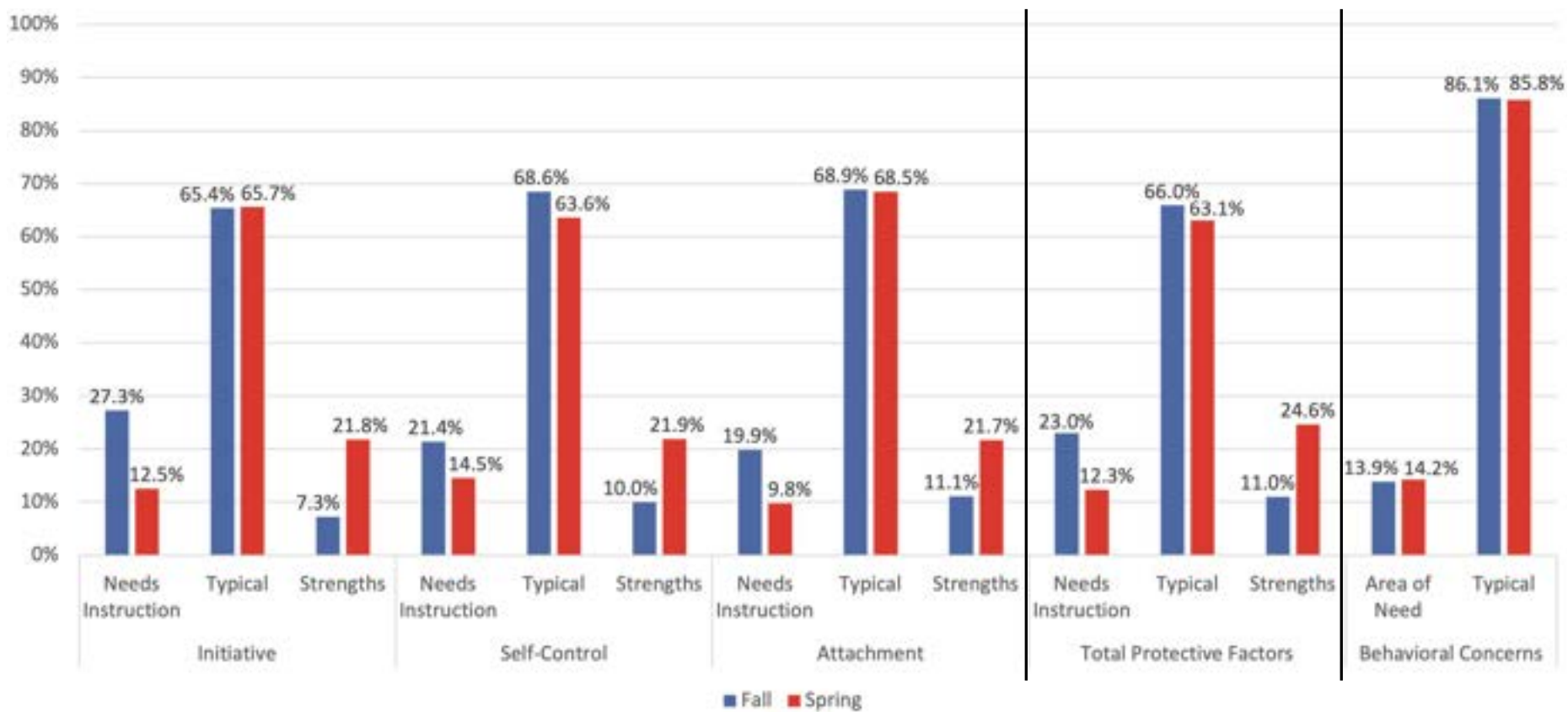
To measure children's social-emotional competencies, Westat analyzed data collected from teacher ratings on the Devereux Early Childhood Assessment (DECA). The DECA is designed based on a strengths-based approach to assessment and as such focuses on building children's social-emotional strengths. It also emphasizes the importance of promoting children's social-emotional competency, because that contributes to building their resilience to overcome adversity. The outcomes related to social-emotional competency that were measured include: Initiative, Self-Control, Attachment, and Behavioral Concerns. Taken together, the outcomes of Initiative, Self-Control, Attachment, combined to form an overall level of social-emotional competencies of Total Protective Factors.

Children (80.5%; n = 1,651) were included in analyses if they had outcome data for both time points. Across all outcomes and timepoints, most children were in the *Typical* level. Results revealed significant positive movement between levels for all outcomes except Behavioral Concerns (see Figure 4). The results showed an increasing

proportion of students moving into the highest level, *Strengths*, for Initiative by 15 percent, Self-Control by 12 percent, Attachment by 11 percent, and Total Protective Factors by 14 percent (see Figure 7). In the normative sample, approximately 16 percent of children demonstrated the highest level, *Strengths* (LeBuffe, Shapiro, & Naglieri, 2022). At the same time, the results showed a declining proportion of students testing at the lowest level, *Needs Instruction*, for Initiative by 15 percent, Self-Control by 7 percent, Attachment by 10 percent, and Total Protective Factors by 11 percent. There was no change for Behavioral Concerns. These findings indicate significant positive change between levels of understanding across all outcomes, except Behavioral Concerns (for more detailed information, see Appendix E Table E-1).

Results revealed that there was significant growth across all outcomes. On average, children grew 5.91 points in Initiative, 3.72 points in Self-Control, 4.32 points in Attachment, and 5.35 points for Total Protective Factors. On average, children decreased 0.87 points in Behavioral Concerns. For this outcome, there was a significant decline between fall and spring, which indicates a reduction in problematic behavior (for more detailed information, see Appendix E Table E-2).

Figure 7. | Proportion of children within each level by social-emotional outcome and time point



Note: Due to rounding, decimals may not agree to the nearest tenths. The Behavioral Concerns outcome has two levels by design. It is intended to measure problematic behaviors, which is reversed from the other outcomes measuring positive behaviors. Therefore, the category *Area of Need* comes from high scores and all other scores fall in the *Typical* range.

