Westat

Improving Lives Through Research®

Pre-K 4 SA Education Centers: Year 11 Supplemental Appendices

January 31, 2025

Authors
Emily Diaz, PhD
Vasiliy Sergueev, BS
Sarah Glidden, BA
Graham Harper, BA
Anne Liu, BA
Lauren Decker-Woodrow, PhD

Prepared For
Early Childhood Education Municipal Development Corporation
7031 S New Braunfels Ave
San Antonio, TX 78223

Prepared By
Westat
1600 Research Boulevard
Rockville, MD 20850-3129

This publication was prepared by Westat under a Professional Services Agreement for the Program Assessment for 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.

Table of Contents

Tables

Table of Contents (continued)

Appendix A

Evaluation Methods

Appendix A
Evaluation Methods

Appendix A provides information on the measures used in the Pre-K 4 SA Year 11 evaluation and details on the analytic approach to the analyses that are described in the body of the report.

Measures

Classroom Assessment Scoring System (CLASS)

The Classroom Assessment Scoring System (CLASS), second edition, (Teachstone, 2023) is an observational system that assesses classroom practices in preschool by measuring the interactions between children and adults. Observations in the Year 11 evaluation consisted of five 20-minute observation periods (or cycles)[1] 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 2023–24 school year. Interactions were measured on 3 domains that were subdivided into 10 total dimensions (see Table A-1 for descriptions of each CLASS dimension). The Emotional Support domain is measured using four dimensions: positive climate, negative climate, educator sensitivity, and regard for child perspectives. The Classroom Organization domain is measured using three dimensions: behavior management, productivity, and instructional learning formats. The Instructional Support domain is measured using three dimensions: concept development, quality of feedback, and language modeling.

CLASS uses a 7-point Likert-type scale for which a score of 1or 2 is considered low-range and generally indicates low quality; a score of 3, 4, or 5 is midrange and indicates midrange quality; and a score of 6 or 7 is consider in the high-range and indicates high quality. Each dimension and domain are assigned a score during each 20-minute cycle. The number of children and adults in the classroom was also recorded during each 20-minute cycle.

[1] There were 100 (20.12%) irregular cycle lengths observed. Of all cycles that were less than 20 minutes, 47 were 15 minutes, 5 were 16 minutes, 3 were 17 minutes, 4 were 18 minutes, and 5 were 19 minutes. Of all cycles that were longer than 20 minutes, 17 were 21 minutes, 11 were 22 minutes, 2 were 23 minutes, 1 was 24 minutes, 3 were 25 minutes, 1 was 26 minutes, and 1 was 35 minutes.

Table A-1. Descriptions of CLASS (second edition) dimensions		
Domain	**Dimension**	**Description**
Emotional Support	Positive climate	Educators foster connections and a sense of belonging between adults and children, among peers, and as a classroom community. Verbal and nonverbal communications between educators and children and among children convey warmth, respect, and collaboration. The educators' interactions enhance each child's enjoyment of the learning setting and their experience of it as a caring community.
	Negative climate	Educators and children express little relational negativity verbally or nonverbally. Educators and children rarely display irritability, anger, or disrespect toward others. Educators do not enact threats or severe punishment that cause disruptions to relationships in the learning setting.
	Educator sensitivity	Educators are aware of and responsive to children's needs—social, emotional, physical, academic, linguistic, and cognitive. The educators' sensitivity supports children's feelings of safety and comfort in the learning setting and facilitates children's ability to actively participate, explore, and take risks.
	Regard for child perspectives	Educators emphasize children's emerging sense of self and help children develop and express their unique interests, motivations, and points of view by providing opportunities for children to experience autonomy and direct their own learning. Children's interests and choices guide classroom experiences and, as a result, children are meaningful contributors to activities.
Classroom Organization	Behavior management	Educators support children's growing behavioral regulation skills by creating developmentally informed, clear, consistent expectations and proactively supporting cooperative behaviors. Children may demonstrate challenging behaviors as they learn these skills, but educators' methods for preventing and positively redirecting these behaviors result in the occurrences being infrequent, mild, and quickly addressed.
	Productivity	Educators use time and structure activities, routines, and transitions so that children have regular, ongoing opportunities to participate and know how to do so.
	Instructional learning formats	Educators facilitate activities by supporting work and play in ways that enhance children's engagement. Educators balance this facilitation with moments of observation as children engage in independent or peer play or work. Educators support children's general engagement and enhance their focus on specific learning objectives within activities. Through these efforts, children remain deeply engaged in work and play, as demonstrated by their active participation and focused attention.
Instructional Support	Concept development	Educators use instructional strategies and activities that help children learn about and understand concepts and content. Educators facilitate learning opportunities that support children's development of thinking skills and creativity. Factual information is taught in the context of these learning opportunities rather than in rote ways that focus only on memorization or recall of information. Educators help children create meaning by linking new concepts and content to prior knowledge and ensuring it is connected to their lived experiences.
	Quality of feedback	Educators provide feedback that builds on children's knowledge and skills in ways that expand understanding or increase persistence. Effective feedback is extended, specific, and individualized, meeting children where they are and scaffolding support as children deepen and refine their learning. Educators also enhance children's motivation and persistence by encouraging and affirming their efforts rather than their work products.

Domain	Dimension	Description
	Language modeling	Educators promote and expand children's language development and verbal and nonverbal communication skills. Educators support children's development in both the language(s) of instruction and children's home language(s). Educators encourage conversations, provide individualized language support, and use varied descriptive language such that children understand and communicate more in the learning setting.

Teaching Strategies Growth, Observation, and Learning (GOLD)

The Growth, Observation, and Learning (or GOLD) assessment (Lambert, 2020) is a teacher-reported measure selected and used by Pre-K 4 SA to collect information on children's progress in 36 objectives across 6 main categories: cognitive, literacy, oral language, mathematics, physical, and social-emotional. The GOLD assessment is conducted three times (fall, winter, and spring) across the school year. The new updated norms for birth through third grade were used as a comparison to Pre-K 4 SA children.

The Woodcock-Johnson Achievement Test (WJ)

The Woodcock-Johnson Tests of Achievement (WJ IV; Schrank, et al., 2014) is an individually administered norm-referenced test that assesses 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 (.97 for Letter-Word and .92 for Applied Problems), and the reported correlations of the WJ IV with other tests of cognitive ability and achievement range from .83 to .86 (McGrew et al., 2014). This measure has been used in numerous large-scale preschool studies (e.g., Early et al., 2007; Wong et al., 2008). For children requiring assessment in Spanish, matching subtests from the Batería III were used (Muñoz-Sandoval et al., 2005). In both versions, the Letter-Word subtest is a test of basic literacy skills involving symbolic learning and the ability to identify 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. With one exception, testing stops when the child scores zero on six successive items across both subtests in English and Spanish. Testing for Applied Problems in English, however, stops when the child scores zero on five successive items (McGrew et al., 2014). To help understand the range of children's age equivalence in early literacy and early numeracy, descriptive information for both subtests based on the language of assessment is provided in Table A-2. There was some variation in children's assessed level of understanding based on the language of assessment. For example, children assessed in English for Letter-Word had a higher maximum (6 years, 4 months) than children assessed in Spanish (6 years, 0 months).

