

# Pre-K 4 SA Education Centers: Year 12 (2024–2025) Supplemental Appendices

February 9, 2026

## Authors

Emily Diaz, PhD  
Vasiliy Sergueev, BS  
Sarah Glidden, BA  
Lauren Decker-Woodrow, PhD

## Prepared For

Early Childhood Education Municipal  
Development Corporation  
3510 S WW White Road  
San Antonio, TX 78222

## Prepared By

Westat  
7501 Wisconsin Avenue  
Suite 1000E  
Bethesda, Maryland 20841-6527

*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

<b>Appendix A</b>	<b>Evaluation Methods</b>	<b>A-1</b>
	Measures	A-1
	Classroom Assessment Scoring System (CLASS)	A-1
	Teaching Strategies Growth, Observation, and Learning (GOLD)	A-3
	The Woodcock-Johnson Achievement Test (WJ IV)	A-3
	Receptive One-Word Picture Vocabulary Test (ROWPVT) and Expressive One-Word Picture Vocabulary Test (EOWPVT)	A-4
	Devereux Early Childhood Assessment (DECA)	A-4
	Analytic Approach	A-5
<b>Appendix B</b>	<b>Additional CLASS Results</b>	<b>B-1</b>
<b>Appendix C</b>	<b>Additional GOLD Results</b>	<b>C-1</b>
<b>Appendix D</b>	<b>Additional WJ IV and Bateria Results</b>	<b>D-1</b>
<b>Appendix E</b>	<b>Additional ROWPVT and EOWPVT Results</b>	<b>E-1</b>
<b>Appendix F</b>	<b>Additional DECA Results</b>	<b>F-1</b>
<b>Appendix References</b>		<b>R-1</b>
<b>Tables</b>		
Table A.1.	Descriptions of CLASS (second edition) dimensions	A-2
Table A.2.	Range of age equivalencies by assessment and language	A-3
Table A.3.	Reliability evidence for the ROWPVT and EOWPVT by version	A-4
Table A.4.	Reliability evidence for the DECA by parent and teacher ratings	A-5
Table B.1.	Average Year 12 CLASS scores by center ( $N = 97$ )	B-2
Table B.2.	Classroom quality results comparing master and non-master teachers	B-3
Table C.1a.	Pre-K 4 SA (3-year-old) and normative sample comparison results for Cognitive, Literacy, and Mathematics GOLD outcomes across time	C-2
Table C.1b.	Pre-K 4 SA (3-year-old) and normative sample comparison results for Oral Language, Physical, and Social-Emotional GOLD outcomes across time	C-3

## Table of Contents (continued)

Table C.2a.	Pre-K 4 SA (4-year-old) and normative sample comparison results for Cognitive, Literacy, and Mathematics GOLD outcomes across time	C-4
Table C.2b.	Pre-K 4 SA (4-year-old) and normative sample comparison results for Oral Language, Physical, and Social-Emotional GOLD outcomes across time	C-5
Table C.3.	Growth results comparing fall 2023 and spring 2024 by GOLD outcome and age level	C-6
Table C.4.	Descriptives of kindergarten readiness by GOLD outcome	C-6
Table C.5a.	Pre-K 4 SA and normative sample comparison results for Cognitive and Literacy GOLD outcomes across time for children attending as both 3-year-olds and 4-year-olds	C-7
Table C.5b.	Pre-K 4 SA and normative sample comparison results for Mathematics and Oral Language GOLD outcomes across time for children attending as both 3-year-olds and 4-year-olds	C-8
Table C.5c.	Pre-K 4 SA and normative sample comparison results for Physical and Social-Emotional GOLD outcomes across time for children attending as both 3-year-olds and 4-year-olds	C-9
Table C.6.	Growth results comparing fall 2023 and spring 2025 by GOLD outcome for children attending as both 3-year-olds and 4-year-olds	C-10
Table C.7.	Descriptives of kindergarten readiness by GOLD outcome for children attending as both 3-year-olds and 4-year-olds	C-10
Table D.1.	Percentage analysis results meeting age equivalency comparing fall 2024 and spring 2025 by subtest and years with Pre-K 4 SA	D-2
Table D.2.	Age equivalencies results comparing fall 2024 and spring 2025 by subtest and years with Pre-K 4 SA	D-3
Table D.3.	Gap analysis age equivalency results comparing fall 2024 and spring 2025 by subtest and years with Pre-K 4 SA	D-4
Table E.1.	Descriptives of vocabulary performance levels by assessment time and years with Pre-K 4 SA	E-2
Table E.2.	Vocabulary growth results comparing fall 2024 and spring 2025 by assessment and years with Pre-K 4 SA	E-4
Table F.1.	Social-emotional results comparing levels in fall 2024 and spring 2025 by outcome and years with Pre-K 4 SA	F-2
Table F.2.	Social-emotional equivalencies results comparing fall 2024 and spring 2025 by outcome and years with Pre-K 4 SA	F-5