Conclusions and Looking Ahead

Overall, results from the Year 10 evaluation indicate Pre-K 4 SA has provided high-quality instructional environments to more than 2,000 predominantly low-income children from across San Antonio. The characteristics of children served were similar to those from previous years. Children's attendance in the program increased compared to last year. However, despite that increase overall attendance still remained lower when compared to years prior to the pandemic. This is not a surprising finding since national trends show similar decreases in school attendance (National Center for Education Statistics, 2022). Classroom quality scores were high (or midrange in the case of Instructional Support), indicating strong teacher-child interaction quality. Moreover, average instructional quality (Instructional Support) increased by one point from last year.

Teacher-reported kindergarten readiness at the end of the pre-K year (GOLD) suggested there was significant growth over time for all six outcomes: cognitive, literacy, mathematics, oral language, physical, and social-emotional.

Early literacy and numeracy results suggested children gained significant understanding across the year. However, most children were not performing at their age level of understanding and in need of additional educational supports. Receptive vocabulary results demonstrated children experienced significant growth and gained additional vocabulary. Moreover, a little over one-quarter (27.7%) of children narrowed their achievement gap when comparing their learning to a normative group of children. Teacher-reported social-emotional results suggested children significantly gain skills and demonstrated significant growth in their understanding over the year. As this is the first year conducting receptive vocabulary and social-emotional analyses, it will be interesting to explore what patterns and trends emerge in future evaluations and whether current findings are replicated, or new ones emerge.

Taken together, the results from the Year 10 evaluation suggest children are benefiting from participation at Pre-K 4 SA centers but significant learning supports may be needed in early literacy, numeracy, and vocabulary as is being seen across the nation. Moreover, this marks the first year Pre-K 4 SA has served 3-year-old children. It will be interesting to explore relationships for children who attend for 2 years compared to a single year to determine what patterns and trends emerge based on increased participation.

Limitations and Recommendations

Related to these findings is one limitation. 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 all 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 and did not experience learning in the context of the pandemic. Normed samples are created to be reflective of the demographic proportions similar to those found in the U.S. Census and were constructed prior to the pandemic during a 'typical' school year. There can be more confidence in interpreting resulting differences on outcomes when a comparison or control group is formed with children who are most like Pre-K 4 SA children and experienced learning during the pandemic. Furthermore, there can be more confidence that differences can be attributed to Pre-K 4 SA and are not a result of other factors.¹⁷



¹⁷ One way to form such a group of children, similar in nature to Pre-K 4 SA children, would be to work with an organization (e.g., Teaching Strategies) to create a matched comparison group from the normed sample of children in the future.

References

- Barnett, W.S. (2011). Four reasons the United States should offer every child a preschool education. In E. Zigler, W.S. Gilliam, and W.S. Barnett (Eds.), *The pre-K debates: Current controversies and issues* (pp. 34–39). Baltimore, MD: Brookes Publishing.
- Bassok, D., and Galdo, E. (2016). Inequality in preschool quality? Community-level disparities in access to high-quality learning environments. *Early Education and Development*, 27(1), 128–144.
- Benjamini, Y., and 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.
- Bloom, H.S., and Weiland, C. (2014). To what extent do the effects of Head Start on enrolled children vary across sites? (working paper). New York, NY: MDRC.
- 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., and Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science*, 6(1), 42–57.
- Carminucci, J., Hodgman, S., Rickles, J., and Garet, M. (2021). Student attendance and enrollment loss in 2020-21. AIR. Available at: https://www.air.org/sites/default/files/2021-07/research-brief-covid-survey-student-attendance-june-2021_0.pdf
- Dunn, L., and Dunn, L. (2019). *Peabody Picture Vocabulary Test, fifth edition* (PPVT -V). Bloomington, MN: NCS Pearson.
- Early Childhood Learning & Knowledge Center (ECLKC). (2020). *A national overview of grantee CLASS scores in 2020*. Washington, DC: U.S. Department of Health & Human Services, ECLKC. Available at: <https://eclkc.ohs.acf.hhs.gov/data-ongoing-monitoring/article/national-overview-grantee-class-scores-2020>
- Heckman, J. J., Moon, S.H., Pinto, R., Savelyev, P.A., and Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, 94(1-2), 114–128.
- Hill, C.J., Gormley, Jr., W.T., and Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Childhood Research Quarterly*, 32, 60–79.
- Lambert, R. (2020). Technical manual for the Teaching Strategies GOLD assessment system (2nd edition). Charlotte, NC: Center for Educational Measurement and Evaluation, University of North Carolina. Available at: https://teachingstrategies.com/wp-content/uploads/2020/10/2020-Tech-Manual_GOLD.pdf
- La Paro, K.M., Pianta, R.C., and Shuhlman, M. (2004). Classroom Assessment Scoring System: Findings from the prekindergarten year. *Elementary School Journal*, 104(5), 409–426.
- LeBuffe, P.A., and Naglieri, J.A. (2012). *Devereux Early Childhood Assessment for preschoolers second edition* (DECA-P2): *User's guide and technical manual*. Lewisville, NC: Kaplan Press.
- LeBuffe, P.A., Shapiro, V.B., and Naglieri, J.A. (2022). *Devereux student strengths assessment*. Lewisville, NC: Kaplan Press.
- LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., and 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.