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

Subtest	English		Spanish	
	Minimum	Maximum	Minimum	Maximum
Letter-Word	2 years, 4 months	6 years, 4 months	2 years, 0 months	6 years, 0 months
Applied Problems	2 years, 0 months	6 years, 5 months	2 years, 0 months	5 years, 11 months

Peabody Picture Vocabulary Test (PPVT)

The Peabody Picture Vocabulary Test, fifth edition (PPVT; Dunn & Dunn, 2019) is a test of receptive vocabulary in standard English. The PPVT has established overall reliability (.97), and test-retest reliability (.88), and concurrent and predictive validity based on moderate effect sizes ranging from .46 to .77 when compared 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 must select the image that matches the word said by the examiner. 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.

This measure was only available in English; it was not possible to assess children in Spanish. The Spanish version, Test de Vocabulario en Imagenes Peabody (Dunn et al., 1986), has been discontinued by the publisher[2]. Starting with the 2024–25 school year, children will be assessed using the Receptive and Expressive One-Word Picture Vocabulary Tests in English and Spanish (Martin, 2013a, 2013b; Martin & Brownell, 2011a, 2011b).

Devereux Early Childhood Assessment (DECA)

The Devereux Early Childhood Assessment (DECA; LeBuffe & Naglieri, 1999, 2012) measures children's social-emotional competencies using parent and teacher reports. Parents and teachers report on the frequency of children's behavior on items constituting two main scales and three subscales. Using a 5-point Likert-type scale, scores 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 (Center for Resilient Children, 2013; LeBuffe & Naglieri, 1999). As shown in Table A-3, internal consistency and test-retest reliability are above .80, which is the suggested standard (Nunnally & Bernstein, 1994). For interrater reliability, the parent values are moderate, and the teacher ratings are substantial (McHugh, 2012).

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

	Total protective factors		Behavioral concerns	
	Parent	Teacher	Parent	Teacher
Internal consistency	.92	.95	.80	.86
Test-retest	.88	.95	.78	.88
Interrater	.51	.72	.46	.70

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 = .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 et al., 2011). Therefore, whenever possible parents and teachers should complete the assessment in the same language.

[2] The testing easels to administer the assessment in Spanish have been discontinued but the score sheets are still available. As the measure is being phased out, this led the team to assess children in English only and use a different measure with an accompanying version for Spanish speaking children for the 2025-25 school year.

Analytic Approach

Research questions were addressed by analyzing study data as well as existing Pre-K 4 SA databases. To protect against the disclosure of confidential data, all results based on less than 10 children or percentages that round to 0 percent or 100 percent are masked. Where needed, additional results may be masked or rounded. To address the first set of research 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 11?), 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 11?), an independent samples *t* test between master and non-master teachers in Pre-K 4 SA classrooms was conducted for each domain and dimension. The Benjamini-Hochberg (1995) technique was applied to correct for multiple hypothesis testing.

The address the third set of research questions (3A: How did Pre-K 4 SA children compare to the normative sample on GOLD outcomes?; 3B: Did Pre-K 4 SA children demonstrate significant improvement on GOLD outcomes?; and 3C: What percentage of Pre-K 4 SA children demonstrated kindergarten readiness as measured by GOLD outcomes?), data collected by Pre-K 4 SA were submitted to Westat. Data across fall, winter, and spring assessments were combined for descriptive and inferential analyses. To answer question 3A, an independent samples *t* test between Pre-K 4 SA and the normative sample was conducted in the 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 improvement. To answer question 3C, descriptive statistics were conducted for the spring assessment time to determine the percentage of 4-year-old children who were ready for kindergarten for each GOLD outcome. The Benjamini-Hochberg (1995) technique was applied to correct for multiple hypothesis testing.

As children were not randomly sampled, demographic tests of differences were conducted to determine if the sample of children included in and excluded from analyses were similar. No significant differences were found between children included and not included in analyses for gender ($\chi^2(1) = 0.0076$, $p = .930$), eligibility based on economically disadvantaged status ($\chi^2(2) = 4.5937$, $p = .101$), and eligibility based on English-language learner status ($\chi^2(1) = .0012$, $p = .972$). However, differences were found based on age ($\chi^2(1) = 184.87$, $p < .001$), race and ethnicity ($\chi^2(5) = 20.645$, $p = .001$), and family military affiliation status ($\chi^2(1) = 4.3103$, $p = .038$). Three-year-old children were more likely to be included in the analysis than 4-year-old children. Children included in analyses were more likely to be Hispanic than children excluded from analyses. Children with a family military affiliation were more likely to be included in analysis than children without a family military affiliation.

To address the fourth set of research questions (4A: What percentage of a random sample of Pre-K 4 SA children performed at or above their age level in early literacy and early numeracy, and to what extent did the percentage change?; 4B: Did a random sample of Pre-K 4 SA children demonstrate significant improvement in early literacy and early numeracy?; and 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. The goal was to determine and compare children's early literacy and early numeracy levels for both the fall and spring assessments. For any research question in the fourth set, children had to have data in both the fall and spring to be included in an outcome analysis. 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; McGrew et al., 2014).[3] For age level, the number of months was converted into a proportional figure (e.g., 4 years and 6 months = 4.5) that compares to their actual age in years and months. Another method of analysis is to convert raw scores into grade-level equivalences. For pre-K children, there is only 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. To answer the first part of research question 4A, descriptive analyses were conducted for each outcome. To answer the second part, the percentages below, on, and above age level were computed and analyzed for each outcome. 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. Analyses were conducted based on the binary indicators: a McNemar test was conducted between the fall and spring for each outcome to determine if there was a significant increase in the percentages over time. For research question 4B, dependent *t* tests were conducted for each outcome between the fall and spring to determine if there were significant improvements over time. For research question 4C, the difference between each child's assessed age and biological age was computed for each outcome in the fall and spring. This difference was used to measure the achievement gap for each child by denoting how many months they were above or below the normative sample in their understanding of early literacy and early numeracy, or if they were on par with the normative sample. A difference of differences analysis was conducted to determine if there was a significant reduction of the achievement gap from the fall to spring. The Benjamini-Hochberg (1995) technique was applied to correct for multiple hypothesis testing.

To address the fifth set of research questions (5A: What were the receptive vocabulary performance levels of a random sample of Pre-K 4 SA children?, 5B: Did a random sample of Pre-K 4 SA children demonstrate significant improvement in receptive vocabulary?; and 5C: What types of receptive vocabulary improvement did a random sample of Pre-K 4 SA children demonstrate?), data collected by Pre-K 4 SA were submitted to Westat and descriptively and inferentially analyzed. It is worth noting that the Spanish version, Test de Vocabulario en Imagenes Peabody (Dunn et al., 1986), has been discontinued by the publisher; therefore, it was not possible to assess children in Spanish.[4] Children had to have data in both the fall and spring assessments 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), and standard scores were converted into performance descriptions based on the technical manual. To address research question 5A, descriptive analyses were conducted for the performance levels for each time point. To address research question 5B, dependent *t* tests were conducted separately for the standard scores and growth scale values between the fall and spring to determine if there were significant increases over time. To address research question 5C and aid the interpretation of changes in standard scores and growth scale values over time, score patterns were descriptively analyzed according to the five patterns detailed in the technical manual.[5] The Benjamini-Hochberg (1995) technique was applied to correct for multiple hypothesis testing.