## Table of Contents (continued)

Table F.3.	Social-emotional results comparing levels from fall 2023 to spring 2025 by age and outcome for children attending as both 3-year-olds and 4-year-olds	F-6
Table F.4.	Social-emotional equivalencies results comparing levels from fall 2023 to spring 2025 by age and outcome for children attending as both 3-year-olds and 4-year-olds	F-7

# Appendix A

## Evaluation Methods

## Appendix A

### Evaluation Methods

Appendix A provides information on the measures used in the Pre-K 4 SA Year 12 evaluation and we describe details on the analytic approach to the analyses in the body of the report.

#### Measures

---

#### Classroom Assessment Scoring System (CLASS)

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 12 evaluation consisted of five 20-minute observation periods (or cycles)<sup>1</sup> followed by 10-minute coding periods. Observers assigned scores during various classroom activities. Westat computed averages across all cycles for an overall quality score.

Observations occurred during the spring of the 2024–25 school year. CLASS measures interactions 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 1 or 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 considered in the high-range and indicates high quality. Observers assigned each dimension and domain a score during each 20-minute cycle. They also recorded the number of children and adults in the classroom during each 20-minute cycle.

---

<sup>1</sup> There was a total of 56 (11.57%) irregular cycle lengths. For cycles below 20 minutes, 44 were 15 minutes, 3 were 17 minutes, 2 were 18 minutes, and 2 were 19 minutes. For cycles above 20 minutes, 2 were 21 minutes, 2 were 22 minutes, and 1 was 30 minutes.

**Table A.1. Descriptions of CLASS (second edition) dimensions**

Domain	Dimension	Description
<b>Emotional Support</b>	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.
<b>Classroom Organization</b>	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.
<b>Instructional Support</b>	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 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. Pre-K 4 SA conducts the GOLD assessment three times (fall, winter, and spring) across the school year. We used the revised norms for birth through third grade as a comparison to Pre-K 4 SA children.

**The Woodcock-Johnson Achievement Test (WJ)**

The 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). Early childhood research uses this measure in numerous large-scale preschool studies (e.g., Early et al., 2007; Wong et al., 2008). For children requiring assessment in Spanish, assessors used the matching subtests from the Bateria III (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, we provide descriptive information for both subtests based on the language of assessment 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, 9 months) than children assessed in Spanish (6 years, 2 months).

Subtest	English		Spanish	
	Minimum	Maximum	Minimum	Maximum
Letter-Word	2 years, 4 months	6 years, 9 months	4 years, 1 month	6 years, 2 months
Applied Problems	2 years, 0 months	8 years, 3 months	2 years, 9 months	6 years, 1 month