- Mashburn, A.J., Pianta, R.C., Hamre, B.K., Downer, J.T., Barbarin, O.A., Bryant, D., Burchinal, M., Early, D.M., and 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.
- McCormick, M. (2022). Building better evidence on pre-K by strengthening assessments of children's skills. Washington, DC: MDRC. Available at: <https://www.mdrc.org/publication/building-better-evidence-pre-k-strengthening-assessments-children-s-skills>.
- Muñoz-Sandoval, A.F., Woodcock, R.W., McGrew, K.S., and Mather, N. (2005). *The Batería III Woodcock-Muñoz: Pruebas de aprovechamiento*. Itasca, IL: Riverside Publishing.
- National Center for Education Statistics (2022). *More than 80 percent of U.S. public schools report pandemic has negatively impacted student behavior and socio-emotional development*. Washington, DC: U.S. Department of Education. Available at [Press Release - More than 80 Percent of U.S. Public Schools Report Pandemic Has Negatively Impacted Student Behavior and Socio-Emotional Development - July 6, 2022](#)
- Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the classroom assessment scoring system) in early childhood education and care settings and child outcomes. *PloS one*, 11(12), e0167660.
- Pianta, R., La Paro, K., and Hamre, B. (2008). *Classroom assessment scoring system*. Baltimore, MD: Brookes Publishing.
- Puma, M., Bell, S., Cook, R., and Heid, C. (2010). *Head Start Impact Study final report*. Washington, DC: Administration for Children & Families, U.S. Department of Health and Human Services.
- Reynolds, A.J., Temple, J.A., White, B.A, Ou, S-R., and Robertson, D.L. (2011). Age-26 cost benefit analysis of the Child-Parent Center Early Education Program. *Child Development*, 82(1), 379–404.
- Rolnick, A., and Grunewald, R. (2003). Early childhood development: Economic development with a high public return. *The Region*, 17(4), 6–12.
- Schrank, F.A., Mather, N., and McGrew, K.S. (2014). *Woodcock-Johnson IV tests of achievement*. Rolling Meadows, IL: Riverside Publishing.
- Schrank, F.A., McGrew, K.S., and LaForte, E.M. (2014). *Technical manual Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.
- Valentino, R. (2018). Will public pre-K really close achievement gaps? Gaps in prekindergarten quality between students and across states. *American Educational Research Journal*, 55(1), 79–116.
- Weiland, C. (2016). Impacts of the Boston prekindergarten program on the school readiness of young children with special needs. *Developmental Psychology*, 52(11), 1763–1776.

Appendix A

Evaluation Methods

Here we provide information on measures used in the Pre-K 4 SA Year 10 evaluation, as well as details on the analytic approach to the analyses described in the body of the report.

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 10 evaluation consisted of five 20-minute cycles,¹⁸ followed by 10-minute coding periods. Scores were assigned during various classroom activities and then averaged across all cycles for an overall quality score.

Observations occurred during the spring of the 2022-23 school year. Interactions were measured on 10 different dimensions (see Table A-1 for descriptions of each CLASS dimension) divided into three larger domains. The Emotional Support domain is measured using four dimensions: positive climate, negative climate, teacher sensitivity, and regard for student perspectives. The CLASS also measures Classroom Organization using three dimensions: behavior management, productivity, and instructional learning formats. Instructional Support is measured using 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, a score of 3, 4, or 5 indicates midrange 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, as well as 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.

¹⁸ Across all observation cycles, there were a total of 25 cycles in which the time was greater than 20 minutes ranging from 21 to 31 minutes.

Teaching Strategies GOLD

The GOLD assessment is a teacher-reported measure selected and used by Pre-K 4 SA to collect information on children's progress on 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 points throughout the year: fall, winter, and spring. The new updated norms for birth through third grade were used as comparison to Pre-K 4 SA children.

The Woodcock-Johnson Achievement Test

The Woodcock-Johnson Test of Achievement-IV (WJ) (Schrank, McGrew, Mather, 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* subtest and the *Applied Problem* subtest. Both subtests demonstrate excellent reliability (.92 for Applied Problems and .97 for Letter-Word). Correlations of the WJ with other tests of cognitive ability and achievement are reported to range from .83 to .86 (McGrew, LaForte, & Schrank, 2014). This measure has been used in numerous large-scale preschool studies (e.g., Early et al., 2007; Wong, Cook, Barnett, & Jung, 2008). 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. The *Applied Problems* subtest is a test of basic analytic skills involving applying simple number concepts and solving math problems.

The child listens to the problem, recognizes the underlying mathematical procedure and steps to solve the problem, and performs the appropriate calculations. In both subtests, items are set in difficulty order, with the easiest first and the most difficult last. Testing stops when the child scores zero on six successive items across both subtests in English and Spanish with one exception. Testing for *Applied Problems* in English stops when the child scores zero on five successive items (McGrew, LaForte, & Schrank, 2014). To help understand the range of children's age equivalence in early literacy and numeracy, descriptive information for both subtests based on the language of assessment is provided in Table A-2. As displayed in Table A-2, there was some variation in children's assessed level of understanding based on language of assessment. For example, children assessed in English had a higher maximum (7 years, 4 months) than children assessed in Spanish (6 years, 2 months) for *Letter-Word*.

Peabody Picture Vocabulary Test

The Peabody Picture Vocabulary Test Fifth Edition (PPVT – V) (Dunn & Dunn, 2019) is a test of receptive vocabulary in standard English. The PPVT has established overall (.97), and test-retest reliability (.88) and concurrent and predictive validity based on moderate effect sizes ranging from .46 to .77 with other language and achievement measures. These measures have been demonstrated to be valid and reliable for ages ranging from 2.5 years to over 90 (Dunn & Dunn, 2019; Pearson Education Inc., 2019). For the test administration, the child is presented pictorial images of words and selects the correct one. The items are set in difficulty order, with the easiest first and the most difficult last. Testing stops when the child scores zero on six successive items.