---

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

[4] The testing easels to administer the assessment in Spanish have been discontinued but the score sheets are still available. As the measure is being phased out, this led the team to assess children in English only and use a different measure with an accompanying version for Spanish speaking children for the 2025-25 school year.

[5] 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.

Finally, to address the sixth set of research questions (6A: What were the levels of Pre-K 4 SA children's social-emotional competence, and to what extent did the levels change?; and 6B: Did Pre-K 4 SA children demonstrate significant improvement in social-emotional competence?), data collected by Pre-K 4 SA were submitted to Westat and descriptively and inferentially analyzed. For any research question in the sixth set, children had to have data in both the fall and spring assessments to be included in an outcome analysis. *T* scores were converted into three categorical levels—Needs Instruction, Typical, and Strengths—based on the technical manual (LeBuffe & Naglieri, 2012). To address the first part of research question 6A, descriptive analyses were conducted for the categorical levels for each assessment point. To address the second part of research question 6A, a Wilcoxon signed-rank test was conducted between the fall and spring 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 the fall and spring. To address research question 6B, dependent *t* tests were conducted for the *T* scores for all but one outcome between the fall and spring to determine if there were significant increases over time. The one exception, Behavioral Concerns, is the outcome in which a decrease or reduction in problematic behaviors is the desired result. The Benjamini-Hochberg (1995) technique was applied to correct for multiple hypothesis testing.

As this was the first year of children attending Pre-K 4 SA for two consecutive years as both three-year-olds and four-year olds, inferential analyses were conducted to determine patterns and trends in children's social-emotional competency over time. Children had to have data in the fall and spring assessments for both years (2022-23 and 2023-24) to be included in an outcome analysis. For this subgroup of children, the same methods to answer research question 6A and 6B were used. *T* scores were converted into three categorical levels—Needs Instruction, Typical, and Strengths—based on the technical manual (LeBuffe & Naglieri, 2012). To address the first part of research question 6A, descriptive analyses were conducted for the categorical levels for each assessment point. To address the second part of research question 6A, a Wilcoxon signed-rank test was conducted between the fall 2022 and spring 2024 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 the fall 2022 and spring 2024. To address research question 6B, dependent *t* tests were conducted for the *T* scores for all but one outcome between the fall 2022 and spring 2024 to determine if there were significant increases over time. The one exception, Behavioral Concerns, is the outcome in which a decrease or reduction in problematic behaviors is the desired result. The Benjamini-Hochberg (1995) technique was applied to correct for multiple hypothesis testing. These longitudinal two year findings were also compared to the 2023-24 results to determine if children in their first year of Pre-K 4 SA demonstrated similar patterns as children attending for two years.

As children were not randomly sampled, demographic tests of differences were conducted to determine if the full sample of children included and excluded from 2023-24 analyses were similar. We found no significant differences were found between children included in and excluded from analyses for gender ($\chi^2(1) = 0.261$, $p = .609$), indicators of race and ethnicity ($\chi^2(5) = 8.720$, $p = .121$), English- language learner status ($\chi^2(1) = 1.502$, $p = .220$), or family military affiliation status ($\chi^2(1) = 3.602$, $p = .058$). Significant differences were found between children included in and excluded from analyses for age ($\chi^2(1) = 5.632$, $p = .018$) and attending based on economic disadvantage ($\chi^2(1) = 11.230$, $p = .004$). Children included in analyses were more likely to be 4 years old than children excluded from analyses. Children included in analyses were more likely to have parents or guardians who paid tuition than children excluded from analyses. This implies the findings are not representative of all children in the education centers.

As children were not randomly sampled, demographic tests of differences were conducted to determine if the sample of children included and excluded from the children attending as both 3-year-olds and 4-year-olds (two year DECA analyses) were similar. We found no significant differences were found between children included in and excluded from analyses for gender ($\chi^2(1) = 0.689$, $p = .407$), indicators of race and ethnicity ($\chi^2(5) = 3.521$, $p = .620$), English- language learner status ($\chi^2(1) = 3.384$, $p = .066$), family military affiliation status ($\chi^2(1) = 0.056$, $p = .812$), age ($\chi^2(1) = 2.020$, $p = .155$), or attending based on economic disadvantage ($\chi^2(1) = <.001$, $p = 1.000$). This implies the findings are representative of all children in the education centers.

Appendix B

Additional Classroom Assessment Scoring System (CLASS) Results

Appendix B
Additional Classroom Assessment Scoring System (CLASS) Results

Appendix B provides additional classroom quality results that address the second set of research questions. 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 11?) by providing descriptives for each center. The overall results are provided in the technical report in Table 4. Table -2 addresses research question 2B (Did master teachers of Pre-K 4 SA classrooms have higher observed teacher–child interaction quality in Year 11?).

Table B-1. Average Year 11 CLASS scores by center

CLASS outcome	East		North		South		West	
	M (SD)	Total range observed	M (SD)	Total range observed	M (SD)	Total range observed	M (SD)	Total range observed
Emotional Support	6.07 (0.79)	(4.45–7.00)	6.44 (0.58)	(5.05–7.00)	6.57 (0.38)	(5.75–7.00)	6.68 (0.43)	(5.45–7.00)
Positive climate	6.04 (1.01)	(3.80–7.00)	6.57 (0.65)	(4.80–7.00)	6.70 (0.39)	(5.60–7.00)	6.70 (0.58)	(4.80–7.00)
Negative climate[a]	6.55 (0.52)	(5.00–7.00)	6.79 (0.33)	(6.00–7.00)	6.83 (0.24)	(6.40–7.00)	6.94 (0.12)	(6.60–7.00)
Educator sensitivity	5.82 (0.95)	(4.20–7.00)	6.12 (0.78)	(4.20–7.00)	6.42 (0.66)	(5.20–7.00)	6.55 (0.57)	(5.20–7.00)
Regard for child perspectives	5.87 (0.89)	(4.00–7.00)	6.28 (0.72)	(4.80–7.00)	6.34 (0.55)	(4.80–7.00)	6.54 (0.63)	(5.00–7.00)
Classroom Organization	5.81 (0.94)	(3.87–7.00)	6.14 (0.70)	(4.27–7.00)	6.50 (0.42)	(5.67–7.00)	6.39 (0.61)	(4.80–7.00)
Behavior management	5.97 (0.90)	(4.20–7.00)	6.27 (0.76)	(4.60–7.00)	6.58 (0.59)	(4.80–7.00)	6.68 (0.38)	(5.60–7.00)
Productivity	5.96 (1.02)	(3.80–7.00)	6.36 (0.74)	(4.20–7.00)	6.66 (0.51)	(5.20–7.00)	6.55 (0.65)	(4.80–7.00)
Instructional learning formats	5.49 (1.06)	(3.60–7.00)	5.77 (0.95)	(3.40–7.00)	6.27 (0.57)	(5.20–7.00)	5.94 (0.89)	(3.80–7.00)
Instructional Support	3.73 (1.27)	(1.73–6.00)	3.90 (1.38)	(1.33–6.07)	4.42 (1.18)	(2.27–6.00)	3.98 (1.32)	(2.40–6.40)
Concept development	3.38 (1.38)	(1.00–6.00)	3.50 (1.46)	(1.00–6.00)	3.83 (1.17)	(1.20–6.00)	3.48 (1.48)	(1.40–6.40)
Quality of feedback	3.90 (1.44)	(1.60–6.20)	4.21 (1.51)	(1.40–6.60)	4.72 (1.33)	(2.60–6.60)	4.31 (1.24)	(2.80–6.60)
Language modeling	3.91 (1.21)	(2.20–5.80)	3.98 (1.29)	(1.60–6.60)	4.72 (1.24)	(2.20–6.60)	4.14 (1.39)	(2.20–6.60)

Notes: Domains are in bold font, and dimensions are in plain font. Because of rounding, decimals may not agree to the nearest hundredths.