## Receptive One-Word Picture Vocabulary Test (ROWPVT) and Expressive One-Word Picture Vocabulary Test (EOWPVT)

The ROWPVT, fourth edition (Martin, 2013b; Martin & Brownell, 2011b), is a norm-referenced test of receptive vocabulary, and the EOWPVT, fourth edition (Martin, 2013a; Martin & Brownell, 2011a), is a norm-referenced test of expressive vocabulary. Both measures are available in English and Spanish (Spanish-Bilingual edition). ROWPVT and EOWPVT have established overall reliability and test-retest reliability. As shown in Table A-3, internal consistency and test-retest reliability are above .80, which is the suggested standard (Nunnally & Bernstein, 1994). The construct validity correlations range from .35 to .95 when compared with other vocabulary measures; the lower end of this range was based on a correlation with the Wechsler Intelligence Scale for Children, fourth edition, Verbal Comprehension Index (Wechsler, 2003). There are some task differences between the two compared assessments that may contribute to the lower correlation; however, both assessments assess similar abilities. These measures are both valid and reliable for ages ranging from 2 years to over 90 (Martin, 2013a, 2013b; Martin & Brownell, 2011a, 2011b). For the ROWPVT administration, the assessor presents the child with pictorial images of words and the child must select the image that matches the word said by the assessor. The items are set in difficulty order, with the easiest first and the most difficult last. Testing stops in the English version when the child scores zero on six out of eight successive items, and in the Spanish-Bilingual version when the child scores zero on four out of six successive items. For the EOWPVT administration, the assessor presents the child with one image at a time and asked to name the image shown. Items are set in difficulty order, with the easiest first and the most difficult last. In both the English and Spanish-Bilingual versions, testing stops when the child scores zero on six successive items.

	ROWPVT (Receptive)		EOWPVT (Expressive)	
	English	Spanish-Bilingual	English	Spanish-Bilingual
<b>Internal consistency<sup>a</sup></b>	.97	.95	.95	.95
<b>Test-retest<sup>b</sup></b>	.91	.91	.97	.97

<sup>a</sup> Findings are based on the median across age groups.

<sup>b</sup> Findings are based on standard scores.

## Devereux Early Childhood Assessment (DECA)

The DECA, second edition (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 reported good reliability evidence (Center for Resilient Children, 2013; LeBuffe & Naglieri, 1999). As shown in Table A-4, 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.4. 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 was 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.

## Analytic Approach

We addressed research questions by analyzing study data as well as existing Pre-K 4 SA databases. To protect against the disclosure of confidential data, we mask all results based on less than 10 children or percentages that round to 0 percent or 100 percent. Where needed, we mask additional results or include rounded percentages. 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 submitted data 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 12?), we descriptively analyzed CLASS observation data. To answer research question 2B (Did master teachers of Pre-K 4 SA classrooms have higher observed teacher–child interaction quality in Year 12?), we conducted an independent samples *t* test between master and non-master teachers in Pre-K 4 SA classrooms for each domain and dimension. We applied the Benjamini-Hochberg (1995) technique 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?), Pre-K 4 SA collected and submitted data to Westat. We combined data across fall, winter, and spring assessments for descriptive and inferential analyses. To answer question 3A, we conducted an independent samples *t* test between Pre-K 4 SA and the normative sample in the fall, winter, and spring for each outcome to determine if there were significant differences. To answer question 3B, we conducted a dependent sample *t* test between the fall and spring for each GOLD assessment outcome to determine if there was significant improvement. To answer question 3C, we conducted descriptive statistics for the spring assessment time to determine the percentage of 4-year-old children who were ready for kindergarten for each GOLD outcome. We applied the Benjamini-Hochberg (1995) technique to correct for multiple hypothesis testing.

Because children can attend Pre-K 4 SA for 2 consecutive years as 3-year-olds and 4-year olds, we conducted inferential analyses to determine patterns and trends in children’s GOLD outcomes over time. We included children in outcome analyses with data in the fall, winter, and spring assessments for both school years (2023–24 and 2024–25). We used the same methods for this subgroup of children as to answer research question 3A, 3B, and 3C. To answer question 3A, we conducted an independent samples *t* test between Pre-K 4 SA and the normative sample in the fall, winter, and spring for each outcome to determine if there were significant differences. To answer question 3B,

we conducted a dependent sample *t* test between the fall and spring for each GOLD assessment outcome to determine if there was significant improvement. To answer question 3C, we conducted descriptive statistics for the spring assessment time to determine the percentage of 4-year-old children who were ready for kindergarten for each GOLD outcome. We applied the Benjamini-Hochberg (1995) technique to correct for multiple hypothesis testing.