Table A-2 | Range of age equivalencies by assessment and language

Outcome	English		Spanish	
	Minimum	Maximum	Minimum	Maximum
Letter-Word	2 years, 4 months	7 years, 4 months	2 years, 5 months	6 years, 2 months
Applied Problems	3 years, 2 months	10 years, 7 months	4 years, 5 months	10 years, 11 months

Devereux Early Childhood Assessment

The DECA (LeBuffe & Naglieri, 1999; LeBuffe, Shapiro, & Naglieri, 2022) measures children's social-emotional competencies through parent and teacher reports. Parents and teachers report on the frequency of children's behavior on items comprising two main scales and three subscales using a 5-point Likert-type scale to indicate how often within the past 4 weeks a child exhibited behaviors described by assessment items (0 = never, 1 = rarely, 2 = occasionally, 3 = frequently, and 4 = very frequently). In the current evaluation, results are based on teacher ratings. In a standardization sample, the DECA was reported to have good reliability evidence (LeBuffe & Naglieri, 1999; Center for Resilient Children, 2013). As displayed in Table A-3, internal consistency and test-retest reliability are above 0.80, which is the suggested standard (Crocker & Algina, 2006; Nunnally & Bernstein, 1994). For inter-rater reliability, the parent values are moderate, and the teacher ratings are substantial (McHugh, 2012).

In addition, the test-retest reliability coefficients for subscales ranged from .86 to .94 (LeBuffe et al., 2009). The agreement among parents and teachers has been shown to be moderate ($r_s = .20-.28$), which is consistent with prior research. The agreement is higher when both parents and teachers complete the assessment in the same language (Crane, Mincic & Winsler, 2011). Therefore, whenever possible parents and teachers should complete the assessment in the same language (LeBuffe & Likins, n.d.).

Analytic Approach

Research questions were addressed by analyzing study data as well as existing Pre K 4 SA databases. To address the first two questions—1A: *What were the reported levels of child attendance*

during the pre-K year?; and 1B: *In what ways have attendance rates changed since the COVID-19 pandemic?*; Pre-K 4 SA data were submitted to Westat and descriptively analyzed. To address research question 2A: *What was the observed teacher-child interaction quality in Pre-K 4 SA classrooms in Year 10?*; CLASS observation data were descriptively analyzed. To answer research question 2B: *Did master teachers of Pre-K 4 SA classrooms have higher observed teacher-child interaction quality in Year 10?*; an independent samples *t*-test between Master and Non-Master teachers in Pre-K 4 SA classrooms was conducted for each domain and dimension. To answer research question 2C: *Were there any significant differences across the education centers?*; a one-way analysis of variance was conducted for each domain and dimension to test for significant differences across the four center locations. The Benjamini-Hochberg (1995) technique was applied to correct for multiple hypothesis testing.

To address the third set of questions, 3A: *How did Pre-K 4 SA children compare to the normed sample on GOLD outcomes over the year?*; and, 3B: *Did Pre-K 4 SA children demonstrate significant growth on GOLD outcomes over the year?*; data collected by Pre-K 4 SA were submitted to Westat. Data across the three timepoints—fall, winter, and spring—were combined for inferential analyses. To answer question 3A an independent samples *t*-test between Pre-K 4 SA and the normative sample was conducted in fall, winter, and spring for each outcome to determine if there were significant differences. To answer question 3B, a dependent sample *t*-test was conducted between the fall and spring for each GOLD assessment outcome to determine if there was significant growth.

Table A-3 | Reliability evidence for the DECA by parent and teacher ratings

Reliability Type	Total Protective Factors		Behavior Concerns	
	Parent	Teacher	Parent	Teacher
Internal consistency	.92	.95	.80	.86
Test-retest	.88	.95	.78	.88
Inter-rater	.51	.72	.46	.70

As children were not randomly sampled, demographic tests of differences were conducted to determine if the sample of children included and excluded from analyses were similar. For 3-year-old children no significant differences were found between children included and not included in analyses for gender ($\chi^2(1) = 2.9664, p = .085$), race and ethnicity ($\chi^2(5) = 2.4730, p = .781$), eligibility based on low-income ($\chi^2(1) = .7782, p = .378$), eligibility based on middle-income ($\chi^2(1) = .2205, p = .639$), language ($\chi^2(1) = .2547, p = .614$), military connection status ($\chi^2(1) = .4827, p = .487$), dual enrollment status ($\chi^2(1) = .5578, p = .455$), and tuition status ($\chi^2(1) = .3739, p = .541$).

For 4-year-old children no significant differences were found between children included and not included in analyses for gender ($\chi^2(1) = .2027, p = .653$), race and ethnicity ($\chi^2(5) = 7.4011, p = .192$), eligibility based on middle-income ($\chi^2(1) = 2.6736, p = .102$), language ($\chi^2(1) = .2347, p = .628$), military connection status ($\chi^2(1) = .9080, p = .341$), and tuition status ($\chi^2(1) = 1.3210, p = .250$); however, differences were found for eligibility based on low-income ($\chi^2(1) = 4.3223, p = .038$) and dual enrollment status ($\chi^2(1) = 3.9344, p = .047$). Children included in analyses were less likely to have attended Pre-K 4 SA based on the low-income eligibility requirement than children excluded from analyses. Children included in analyses were less likely to be dually enrolled in a partner district than children excluded from analyses.

To address the fourth set of questions, 4A: *What proportion of a random sample of Pre-K 4 SA children performed at or above their age level in early literacy and early numeracy over the year?*; 4B: *Did a random sample of Pre-K 4 SA children demonstrate significant growth in early literacy and early numeracy over the year?*; 4C: *Did a random sample of Pre-K 4 SA children experience accelerated learning to help narrow achievement gaps in early literacy and early numeracy?*; data collected by Pre-K 4 SA were submitted to Westat and descriptively and inferentially analyzed. Children had to have data in both time points to be included in an outcome analysis of any research question in the fourth set. Raw scores were converted into age