M = mean; SD = standard deviation.

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

Table B-2. Classroom quality results comparing master and non-master teachers									
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	100	6.36	6.56	0.20	1.75	98.00	.084	Nonsignificant	No difference
Positive climate	100	6.42	6.62	0.20	1.42	97.94	.158	Nonsignificant	No difference
Negative climate	100	6.73	6.85	0.12	1.75	96.16	.083	Nonsignificant	No difference
Educator sensitivity	100	6.11	6.39	0.27	1.78	96.51	.078	Nonsignificant	No difference
Regard for child perspectives	100	6.17	6.38	0.21	1.47	96.96	.144	Nonsignificant	No difference
Classroom Organization	100	6.13	6.32	0.20	1.40	98.00	.165	Nonsignificant	No difference
Behavior management	100	6.30	6.48	0.18	1.27	95.75	.210	Nonsignificant	No difference
Productivity	100	6.30	6.50	0.20	1.37	97.43	.174	Nonsignificant	No difference
Instructional learning formats	100	5.78	5.99	0.20	1.14	96.20	.260	Nonsignificant	No difference
Instructional Support	100	4.00	4.01	0.01	0.04	91.25	.970	Nonsignificant	No difference
Concept development	100	3.60	3.47	−0.14	−0.49	91.68	.625	Nonsignificant	No difference
Quality of feedback	100	4.27	4.31	0.05	0.17	92.80	.862	Nonsignificant	No difference
Language modeling	100	4.14	4.26	0.12	0.44	88.88	.660	Nonsignificant	No difference

Note: *df* = degrees of freedom. Domains are in bold font, and dimensions are in plain font. Because of rounding, decimals may not agree to the nearest hundredths.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the group whose score was greater (i.e., the "favored" group) is listed in this column. If there was no statistically significant difference, this column states that there was "no difference."

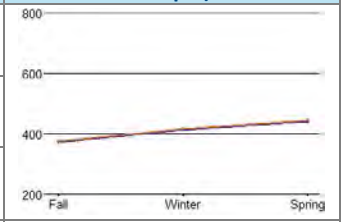
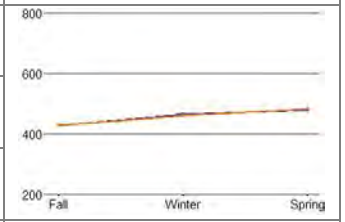
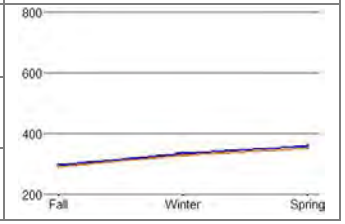
Appendix C

Additional Growth, Observation, and Learning (GOLD) Results

Appendix C
Additional Classroom Assessment Scoring System (CLASS) Results

Appendix C provides additional GOLD results that address the third set of research questions. There are four tables provided. Tables C-1 and C-2 address research question 3A (How did Pre-K 4 SA children compare to the normative sample on GOLD outcomes?). The results are presented separately for 3-year-old children and 4-year-old children as the norms vary depending on age level. Table C-3 addresses research question 4B (Did Pre-K 4 SA children demonstrate significant improvement on GOLD outcomes?). Table C-4 addresses research question 3C (What percentage of Pre-K 4 SA children demonstrated kindergarten readiness as measured by GOLD outcomes?).

Table C-1a. Pre-K 4 SA (3-year-old) and normative sample comparison results for six GOLD outcomes across time

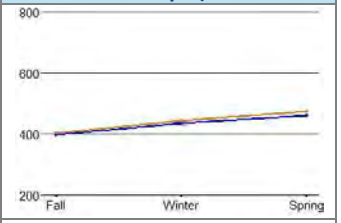
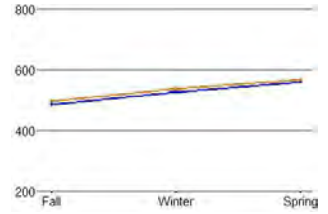
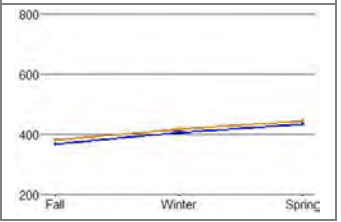
Outcome	Assessment time	Pre-K 4 SA mean	Normed mean	Gap (3-year-old–normed)	*t*-test statistic	*df*	Initial *p* value	Adjusted significance	Group favored[a]	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normative sample)
Cognitive	Fall	372.49	374.62	−2.13	−1.184	1830.790	.2365	Nonsignificant	No difference	
	Winter	413.58	415.58	−2.00	−1.139	1787.890	.2548	Nonsignificant	No difference	
	Spring	440.61	444.42	−3.81	−2.099	1909.270	.0360	Nonsignificant	No difference	
Literacy	Fall	426.27	428.30	−2.03	−1.062	1513.080	.2886	Nonsignificant	No difference	
	Winter	462.85	461.74	1.11	0.740	1624.450	.4593	Nonsignificant	No difference	
	Spring	478.99	481.08	−2.09	−1.323	1532.010	.1859	Nonsignificant	No difference	
Mathematics	Fall	296.34	290.08	6.26	3.562	1802.590	.0004	Significant	Pre-K 4 SA	
	Winter	334.41	327.62	6.79	4.463	1838.370	<.0001	Significant	Pre-K 4 SA	
	Spring	359.95	352.05	7.90	5.260	1874.330	<.0001	Significant	Pre-K 4 SA	

Note: Group mean information is presented in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth.

df = degrees of freedom.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the group whose score was greater (i.e., the "favored" group) is listed in this column. If there was no statistically significant difference, this column states that there was "no difference."

Table C-1b. Pre-K 4 SA (3-year-old) and normative sample comparison results for six GOLD outcomes across time

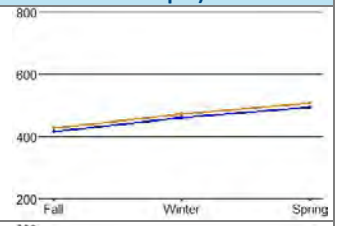
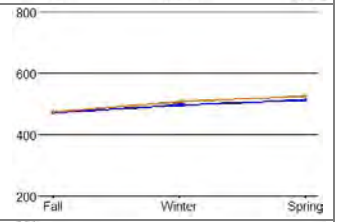
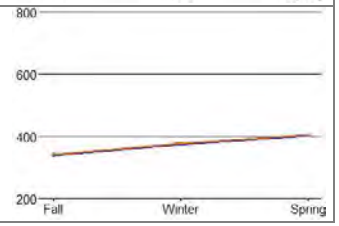
Outcome	Assessment time	Pre-K 4 SA mean	Normed mean	Gap (3-year-old–normed)	*t*-test statistic	*df*	Initial *p* value	Adjusted significance	Group favored[a]	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normative sample)
Oral language	Fall	396.20	402.14	−5.94	−2.292	1383.800	.0220	Significant	Normed	
	Winter	434.65	442.82	−8.17	−3.275	1400.830	.0011	Significant	Normed	
	Spring	460.88	474.07	−13.19	−4.683	1124.170	<.0001	Significant	Normed	
Physical	Fall	484.73	496.07	−11.34	−5.577	1793.640	<.0001	Significant	Normed	
	Winter	526.45	537.37	−10.92	−5.434	1613.040	<.0001	Significant	Normed	
	Spring	558.54	567.59	−9.05	−4.617	1825.830	<.0001	Significant	Normed	
Social-emotional	Fall	368.24	379.88	−11.64	−6.770	1692.760	<.0001	Significant	Normed	
	Winter	406.93	416.85	−9.92	−5.883	1637.470	<.0001	Significant	Normed	
	Spring	432.43	442.81	−10.38	−6.388	1999.960	<.0001	Significant	Normed	