We conducted demographic tests of differences to determine if the sample of children included in and excluded from GOLD analyses were similar. We found no significant differences between children included and not included in analyses for age ( $\chi^2(1, N = 2,041) = 1.0723, p = .300$ ), gender ( $\chi^2(1, N = 2,041) = 0.8263, p = .363$ ), and eligibility to attend based on income or paying tuition ( $\chi^2(1, N = 2,041) = 2.2423, p = .134$ ). There were, however, significant differences between children included in and excluded from analyses for other demographic characteristics. We found a significant differences based on race and ethnicity<sup>2</sup> ( $\chi^2(5, N = 2,041) = 17.470, p = .004$ ), eligibility based on English-language learner status ( $\chi^2(1, N = 2,041) = 4.1079, p = .043$ ), and family military affiliation status ( $\chi^2(1, N = 2,041) = 4.2330, p = .040$ ). Children included in analyses were more likely to be Hispanic than children excluded from analyses. Children with an English-language learner status were less likely to be included in the analysis than children without an English-language learner status. Children with a family military affiliation were less likely to be included in analysis than children without a family military affiliation. This implies the GOLD findings are not representative of all children in the education centers.

We conducted demographic tests of differences to determine if the sample of children included and excluded from the children attending as both 3-year-olds and 4-year-olds (for two school year GOLD analyses) were similar. We found no significant differences between children included and not included in analyses for gender ( $\chi^2(1, N = 759) = 0.281, p = .596$ ), race and ethnicity<sup>3</sup> ( $\chi^2(5, N = 759) = 10.374, p = .065$ ), family military affiliation status ( $\chi^2(1, N = 759) = 0.556, p = .456$ ), and eligibility to attend based on income or paying tuition ( $\chi^2(1, N = 759) = 0.025, p = .876$ ). There were, however, significant differences between children included in and excluded from analyses for other demographic characteristics. We found a significant difference based on eligibility based on English-language learner status ( $\chi^2(1, N = 759) = 6.513, p = .012$ ). Children with an English-language learner status were less likely to be included in the analysis than children without an English-language learner status. This implies the two school year GOLD findings are not representative of all children in the education centers.

To address the fourth set of research questions (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?; 4B: Did Pre-K 4 SA children demonstrate significant improvement in early literacy and early numeracy?; and 4C: Did Pre-K 4 SA children experience accelerated learning to help narrow achievement gaps in early literacy and early numeracy?), Pre-K 4 SA collected and submitted data from a random sample<sup>4</sup> to Westat. We descriptively and inferentially analyzed the data. 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, we included children with data in both the fall and spring in an outcome analysis. We converted raw scores into

---

<sup>2</sup> Due to low cell counts, we conducted a Fisher's exact test for race and ethnicity in addition to the chi-square test. Both the chi-square and Fisher's exact test found a statistically significant difference based on race and ethnicity.

<sup>3</sup> Due to low cell counts, we conducted a Fisher's exact test for race and ethnicity in addition to the chi-square test. Both the chi-square and Fisher's exact test found no significant difference based on race and ethnicity.

<sup>4</sup> We found no significant differences between the random sample and the population for the following variables: gender, classroom type, center, age group, and strata.

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).<sup>5</sup> 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. Because 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, we conducted descriptive analyses for each outcome. To answer the second part, we computed the percentages below, on, and above age level for each outcome. We created a binary indicator for each outcome to determine if a child's assessed age level was either below or on and above their actual age level. We conducted analyses based on the binary indicators: a McNemar test 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, we conducted dependent *t* tests for each outcome between the fall and spring to determine if there were significant improvements over time. For research question 4C, we computed the difference between each child's assessed age and biological age for each outcome in the fall and spring. We used this difference 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. We conducted a difference of differences analysis to determine if there was a significant reduction of the achievement gap from the fall to spring. We applied the Benjamini-Hochberg (1995) technique to correct for multiple hypothesis testing.

To address the fifth set of research questions (5A: What are the receptive and expressive vocabulary performance levels of children, and to what extent do the performance levels change over the year?; and 5B: Did children demonstrate significant growth in receptive and expressive vocabulary?) Pre-K 4 SA collected and submitted data from a random sample<sup>6</sup> to Westat. We descriptively and inferentially analyzed the data. We included children with data in both the fall and spring for each assessment in the analysis. We converted raw scores into performance levels (J. Reeder, personal communication, August 28, 2024) and standard scores based on the norms provided in the technical manuals (Martin, 2013a, 2013b; Martin & Brownell, 2011a, 2011b). To address the first part of research question 5A, we conducted descriptive analyses to determine the categorical levels for each type of assessment (ROWPVT and EOWPVT). To address the second part of research question 5A, we conducted an inferential analysis using a Wilcoxon signed-rank test to determine if there were significant differences across the performance levels over time. To address research question 5B, we conducted dependent *t* tests for the standard scores between the fall and spring to determine if there were significant increases over time. We applied the Benjamini-Hochberg (1995) technique to correct for multiple hypothesis testing.