levels (measured in years and months) based on norms provided in the technical manual for each outcome and language of assessment (English and Spanish)¹⁹ (McGrew, LaForte, & Schrank, 2014). For the age level, the number of months was translated into the proportion of the year (e.g., 4 years and 6 months = 4.5) to compare to their actual age in years and months. Another method of analysis is to convert raw scores into grade-level equivalencies. For pre-K children, there is a single grade level available: below kindergarten. As no finer levels (e.g., pre-K and 2 months) were available, and it would not be possible to determine growth over time, it was determined to use only age equivalencies. A binary indicator was created for each outcome to determine if a child's assessed age level was either below or on/above their actual age level. To answer the first part of research question 4A, descriptive analyses were conducted for each outcome. To answer the second part, the proportions below, on, and above were computed and analyzed for each outcome. Analyses were conducted based on the binary indicators: a McNemar test was conducted between fall 2022 and spring 2023 for each outcome to determine if there was a significant increase in the proportions over time. For 4B, dependent *t*-tests were conducted for each outcome between fall 2022 and spring 2023 to determine if there were significant increases over time. For 4C, the difference between each child's assessed age and biological age was computed for each outcome in fall 2022 and spring 2023. This difference was used to measure the achievement gap for each child denoting how many months they were above, below, or on par in their understanding of early literacy and numeracy. A difference in differences analysis was conducted to determine if there was a significant reduction of the achievement gap from fall to spring.

To address the fifth research question, 5A: *What were the performance levels of a random sample of Pre-K 4 SA children over the year?*; 5B: *Did a random sample of Pre-K 4 SA children demonstrate significant growth in vocabulary over the year?*; and 5C: *What types of vocabulary growth did a random sample of Pre-K 4 SA children demonstrate over the year?*; data collected by Pre-K 4 SA were submitted to Westat

¹⁹The age norms differ depending on whether a child was assessed in English or Spanish.

and descriptively and inferentially analyzed. Children had to have data in both time points to be included in the analysis. Raw scores were converted into standard scores and growth scale values based on the norms provided in the technical manual (Dunn & Dunn, 2019). Standard scores were converted into performance descriptions based on the technical manual. To address 5A, descriptive analyses were conducted for the performance description for each time point. To address 5B, dependent *t*-tests were conducted separately for the standard scores and growth scale values between fall 2022 and spring 2023 to determine if there were significant increases over time. To address 5C and aid the interpretation of changes in standard scores and growth scale values over time, score patterns were descriptively analyzed according to the patterns detailed in the technical manual.²⁰

Finally, to address the sixth research question, 6A: *What were the levels of Pre-K 4 SA children's social-emotional understanding over the year? To what extent did the levels change over the year?*; 6B: *Did Pre-K 4 SA children demonstrate significant growth in social-emotional learning over the year?*; data collected by Pre-K 4 SA were submitted to Westat and descriptively and inferentially analyzed.

As children were not randomly sampled, demographic tests of differences were conducted to determine if the sample of children included and excluded from analyses were similar. No significant differences were found between children included and not included in analyses for

gender ($\chi^2(1) = 2.023, p = .155$), low-income eligibility criteria ($\chi^2(1) = 1.624, p = .202$), English language learner status ($\chi^2(1) = 0.861, p = .353$), middle-income eligibility criteria ($\chi^2(1) = 0.921, p = .337$), military affiliation eligibility criteria ($\chi^2(1) = 0.183, p = .669$), affiliated with a partner district ($\chi^2(1) = 2.183, p = .140$), attending based on tuition status ($\chi^2(1) = 0.488, p = .485$) and indicators of race and ethnicity (African American: $\chi^2(1) = 2.077, p = .150$; Hispanic: $\chi^2(1) = 0.214, p = .643$; White: $\chi^2(1) = 1.213, p = .271$). This implies the findings were representative of all Pre-K 4 SA children. Children had to have data in both time points²¹ to be included in an outcome analysis of any research question in the fifth set. T-scores were converted into three categorical levels—*needs instruction, typical, and strengths*—based on the technical manual (LeBuffe & Naglieri, 2012). To address 6A, descriptive analyses were conducted for the categorical levels for each time point. A Wilcoxon signed-rank test was conducted between fall 2022 and spring 2023 for each outcome except Behavioral Concerns to determine if there were significant differences in the distribution of categories over time. As Behavioral Concerns is binary, a McNemar test was conducted between fall 2022 and spring 2023. To address 6B, dependent *t*-tests were conducted for the T-scores of each outcome between fall 2022 and spring 2023 to determine if there were significant increases over time for all but one outcome. The one exception, Behavioral Concerns, is the outcome in which a decrease or reduction in problematic behaviors is the desired result.

²⁰ For the category "Standard score does not change, growth scale value increases very little" a value of six was used to quantify very little increase for the growth scale value increase as no exact numerical value was provided in the technical manual.

²¹ Eight (0.5%) children had multiple assessments in fall 2022 or spring 2023 and were excluded from analyses. Their data was inconsistent within the time point and would have led to inconsistent findings.

Appendix B

Additional CLASS Results

The purpose of Appendix B is to provide additional classroom quality results answering research question two. There are two tables provided. Table B-1 addresses research question 2A: *What was the observed teacher-child interaction quality of Pre-K 4 SA classrooms in Year 10?* by providing descriptives for each center. Table B-2 addresses research question 2B: *Did master teachers of Pre-K 4 SA classrooms have higher observed teacher-child interaction quality in Year 10?*