Note: Group mean information is presented in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth.

df = degrees of freedom.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the group whose score was greater (i.e., the "favored" group) is listed in this column. If there was no statistically significant difference, this column states that there was "no difference."

Table C-2a. Pre-K 4 SA (4-year-old) and normative sample comparison results for six GOLD outcomes across time

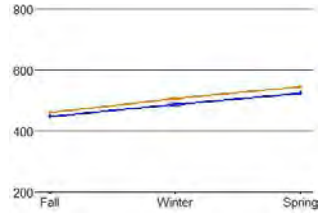
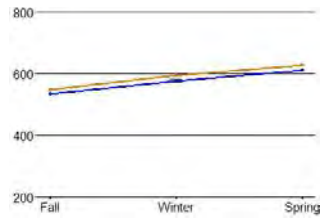
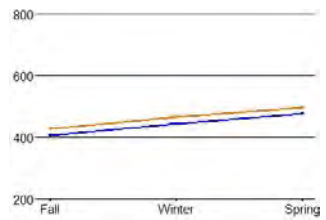
Outcome	Assessment time	Pre-K 4 SA mean	Normed mean	Gap (4-year-old–normed)	*t*-test statistic	*df*	Initial *p* value	Adjusted significance	Group favored[a]	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normative sample)
Cognitive	Fall	416.76	427.94	−11.18	−5.293	741.718	<.0001	Significant	Normed	
	Winter	459.63	472.29	−12.66	−5.763	679.139	<.0001	Significant	Normed	
	Spring	493.81	506.46	−12.65	−5.068	669.688	<.0001	Significant	Normed	
Literacy	Fall	470.02	474.47	−4.45	−2.274	631.083	.0233	Significant	Normed	
	Winter	497.02	506.52	−9.50	−5.590	592.689	<.0001	Significant	Normed	
	Spring	512.90	527.01	−14.11	−6.700	561.755	<.0001	Significant	Normed	
Mathematics	Fall	340.08	341.41	−1.33	−0.660	677.008	.5093	Nonsignificant	No difference	
	Winter	375.96	379.14	−3.18	−1.770	673.124	.0773	Nonsignificant	No difference	
	Spring	402.33	405.25	−2.92	−1.400	647.479	.1632	Nonsignificant	No difference	

Note: Group mean information is presented in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth.

df = degrees of freedom.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the group whose score was greater (i.e., the "favored" group) is listed in this column. If there was no statistically significant difference, this column states that there was "no difference."

Table C-2b. Pre-K 4 SA (4-year-old) and normative sample comparison results for six GOLD outcomes across time

Outcome	Assessment time	Pre-K 4 SA mean	Normed mean	Gap (4-year-old–normed)	*t*-test statistic	*df*	Initial *p* value	Adjusted significance	Group favored[a]	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normative sample)
Oral language	Fall	447.82	461.29	−13.47	−4.249	582.008	<.0001	Significant	Normed	
	Winter	488.24	507.14	−18.90	−5.683	567.354	<.0001	Significant	Normed	
	Spring	523.94	544.96	−21.02	−4.706	531.842	<.0001	Significant	Normed	
Physical	Fall	533.57	547.91	−14.34	−6.005	701.476	<.0001	Significant	Normed	
	Winter	576.13	593.33	−17.20	−7.163	658.008	<.0001	Significant	Normed	
	Spring	612.22	628.50	−16.28	−6.678	680.323	<.0001	Significant	Normed	
Social-emotional	Fall	405.83	426.44	−20.61	−9.623	667.520	.0010	Significant	Normed	
	Winter	443.84	466.51	−22.67	−10.668	659.600	<.0001	Significant	Normed	
	Spring	475.96	497.35	−21.39	−9.493	697.621	<.0001	Significant	Normed	

Note: Group mean information is presented in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth.

df = degrees of freedom.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the group whose score was greater (i.e., the "favored" group) is listed in this column. If there was no statistically significant difference, this column states that there was "no difference."

Table C-3. Growth results comparing fall 2023 and spring 2024 by GOLD outcome and age level										
Age level	Outcome	Sample size	Fall 2023 mean	Spring 2024 mean	Growth (spring 2024–fall 2023)	*t*-test statistic	*df*	Initial *p* value	Adjusted significance	Time favored[a]
3-year-old	Cognitive	945	372.49	440.61	68.12	−59.086	944	<.0001	Significant	Spring 2024
	Literacy	894	426.27	478.99	52.72	−47.135	893	<.0001	Significant	Spring 2024
	Mathematics	947	296.34	359.95	63.61	−57.961	946	<.0001	Significant	Spring 2024
	Oral language	887	396.20	460.88	64.68	−34.225	886	<.0001	Significant	Spring 2024
	Physical	953	484.73	558.54	73.81	−50.374	952	<.0001	Significant	Spring 2024
	Social-emotional	950	368.24	432.43	64.19	−59.460	949	<.0001	Significant	Spring 2024
4-year-old	Cognitive	502	416.76	493.81	77.05	−41.227	501	<.0001	Significant	Spring 2024
	Literacy	473	470.02	512.90	42.88	−31.941	472	<.0001	Significant	Spring 2024
	Mathematics	502	340.08	402.33	62.25	−38.754	501	<.0001	Significant	Spring 2024
	Oral language	469	447.82	523.94	76.12	−25.898	468	<.0001	Significant	Spring 2024
	Physical	503	533.57	612.22	78.65	−41.209	502	<.0001	Significant	Spring 2024
	Social-emotional	502	405.83	475.96	70.13	−41.070	501	<.0001	Significant	Spring 2024

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

df = degrees of freedom.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the time whose score was greater (i.e., 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-4. Descriptives of kindergarten readiness by GOLD outcome		
Outcome	Sample size	Percentage ready for kindergarten
Cognitive	502	75.50
Literacy	473	61.52
Mathematics	502	67.93
Oral language	469	76.12
Physical	503	82.31
Social-emotional	502	75.10