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?), Pre-K 4 SA collected and submitted data to Westat. We descriptively and inferentially analyzed the data. For any research question in the sixth set, we included children with data in both the fall and spring assessments for each outcome analysis. We converted *T* scores into three categorical levels—Needs Instruction, Typical, and Strengths—based on the technical manual (LeBuffe & Naglieri, 2012). To

---

<sup>5</sup> The age norms differ depending on whether a child was assessed in English or Spanish.

<sup>6</sup> We found no significant differences between the random sample and the population for the following variables: gender, classroom type, center, age group, and strata.

address the first part of research question 6A, we conducted descriptive analyses for the categorical levels for each assessment point. To address the second part of research question 6A, we conducted a Wilcoxon signed-rank test 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. Because Behavioral Concerns is binary, we conducted a McNemar test between the fall and spring. To address research question 6B, we conducted dependent *t* tests 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. We applied the Benjamini-Hochberg (1995) technique to correct for multiple hypothesis testing.

As this was the second year of children attending Pre-K 4 SA for two consecutive years as 3-year-olds and then 4-year olds, we conducted inferential analyses to determine patterns and trends in children’s social-emotional competency over time. We included children with data in the fall and spring assessments for both school years (2023–24 and 2024–25) in each outcome analysis. For this subgroup of children, we used the same methods used to answer research question 6A and 6B. We converted *T* scores 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, we conducted descriptive analyses for the categorical levels for each assessment point. To address the second part of research question 6A, we conducted a Wilcoxon signed-rank test between the fall 2023 and spring 2025 for each outcome except Behavioral Concerns to determine if there were significant differences in the distribution of categories over time. Because Behavioral Concerns is binary, we conducted a McNemar test between the fall 2023 and spring 2025. To address research question 6B, we conducted dependent *t* tests for the *T* scores for all but one outcome between the fall 2023 and spring 2025 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. We applied the Benjamini-Hochberg (1995) technique to correct for multiple hypothesis testing. We also compared these longitudinal two year findings to the 2024–25 results to determine if children in their first year of Pre-K 4 SA demonstrated similar patterns as children attending for two years.

We conducted demographic tests of differences to determine if the sample of children included in and excluded from DECA analyses were similar. We found no significant differences between children included in and excluded from analyses for gender ( $\chi^2(1, N = 2,041) = 0.339, p = .561$ ), English-language learner status ( $\chi^2(1, N = 2,041) = 3.304, p = .069$ ), family military affiliation status ( $\chi^2(1, N = 2,041) = 1.559, p = .212$ ), or age ( $\chi^2(1, N = 2,041) = 0.962, p = .327$ ). There were, however, significant differences between children included in and excluded from analyses for other demographic characteristics. We found a significant difference between children included in and excluded from analyses for indicators of race and ethnicity<sup>7</sup> ( $\chi^2(5, N = 2,041) = 24.435, p < .001$ ), eligibility to attend based on income or paying tuition ( $\chi^2(1, N = 2,041) = 3.859, p = .049$ ). Children included in analyses were more likely to have parents or guardians who paid tuition than children excluded from analyses. Children included in analyses were more likely to be of Hispanic or Black race/ethnicity. This implies the DECA findings are not representative of all children in the education centers.

We conducted demographic tests of differences to determine if the sample of children included and excluded from the children attending as both 3-year-olds and 4-year-olds (for two school year DECA

---

<sup>7</sup> Due to low cell counts, we conducted a Fisher’s exact test for race in addition to the chi-square test. Both the chi-square and Fisher’s exact test found a statistically significant difference based on race.

analyses) were similar. We found no significant differences between children included in and excluded from analyses for gender ( $\chi^2(1, N = 759) = 0.067, p = .796$ ), English-language learner status ( $\chi^2(1, N = 759) = 2.027, p = .155$ ), family military affiliation status ( $\chi^2(1, N = 759) = 0.584, p = .445$ ), eligibility to attend based on income or paying tuition ( $\chi^2(1, N = 759) = 2.312, p = 0.128$ ), or indicators of race and ethnicity<sup>8</sup> ( $\chi^2(5, N = 759) = 9.559, p = .089$ ). This implies the findings are representative of all children in the education centers.