Table B-1 | Average Year 10 CLASS scores by center

CLASS Domain/Dimension	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	6.25 (0.55)	(5.10 - 7.00)	6.63 (0.35)	(6.00 - 7.00)	6.44 (0.52)	(5.30 - 7.00)	6.43 (0.42)	(5.70 - 7.00)
Positive climate	6.11 (0.80)	(4.60 - 7.00)	6.64 (0.43)	(6.00 - 7.00)	6.43 (0.73)	(5.00 - 7.00)	6.36 (0.58)	(5.20 - 7.00)
Negative climate ^a	6.85 (0.25)	(6.20 - 7.00)	6.97 (0.07)	(6.80 - 7.00)	6.98 (0.06)	(6.80 - 7.00)	6.98 (0.08)	(6.60 - 7.00)
Teacher sensitivity	5.99 (0.77)	(4.20 - 7.00)	6.38 (0.64)	(5.20 - 7.00)	6.22 (0.66)	(4.80 - 7.00)	6.10 (0.62)	(5.00 - 7.00)
Regard for student perspectives	6.07 (0.59)	(5.00 - 7.00)	6.53 (0.43)	(5.80 - 7.00)	6.13 (0.77)	(4.20 - 7.00)	6.30 (0.61)	(5.00 - 7.00)
Classroom Organization	5.77 (0.89)	(3.73 - 6.93)	6.29 (0.57)	(5.20 - 7.00)	6.04 (0.71)	(4.60 - 7.00)	6.05 (0.60)	(4.67 - 6.87)
Behavior management	5.62 (0.88)	(4.00 - 7.00)	6.14 (0.80)	(4.60 - 7.00)	5.91 (0.82)	(4.40 - 7.00)	5.90 (0.80)	(3.60 - 7.00)
Productivity	6.05 (0.94)	(3.80 - 7.00)	6.50 (0.56)	(5.20 - 7.00)	6.24 (0.61)	(4.60 - 7.00)	6.30 (0.59)	(5.00 - 7.00)
Instructional learning formats	5.64 (0.98)	(3.40 - 7.00)	6.25 (0.50)	(5.00 - 7.00)	5.96 (0.84)	(4.20 - 7.00)	5.94 (0.64)	(4.60 - 6.80)
Instructional Support	3.66 (1.34)	(1.67 - 6.00)	4.27 (1.34)	(1.73 - 6.60)	4.25 (1.34)	(1.60 - 6.40)	4.76 (0.95)	(2.80 - 6.73)
Concept development	3.47 (1.38)	(1.60 - 6.00)	4.04 (1.29)	(1.80 - 6.40)	4.07 (1.34)	(1.80 - 6.00)	4.55 (1.04)	(2.60 - 6.60)
Quality of feedback	3.67 (1.51)	(1.20 - 6.20)	4.45 (1.48)	(1.60 - 6.80)	4.20 (1.59)	(1.20 - 6.40)	4.96 (1.07)	(3.00 - 7.00)
Language modeling	3.83 (1.23)	(2.00 - 6.60)	4.33 (1.37)	(1.80 - 7.00)	4.49 (1.31)	(1.80 - 6.80)	4.77 (0.92)	(2.80 - 6.80)

Note: M = mean; SD = standard deviation. Domains in bold font and Dimensions in plain font.

^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.

Table B-2 | Classroom quality results comparing Master and Non-Master teachers

CLASS Domain/ Dimension	Sample Size	Non-Master Teacher Mean	Master Teacher mean	Difference (Master-Non Master)	t-test statistic	df	Initial p-value	Adjusted Significance	Group favored^a
Emotional Support	90	6.35	6.55	0.20	-2.08	87.94	0.04	Non-significant	No difference
Positive Climate	90	6.28	6.52	0.24	-1.79	87.99	0.08	Non-significant	No difference
Negative Climate	90	6.93	6.97	0.05	-1.73	73.86	0.09	Non-significant	No difference
Teacher Sensitivity	90	6.03	6.35	0.31	-2.24	84.57	0.03	Non-significant	No difference
Regard for Student Perspectives	90	6.17	6.37	0.20	-1.54	87.97	0.13	Non-significant	No difference
Classroom Organization	90	5.88	6.25	0.37	-2.55	87.63	0.01	Significant	Master Teacher
Behavior Management	90	5.74	6.10	0.36	-2.05	79.29	0.04	Significant	Master Teacher
Productivity	90	6.13	6.46	0.33	-2.38	87.13	0.02	Significant	Master Teacher
Instructional Learning Formats	90	5.77	6.18	0.41	-2.68	87.74	0.01	Significant	Master Teacher
Instructional Support	90	4.21	4.28	0.06	-0.23	78.54	0.82	Non-significant	No difference
Concept Development	90	3.99	4.10	0.11	-0.38	79.51	0.70	Non-significant	No difference
Quality of Feedback	90	4.30	4.36	0.06	-0.18	77.24	0.85	Non-significant	No difference
Language Modeling	90	4.35	4.37	0.03	-0.10	77.90	0.92	Non-significant	No difference

Note: df = degrees of freedom. Domains in bold font and Dimensions in plain font. Due to rounding, decimals may not agree to the nearest hundredths. The Adjusted Significance column indicates significance levels (p-values) after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg technique (1995).

^a If a statistically 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.”

Appendix C

Additional Woodcock-Johnson and Bateria Results

The purpose of Appendix C is to provide additional direct child assessment—Woodcock-Johnson and Bateria—results answering research question 4. There are three tables provided. Table C-1 addresses both parts of research question 4A: *What proportion of Pre-K 4 SA children performed at or above their age level in early literacy and early numeracy over the year?* To what extent did the proportion change over the year? Table C-2 addresses research question 4B: *Did a random sample of Pre-K 4 SA children demonstrate significant growth in early literacy and early numeracy over the year?*. Table C-3 addresses research question 4C: *Did a random sample of Pre-K 4 SA children experience accelerated learning to help narrow achievement gaps in early literacy and early numeracy?*.

Table C-1 | Proportion analysis results meeting age equivalency comparing fall 2022 and spring 2023

Outcome	Sample Size	Fall 2022 proportion	Spring 2023 proportion	Difference (Spring 2023 – Fall 2022)	χ^2 statistic	df	Initial p-value	Adjusted Significance	Time favored ^a
Letter-Word	60	0.42	0.33	-0.08	2.273	1	0.132	Not Significant	No Change
Applied Problems	60	0.15	0.15	0.00	0.000	1	1.000	Not Significant	No Change

Note: due to rounding, decimals may not agree to the nearest hundredths. df = degrees of freedom. Letter-Word measures early literacy skills and Applied Problems measures early numeracy skills.