Appendix D

Additional Woodcock-Johnson and Batería Results

Appendix D
Additional Woodcock-Johnson and Batería Results

Appendix D provides additional direct child assessment results from the Woodcock-Johnson and Batería that address the fourth set of research questions. There are three tables provided. Table D-1 addresses both parts of research question 4A (What percentage of Pre-K 4 SA children performed at or above their age level in early literacy and early numeracy, and to what extent did the percentage change?). Table D-2 addresses research question 4B (Did a random sample of Pre-K 4 SA children demonstrate significant improvement in early literacy and early numeracy?). Table D-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 D-1. Percentage analysis results meeting age equivalency comparing fall 2023 and spring 2024 by subtest and years with Pre-K 4 SA										
Outcome	**Years with Pre-K 4 SA**	**Sample size**	**Fall 2023 percentage**	**Spring 2024 percentage**	**Difference (spring 2024–fall 2023)**	**χ^2 statistic**	***df***	**Initial *p* value**	**Significance**	**Time favored[a]**
Letter-Word	New children	39	41.03	38.46	-2.57	0.00	38	1.0000	Nonsignificant	No Change
	Returning children	27	29.63	18.52	-11.11	29.63	26	.2482	Nonsignificant	No Change
	Total	66	36.36	30.30	-6.06	1.13	65	.2888	Nonsignificant	No Change
Applied Problems	New children	39	15.38	35.90	20.52	4.90	38	.0268	Significant	Spring 2024
	Returning children	27	25.92	37.04	11.12	1.33	26	.2482	Nonsignificant	No Change
	Total	66	19.70	36.36	16.66	7.69	65	.0056	Significant	Spring 2024

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

df = degrees of freedom.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the time whose score was greater (i.e., 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-2. Age equivalencies results comparing fall 2023 and spring 2024 by subtest and years with Pre-K 4 SA										
Outcome	**Years with Pre-K 4 SA**	**Sample size**	**Fall 2023 mean**	**Spring 2024 mean**	**Growth (spring 2024–fall 2023)**	***t*-test statistic**	***df***	**Initial *p* value**	**Adjusted significance**	**Time favored[a]**
Letter-Word	New children	39	3 years, 10 months	4 years, 2 months	3 months[b]	3.05	38	.0041	Significant	Spring 2024
	Returning children	27	4 years, 0 months	4 years, 5 months	5 months	5.35	26	<.0001	Significant	Spring 2024
	Total	66	3 years, 11 months	4 years, 3 months	4 months	5.42	65	<.0001	Significant	Spring 2024
Applied Problems	New children	39	3 years, 2 months	3 years, 11 months	9 months	8.40	38	<.0001	Significant	Spring 2024
	Returning children	27	4 years, 2 months	4 years, 8 months	7 months[b]	3.99	26	.0005	Significant	Spring 2024
	Total	66	3 years, 7 months	4 years, 3 months	8 months	8.60	65	<.0001	Significant	Spring 2024

Note: Letter-Word measures early literacy skills, and Applied Problems measures early numeracy skills.

df = degrees of freedom.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the time whose score was greater (i.e., the "favored" time) is listed in this column. If there was no statistically significant difference, this column states that there was "no difference."

[b] Because of rounding, the gap is not the exact difference between fall and spring.

Table D-3. Gap analysis age equivalency results comparing fall 2023 and spring 2024 by subtest and years with Pre-K 4 SA										
Outcome	**Years with Pre-K 4 SA**	**Sample size**	**Average gap between age equivalency and actual age, fall 2023**	**Average gap between age equivalency and actual age, spring 2024**	**Gap (spring 2024–fall 2023)**	***t*-test statistic**	***df***	**Initial *p* value**	**Adjusted significance**	**Gap reduced[a]**
Letter-Word	New children	39	−1 month	−3 months	−3 months[b]	−2.61	38	.0129	Significant	No
	Returning children	27	−8 months	−8 months	−1 month[b]	−0.75	26	.4600	Nonsignificant	No Change
	Total	66	−3 months	−5 months	−2 months	−2.59	65	.0119	Significant	No
Applied Problems	New children	39	−9 months	−6 months	3 months	2.68	38	.0107	Significant	Yes
	Returning children	27	−6 months	−5 months	1 month	0.41	26	.6859	Nonsignificant	No Change
	Total	66	−7 months	−6 months	2 months[b]	2.11	65	.0389	Nonsignificant	No Change

Note: Letter-Word measures early literacy skills, and Applied Problems measures early numeracy skills.

df = degrees of freedom.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, 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 change."

[b] Because of rounding, the gap is not the exact difference between fall and spring.

Appendix E

Additional Peabody Picture Vocabulary Test (PPVT) Results

Appendix E
Additional Peabody Picture Vocabulary Test (PPVT) Results

Appendix E provides additional direct child assessment results from the Peabody Picture Vocabulary Test to address the fifth set of research questions. There are three tables provided. Table E-1 addresses research question 5A (What were the receptive vocabulary performance levels of a random sample of Pre-K 4 SA children?). Table E-2 addresses research question 5B (Did a random sample of Pre-K 4 SA children demonstrate significant improvement in receptive vocabulary?). Table E-3 addresses research question 5C (What types of receptive vocabulary improvement did a random sample of Pre-K 4 SA children demonstrate?).

Table E-1. Descriptives of vocabulary performance levels by assessment time and years with Pre-K 4 SA

Level	Years with Pre-K 4 SA	Sample size	Fall 2023 percentage	Spring 2024 percentage	Growth (spring 2024–fall 2023)
Well below expected	New children	40	≤3.00	≤3.00	−2.50
Below expected			15.00	17.50	2.50
Expected			62.50	80.00	17.50
Above expected			≤18.00	≤3.00	−15.00
Well above expected			≤8.00	≤3.00	−2.50
Well below expected	Returning children	20	≤12.00	≤4.00	−4.00
Below expected			16.00	20.00	4.00
Expected			76.00	76.00	0.00
Above expected			≤4.00	≤4.00	0.00
Well above expected			≤4.00	≤4.00	0.00
Well below expected	Total	65	≤5.00	≤2.00	−3.08
Below expected			15.39	18.46	3.07
Expected			67.69	78.46	10.77
Above expected			≤11.00	≤2.00	−9.23
Well above expected			≤5.00	≤2.00	−1.54

Note: Because of rounding, decimals may not agree to the nearest hundredths, and percentages may sum to more than 100 percent.

Table E-2. Vocabulary growth results comparing fall 2023 and spring 2024 by score type and years with Pre-K 4 SA										
Score type	**Years with Pre-K 4 SA**	**Sample size**	**Fall 2023 mean**	**Spring 2024 mean**	**Growth (spring 2024–fall 2023)**	***t*-test statistic**	***df***	**Initial *p* value**	**Adjusted significance**	**Time favored[a]**
Standard Score	New children	40	97.53	97.55	0.02	0.02	39	.988	Nonsignificant	None
	Returning children	25	92.80	93.00	0.20	0.12	24	.908	Nonsignificant	None
	Total	65	95.71	95.80	0.09	0.08	64	.938	Nonsignificant	None
Growth Scale Value	New children	40	460.93	466.95	6.02	5.32	39	<.001	Significant	Spring 2024
	Returning children	25	466.48	471.20	4.72	4.58	24	<.001	Significant	Spring 2024
	Total	65	463.06	468.59	5.53	6.90	64	<.001	Significant	Spring 2024

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

df = degrees of freedom.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the time whose score was greater (i.e., the "favored" time) is listed in this column. If there was no statistically significant difference, this column states that there was "no difference."

Table E-3. Descriptives of vocabulary score growth comparing fall 2023 and spring 2024 by years with Pre-K 4 SA			
Score growth	**Years with Pre-K 4 SA**	***N***	**Percentage**
Stagnant/losing and widened gap	New children	*	*
Learning but widened gap		12	30.00
Learning slowed but still on par		*	*
Learning		16	40.00
Learning and narrowed gap		*	*
Stagnant/losing and widened gap	Returning children	*	*
Learning but widened gap		*	*
Learning slowed but still on par		*	*
Learning		10	40.00
Learning and narrowed gap		*	*
Stagnant/losing and widened gap	Total	12	18.46
Learning but widened gap		17	26.15
Learning slowed but still on par		*	*
Learning		26	40.00
Learning and narrowed gap		*	*

Note: Because of rounding, decimals may not agree to the nearest hundredths, and percentages may sum to more than 100 percent. * = To protect confidentiality, sample sizes and percentages are masked.

Appendix F

Additional Devereux Early Childhood Assessment (DECA) Results

Appendix F
Additional Devereux Early Childhood Assessment (DECA) Results

Appendix F provides additional social-emotional assessment results to address the sixth set of research questions. There are four tables provided. Tables F-1 and F-2 present findings for the 2023-24 school year. Tables F-3 and F-4 present findings for children who attended for two years (2022-23 and 2023-24) when they were 3- and 4-year-olds. Tables F-1 and F-3 address both parts of research question 6A (What were the levels of Pre-K 4 SA children's social-emotional competence, and to what extent did the levels change?). Table F-1 displays the findings based on the 2023-24 school year. Table F-3 displays the findings for children who attended for two years (2022-23 and 2023-24) when they were 3- and 4-year-olds. Tables F-2 and F-4 addresses research question 6B (Did Pre-K 4 SA children demonstrate significant improvement in social-emotional competence?). Table F-2 displays the findings based on the 2023-24 school year. Table F-4 displays the findings for children who attended for two years (2022-23 and 2023-24) when they were 3- and 4-year-olds.

Table F-1.	Social-emotional results comparing levels in fall 2023 and spring 2024 by outcome and years with Pre-K 4 SA									
Outcome	**Years with Pre-K 4 SA**	**Level**	**Sample size**	**Fall 2023 percentage**	**Spring 2024 percentage**	**Growth (spring 2024–fall 2023)**	**χ^2 statistic**	**Initial *p* value**	**Adjusted significance**	**Time favored[a]**
Initiative	New children	Needs Instruction	1,352	32.84	13.76	−19.08	14,518	<.0001	Significant	Spring 2024
		Typical		61.98	71.08	9.10				
		Strengths		5.18	15.16	9.98				
	Returning children	Needs Instruction	421	12.59	3.09	−9.50	949	<.0001	Significant	Spring 2024
		Typical		73.63	62.47	−11.16				
		Strengths		13.78	34.44	20.66				
	Total	Needs Instruction	1,773	28.03	11.22	−16.81	22,977	<.0001	Significant	Spring 2024
		Typical		64.75	69.04	4.29				
		Strengths		7.22	19.74	12.52				
Self-control	New children	Needs Instruction	1,352	24.11	15.98	−8.13	20,811	<.0001	Significant	Spring 2024
		Typical		66.86	67.23	0.37				
		Strengths		9.02	16.79	7.77				
	Returning children	Needs Instruction	421	14.73	10.21	−4.52	1,871	<.0001	Significant	Spring 2024
		Typical		69.36	59.38	−9.98				
		Strengths		15.91	30.40	14.49				
	Total	Needs Instruction	1,773	21.88	14.61	−7.27	35,113	<.0001	Significant	Spring 2024
		Typical		67.46	65.37	−2.09				
		Strengths		10.66	20.02	9.36				

Outcome	Years with Pre-K 4 SA	Level	Sample size	Fall 2023 percentage	Spring 2024 percentage	Growth (spring 2024–fall 2023)	χ^2 statistic	Initial *p* value	Adjusted significance	Time favored[a]
Attachment	New children	Needs Instruction	1,352	22.78	9.69	−13.09	21,255	<.0001	Significant	Spring 2024
		Typical		70.86	72.04	1.18				
		Strengths		6.36	18.27	11.91				
	Returning children	Needs Instruction	421	9.26	4.99	−4.27	1,787	<.0001	Significant	Spring 2024
		Typical		77.91	65.32	−12.59				
		Strengths		12.83	29.69	16.86				
	Total	Needs Instruction	1,773	19.57	8.57	−11.00	35,327	<.0001	Significant	Spring 2024
		Typical		72.53	70.45	−2.08				
		Strengths		7.90	20.98	13.08				
Total Protective Factors	New children	Needs Instruction	1,352	28.11	12.72	−15.39	17,100	<.0001	Significant	Spring 2024
		Typical		65.09	67.75	2.66				
		Strengths		6.81	19.53	12.72				
	Returning children	Needs Instruction	421	12.59	5.23	−7.36	1,716	<.0001	Significant	Spring 2024
		Typical		70.55	56.53	−14.02				
		Strengths		16.86	38.24	21.38				
	Total	Needs Instruction	1,773	24.42	10.94	−13.48	29,610	<.0001	Significant	Spring 2024
		Typical		66.38	65.09	−1.29				
		Strengths		9.19	23.97	14.78				

Outcome	Years with Pre-K 4 SA	Level	Sample size	Fall 2023 percentage	Spring 2024 percentage	Growth (spring 2024–fall 2023)	χ^2 statistic	Initial *p* value	Adjusted significance	Time favored[a]
Behavioral Concerns	New children	Typical	1,352	82.40	82.47	0.07	0.00	1.00	Nonsignificant	No difference
		Area of Need		17.60	17.53	−0.07				
	Returning children	Typical	421	84.56	86.22	1.66	0.77	0.381	Nonsignificant	No difference
		Area of Need		15.44	13.78	−1.66				
	Total	Typical	1,773	82.91	83.36	0.45	0.20[b]	0.657	Nonsignificant	No difference
		Area of Need		17.09	16.63	−0.46				

Note: Because of rounding, decimals may not agree to the nearest hundredths, and percentages may sum to more than 100 percent.

df = degrees of freedom.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the time whose score was greater (i.e., the "favored" time) is listed in this column. If there was no statistically significant difference, this column states that there was "no difference."

[b] A McNemar's test was conducted for Behavioral Concerns because it had two classifications. The statistic is chi-square with 1 degree of freedom.

Table F-2. Social-emotional equivalencies results comparing fall 2023 and spring 2024 by outcome and years with Pre-K 4 SA										
Outcome	**Years with Pre-K 4 SA**	**Sample size**	**Fall 2023 mean**	**Spring 2024 mean**	**Growth (spring 2024–fall 2023)**	***t*-test statistic**	***df***	**Initial *p* value**	**Adjusted significance**	**Time favored[a]**
Initiative	New children	1,352	45.48	51.10	5.62	25.27	1351	<.0001	Significant	Spring 2024
	Returning children	421	51.27	57.12	5.85	13.85	420	<.0001	Significant	Spring 2024
	Total	1,773	46.85	52.53	5.68	28.81	1772	<.0001	Significant	Spring 2024
Self-control	New children	1,352	46.71	50.34	3.63	16.60	1351	<.0001	Significant	Spring 2024
	Returning children	421	50.08	54.15	4.07	10.51	420	<.0001	Significant	Spring 2024
	Total	1,773	47.51	51.25	3.74	19.61	1772	<.0001	Significant	Spring 2024
Attachment	New children	1,352	46.98	51.73	4.75	19.58	1351	<.0001	Significant	Spring 2024
	Returning children	421	50.61	54.96	4.35	9.90	420	<.0001	Significant	Spring 2024
	Total	1,773	47.84	52.50	4.66	21.92	1772	<.0001	Significant	Spring 2024
Total Protective Factors	New children	1,352	45.84	51.26	5.42	24.11	1351	<.0001	Significant	Spring 2024
	Returning children	421	50.80	56.26	5.46	13.31	420	<.0001	Significant	Spring 2024
	Total	1,773	47.02	52.56	5.54	27.55	1772	<.0001	Significant	Spring 2024
Behavioral Concerns	New children	1,352	50.06	49.12	−0.94[b]	−4.38	1351	<.0001	Significant	Spring 2024
	Returning children	421	47.86	46.44	−1.42[b]	−3.77	420	<.0001	Significant	Spring 2024
	Total	1,773	49.54	48.49	−1.05[b]	−5.65	1772	<.0001	Significant	Spring 2024