^a If a statistically significant difference was found, the time whose score was greater (the “favored” time) is listed in this column. If there was no statistically significant difference, this column states that there was “no difference.”

Table C-2 | Age equivalencies results comparing fall 2022 and spring 2023 by subtest

Outcome	Sample Size	Fall 2022 mean	Spring 2023 mean	Growth (Spring 2023 – Fall 2022)	t-test statistic	df	Initial p-value	Adjusted Significance	Time favored ^a
Letter-Word	60	4 years, 2 months	4 years, 6 months	4 months	4.89	59	<.0001	Significant	Spring 2022
Applied Problems	60	3 years, 5 months	4 years, 0 months	6 months ^b	6.92	59	<.0001	Significant	Spring 2022

Note: df = degrees of freedom. Letter-Word measures early literacy skills and Applied Problems measures early numeracy skills.

^a If a statistically significant difference was found, the time whose score was greater (the “favored” time) is listed in this column. If there was no statistically significant difference, this column states that there was “no difference.”

^b Due to rounding, the gap is not the exact difference between fall and spring.

Table C-3 | Gap analysis age equivalency results comparing fall 2022 and spring 2023 by subtest

Outcome	Sample Size	Average gap between age equivalency and actual age Fall 2022	Average gap between age equivalency and actual age Spring 2023	Gap (Spring 2023 – Fall 2022)	t-test statistic	df	Initial p-value	Adjusted Significance	Gap reduced ^a
Letter-Word	60	-1 month	-3 months	-1 month ^b	-1.97	59	0.053	Not significant	No difference
Applied Problems	60	-10 months	-9 months	Less than a month	0.97	59	0.334	Not significant	No difference

Note: df = degrees of freedom. Letter-Word measures early literacy skills and Applied Problems measures early numeracy skills.
^aIf a statistically significant difference was found, this indicates if the gap was reduced and by how many months. If there was no statistically significant difference, this column states that there was “no difference.”
^b Due to rounding, the gap is not the exact difference between fall and spring.



Appendix D

Additional Peabody Picture Vocabulary Results

The purpose of Appendix D is to provide additional direct child assessment—Peabody Picture Vocabulary Test—results answering research question 5. There are three tables provided. Table D-1 addresses research question 5A: *What were the performance levels of a random sample of Pre-K 4 SA children over the year?* Table D-2 addresses research question 5B: *Did a random sample of Pre-K 4 SA children demonstrate significant growth in vocabulary over the year?* Table D-3 addresses research question 5C: *What types of vocabulary growth did a random sample of Pre-K 4 SA children demonstrate over the year?*

Table D-1 | Descriptives of vocabulary performance levels by timepoint

Level	Sample Size	Fall 2022 proportion	Spring 2023 proportion	Difference (Spring 2023 – Fall 2022)
Well below expected	47	0.00	0.00	0.00
Below expected		0.30	0.28	-0.02
Expected		0.68	0.66	-0.02
Above expected		0.00	0.06	0.06
Well above expected		0.02	0.00	-0.02

Note: Due to rounding, decimals may not agree to the nearest hundredths.

Table D-2 | Vocabulary growth results comparing fall 2022 and spring 2023 by score type

Score type	Sample Size	Fall 2022 mean	Spring 2023 mean	Growth (Spring 2023 – Fall 2022)	t-test statistic	df	Initial p-value	Adjusted Significance	Time favored ^a
Standard Score	47	92.30	92.49	0.19	0.15	46	0.885	Not significant	No Difference
Growth Scale Value	47	461.34	466.68	5.34	6.09	46	<.0001	Significant	Spring 2023

Note: df = degrees of freedom. Due to rounding, decimals may not agree to the nearest hundredths.

^aIf a statistically significant difference was found, the time whose score was greater (the “favored” time) is listed in this column. If there was no statistically significant difference, this column states that there was “no difference.”

Table D-3 | Descriptives of vocabulary score change comparing fall 2022 and spring 2023

Type of Change	N	Percentage
Children did not demonstrate their best performance or are losing previously mastered vocabulary and are not acquiring skills as quickly as other children who are the same age. Therefore, the gap between these children's performance and the performance of children following a typical trajectory has widened.	8	17.02
Children are learning new vocabulary or improving their understanding of current vocabulary. However, they are not acquiring skills as quickly as other children who are the same age. Therefore, the gap between these children's performance and the performance of children following a typical trajectory has widened.	11	23.40
Children are learning new vocabulary or improving their understanding of current vocabulary very slowly and at the same time maintained the same rating as children of their same age.	4	8.51
Children are learning new vocabulary or improving their understanding of current vocabulary.	13	27.66
Children are learning new vocabulary or improving their understanding of current vocabulary and the gap between the individual's performance and the performance of typically developing children of the same age has narrowed.	11	23.40
Total	47	100

Note: N = sample size. Due to rounding, decimals may not agree to the nearest hundredths.