Note: *df* = degrees of freedom. Because of rounding, decimals may not agree to the nearest hundredth.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the time whose score was greater (i.e., 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 2024, indicating a positive finding.

Table F-3. Social-emotional results comparing levels from fall 2022 to spring 2024 by age and outcome for children attending as both 3-year-olds and 4-year-olds

Outcome	Level	Sample size	3-year-old		4-year-old		Growth (spring 2024–fall 2022)	χ^2 statistic	Initial *p* value	Adjusted significance	Time favored[a]
			Fall 2022 percentage	Spring 2023 percentage	Fall 2023 percentage	Spring 2024 percentage					
Initiative	Needs Instruction	420	35.48	19.76	11.90	3.33	-32.15	663	<.0001	Significant	Spring 2024
	Typical		61.67	67.86	73.81	62.62	0.95				
	Strengths		2.86	12.38	14.29	34.05	31.19				
Self-Control	Needs Instruction	420	21.90	16.67	15.00	10.71	-11.19	1260	<.0001	Significant	Spring 2024
	Typical		71.43	65.24	69.05	60.00	-11.43				
	Strengths		6.67	18.10	15.95	29.29	22.62				
Attachment	Needs Instruction	420	21.67	9.52	9.29	5.00	-16.67	2288	<.0001	Significant	Spring 2024
	Typical		69.05	71.90	76.90	65.00	-4.05				
	Strengths		9.29	18.57	13.81	30.00	20.71				
Total Protective Factors	Needs Instruction	420	28.33	15.95	12.38	5.71	-22.62	1379	<.0001	Significant	Spring 2024
	Typical		64.76	67.38	70.00	56.67	-8.09				
	Strengths		6.90	16.67	17.62	37.62	30.72				
Behavioral Concerns	Typical	420	85.24	84.52	84.29	85.71	0.47	0.02	0.89	Not significant	No difference
	Area of Need		14.76	15.48	15.71	14.29	-0.47				

Note: Because of rounding, decimals may not agree to the nearest hundredths, and percentages may sum to more than 100 percent.

df = degrees of freedom.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the time whose score was greater (i.e., the "favored" time) is listed in this column. If there was no statistically significant difference, this column states that there was "no difference."

[b] A McNemar's test was conducted for Behavioral Concerns because it had two classifications. The statistic is chi-square with 1 degree of freedom.

Table F-4. Social-emotional equivalencies results comparing levels from fall 2022 to spring 2024 by age and outcome for children attending as both 3-Year-Olds and 4-Year-Olds

Outcome	Sample size	3-year-old		4-year-old		Growth	*t*-test statistic	*df*	Initial *p* value	Adjusted significance	Time favored[a]
		Fall 2022 mean	Spring 2023 mean	Fall 2023 mean	Spring 2024 mean						
Initiative	420	44.11	49.46	51.44	57.05	12.94	27.74	419	<.0001	Significant	Spring 2024
Self-control	420	46.5	49.85	50.09	53.96	7.47	16.92	419	<.0001	Significant	Spring 2024
Attachment	420	47.82	52.33	50.77	54.96	7.15	13.68	419	<.0001	Significant	Spring 2024
Total Protective Factors	420	45.54	50.65	50.92	56.19	10.65	22.98	419	<.0001	Significant	Spring 2024
Behavioral Concerns	420	49.73	49.22	47.96	46.57	-3.15	-7.46	419	<.0001	Significant	Spring 2024

Note: *df* = degrees of freedom. Because of rounding, decimals may not agree to the nearest hundredth.

[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the time whose score was greater (i.e., 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 2024, indicating a positive finding.

Appendix References

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

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.* Center for Resilient Children. https://centerforresilientchildren.org/wp-content/uploads/2014/01/DECA-P2-full-version.pdf.

Crane, J., Mincic, M. S., & 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 & Development, 22*(3), 520–47.

Dunn, L., & Dunn, L. (2019). *Peabody Picture Vocabulary Test, fifth edition (PPVT -V).* NCS Pearson.

Dunn, L. M., Lugo, D. E., Padilla, E. R., & Dunn, L. M. (1986). *Test de Vocabulario en Imagenes Peabody - TVIP.* American Guidance Services.

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., & Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, *78*(2*)*, 558–80.

Lambert, R. (2020). *Technical manual for the Teaching Strategies GOLD assessment system* (2nd ed.). Center for Educational Measurement and Evaluation, University of North Carolina. https://teachingstrategies.com/wp-content/uploads/2020/10/2020-Tech-Manual_GOLD.pdf.

LeBuffe, P. A., & Naglieri, J. A. (1999). *The Devereux early childhood assessment*. Kaplan Press.

LeBuffe, P. A., & Naglieri, J. A. (2012). *Devereux Early Childhood Assessment for preschoolers, second edition (DECA-P2): User's guide and technical manual*. Kaplan Press.

LeBuffe, P. A., Shapiro, V. B., & Naglieri, J. A. (2009). *Devereux student strengths assessment.* Kaplan Press.

Martin, N. A. (2013a). *Expressive One-Word Picture Vocabulary Test (Spanish-Bilingual Edition)*. Academic Therapy Publications, Inc.

Martin, N. A. (2013b). *Receptive One-Word Picture Vocabulary Test (Spanish-Bilingual Edition)*. Academic Therapy Publications, Inc.

Martin, N. A., & Brownell, R. (2011a). *Expressive One-Word Picture Vocabulary Test*. Academic Therapy Publications, Inc.

Martin, N. A., & Brownell, R. (2011b). *Receptive One-Word Picture Vocabulary Test*. Academic Therapy Publications, Inc.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual Woodcock-Johnson IV*. Riverside Publishing.

McHugh, M.L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, *22*(3), 276–82.

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *The Batería III Woodcock-Muñoz: Pruebas de aprovechamiento*. Riverside Publishing.

Nunnally, J. & Bernstein L. (1994). *Psychometric theory*. McGraw-Hill Higher, Inc.

Pearson Education, Inc. (2019). *PPVT by the numbers*. Pearson Education, Inc. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/PPVT5-EVT3-By-the-Numbers-Infographic.pdf.

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

Teachstone. (2023). *Classroom Assessment Scoring System, 2nd Edition: Pre-K–3rd Observation Field Guide.* Teachstone.

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: The Journal of the Association for Public Policy Analysis and Management*, *27*(1), 122–54.