Appendix E

Additional Devereux Early Childhood Assessment Results

The purpose of Appendix E is to provide additional social-emotional assessment—results answering research question 6. There are three tables provided. Table E-1 addresses both parts of research question 6A: *What were the levels of Pre-K 4 SA children’s social-emotional understanding over the year? To what extent did the levels change over the year?* Table E-2 addresses research question 6B: *Did Pre-K 4 SA children demonstrate significant growth in social-emotional learning over the year?*

Table E-1 | Social-emotional results comparing levels in fall 2022 and spring 2023 by outcome

Outcome	Level	Sample Size	Fall 2022 proportion	Spring 2023 proportions	Difference (Spring 2023-Fall2022)	S statistic	Initial p-value	Adjusted Significance	Time favored ^a
Initiative	Needs Instruction	1651	0.27	0.13	-0.15	70,681	<.0001	Significant	Spring 2023
	Typical		0.65	0.66	0.00				
	Strengths		0.07	0.22	0.15				
Self-Control	Needs Instruction	1651	0.21	0.15	-0.07	37,753	<.0001	Significant	Spring 2023
	Typical		0.69	0.64	-0.05				
	Strengths		0.10	0.22	0.12				
Attachment	Needs Instruction	1651	0.20	0.10	-0.10	46,027	<.0001	Significant	Spring 2023
	Typical		0.69	0.69	0.00				
	Strengths		0.11	0.22	0.11				
Total Protective Factors	Needs Instruction	1651	0.23	0.12	-0.11	55,844	<.0001	Significant	Spring 2023
	Typical		0.66	0.63	-0.03				
	Strengths		0.11	0.25	0.14				
Behavioral Concerns	Typical	1651	0.86	0.86	0.00	0.2022 ^b	0.6529	Not significant	No Difference
	Area of Need		0.14	0.14	0.00				

Note: due to rounding, decimals may not agree to the nearest hundredths. df = degrees of freedom.

^a If a statistically significant difference was found, the time whose score was greater (the “favored” time) is listed in this column. If there was no statistically significant difference, this column states that there was “no difference.”

^b McNemar’s test was conducted for Behavioral Concerns because it had two possible classifications. The test statistic is a Chi-square with 1 degree of freedom.

Table E-2 | Social-emotional equivalencies results comparing fall 2022 and spring 2023 by outcome

Outcome	Sample Size	Fall 2022 mean	Spring 2023 mean	Growth (Spring 2023 – Fall 2022)	t-test statistic	df	Initial p-value	Adjusted Significance	Time favored ^a
Initiative	1651	47.04	52.95	5.91	29.86	1650	<.0001	Significant	Spring 2023
Self-Control	1651	47.90	51.62	3.72	19.79	1650	<.0001	Significant	Spring 2023
Attachment	1651	48.46	52.78	4.32	19.63	1650	<.0001	Significant	Spring 2023
Total Protective Factors	1651	47.48	52.83	5.35	27.15	1650	<.0001	Significant	Spring 2023
Behavioral Concerns	1651	48.68	47.80	-0.87	-4.66	1650	<.0001	Significant	Spring 2023 ^b

Note: df = degrees of freedom.

^a If a statically significant difference was found, the time whose score was greater (the “favored” time) is listed in this column. If there was no statistically significant difference, this column states that there was “no difference.”

^b For the Behavioral Concerns outcome, negative growth means that behavioral issues decreased in spring 2023, indicating a positive finding.



Appendix References

- Benjamini, Y., and 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.
- Center for Resilient Children. (2013). *Devereux Center for Resilient Children (DCRC) assessment tools: the Devereux Early Childhood Assessment for Preschoolers, second edition (DECA-P2) summary table*. Villanova, PA: Author. Available at: <https://centerforresilientchildren.org/wp-content/uploads/2014/01/DECA-P2-full-version.pdf>.
- Crane, J., Mincic, M.S., and Winsler, A. (2011). Parent–teacher agreement and reliability on the Devereux Early Childhood Assessment (DECA) in English and Spanish for ethnically diverse children living in poverty. *Early Education and Development*, 22(3), 520–547.
- Crocker, L., and Algina, J. (2006). *Introduction to classical and modern test theory*. Boston, MA: Cengage Learning.
- Dunn, L., and Dunn, L. (2019). *Peabody Picture Vocabulary Test, fifth edition (PPVT -V)*. Bloomington, MN: NCS Pearson.
- Early, D.M., Maxwell, K.L., Burchinal, M., Alva, S., Bender, R.H., Bryant, D., Cai, K., Clifford, R.M., Ebanks, C., Griffin, J.A., Henry, G.T., Howes, C., Irondo-Perez, J., Jeon, H., Mashburn, A.J., Peisner-Feinberg, E., Pianta, R.C., Vandergrift, N., and Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, 78(2), 558–580.
- LeBuffe, P.A., and Naglieri, J.A. (1999). *The Devereux early childhood assessment*. Lewisville, NC: Kaplan Press.
- LeBuffe, P.A., and Naglieri, J.A. (2012). *Devereux Early Childhood Assessment for preschoolers second edition (DECA-P2): User's guide and technical manual*. Lewisville, NC: Kaplan Press.
- LeBuffe, P.A., Shapiro, V.B., & Naglieri, J.A. (2009). *Devereux student strengths assessment*. Lewisville, NC: Kaplan Press.
- LeBuffe, P.A., Shapiro, V.B., and Naglieri, J.A. (2022). *Devereux student strengths assessment*. Lewisville, NC: Kaplan Press.
- McHugh, M.L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276–282.
- Muñoz-Sandoval, A.F., Woodcock, R.W., McGrew, K.S., and Mather, N. (2005). *The Batería III Woodcock-Muñoz: Pruebas de aprovechamiento*. Itasca, IL: Riverside Publishing.
- Nunnally, J. & Bernstein L. (1994). *Psychometric theory*. New York: McGraw-Hill Higher, Inc.
- Pearson Education, Inc. (2019). PPVT by the numbers. Bloomington, MN: Author. Available at: <https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/PPVT5-EVT3-By-the-Numbers-Infographic.pdf>
- Pianta, R., La Paro, K., and Hamre, B. (2008). *Classroom assessment scoring system*. Baltimore, MD: Brookes Publishing.
- Schrank, F.A., McGrew, K.S., and LaForte, E.M. (2014). *Technical manual Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside Publishing.
- Schrank, F.A., McGrew, K.S., Mather, N., Wendling, B.J., and LaForte, E.M. (2014). *Woodcock-Johnson IV Tests of Achievement: Form A*. Rolling Meadows, IL: Riverside Publishing.
- Wong, V.C., Cook, T.D., Barnett, W.S., and Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, 27(1), 122–154.