---

<sup>8</sup> Due to low cell counts, we conducted a Fisher's exact test for race in addition to the chi-square test. Both the chi-square and Fisher's exact test did not find a statistically significant difference based on race.

# 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. Across the four education centers, observed evaluated nearly all Pre-K 4 SA classrooms ( $n = 97$ ) during Year 12.<sup>9</sup> 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 12?) by providing descriptives for each center. We provide the overall results in the technical report in Table 4. 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 12?).

---

<sup>9</sup> CLASS data were available on 97 of the 100 classrooms in Year 12. For one classroom, both teachers in the pair left Pre-K 4 SA and were not able to be observed. For the other two classrooms, one teacher in each pair was on leave and not able to be observed.

**Table B.1. Average Year 12 CLASS scores by center (N = 97)**

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
<b>Emotional Support</b>	6.56 (0.32)	(5.95–7.00)	6.51 (0.47)	(5.38–7.00)	6.60 (0.27)	(6.18–7.00)	6.70 (0.23)	(6.25–7.00)
Positive climate	6.64 (0.42)	(5.60–7.00)	6.59 (0.44)	(5.40–7.00)	6.65 (0.32)	(6.00–7.00)	6.66 (0.35)	(6.00–7.00)
Negative climate <sup>a</sup>	6.95 (0.12)	(6.60–7.00)	6.95 (0.11)	(6.60–7.00)	6.95 (0.10)	(6.60–7.00)	6.98 (0.09)	(6.60–7.00)
Educator sensitivity	6.15 (0.81)	(4.00–7.00)	6.13 (0.84)	(4.50–7.00)	6.36 (0.55)	(5.50–7.00)	6.53 (0.36)	(6.00–7.00)
Regard for child perspectives	6.50 (0.42)	(5.60–7.00)	6.37 (0.64)	(4.80–7.00)	6.46 (0.44)	(5.60–7.00)	6.65 (0.52)	(5.40–7.00)
<b>Classroom Organization</b>	6.42 (0.42)	(5.67–7.00)	6.50 (0.50)	(5.20–7.00)	6.41 (0.32)	(5.87–7.00)	6.33 (0.41)	(5.67–7.00)
Behavior management	6.39 (0.60)	(5.20–7.00)	6.52 (0.70)	(4.20–7.00)	6.47 (0.55)	(4.80–7.00)	6.51 (0.50)	(5.40–7.00)
Productivity	6.66 (0.35)	(5.80–7.00)	6.66 (0.41)	(5.60–7.00)	6.69 (0.31)	(6.00–7.00)	6.45 (0.41)	(5.60–7.00)
Instructional learning formats	6.21 (0.57)	(5.20–7.00)	6.30 (0.55)	(5.00–7.00)	6.06 (0.53)	(5.00–7.00)	6.04 (0.68)	(4.60–7.00)
<b>Instructional Support</b>	3.96 (0.84)	(2.47–5.47)	4.10 (0.89)	(2.40–5.87)	3.79 (0.71)	(2.60–5.13)	3.82 (0.81)	(2.73–5.73)
Concept development	3.31 (1.05)	(1.60–5.00)	3.47 (1.25)	(1.20–5.80)	3.14 (0.77)	(2.20–5.00)	3.21 (0.99)	(1.60–5.20)
Quality of feedback	4.02 (1.08)	(2.20–6.40)	4.24 (1.14)	(2.40–6.40)	3.90 (0.97)	(2.40–6.00)	3.95 (0.97)	(2.60–6.20)
Language modeling	4.55 (0.90)	(2.60–5.80)	4.57 (0.74)	(3.00–6.00)	4.34 (0.79)	(2.80–5.60)	4.31 (0.82)	(3.00–6.00)

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

M = mean; SD = standard deviation.

<sup>a</sup> 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 <sup>a</sup>
<b>Emotional Support</b>	97	6.57	6.62	0.05	-0.713	85.23	0.4777	Not significant	No difference
Positive climate	97	6.66	6.61	-0.05	0.716	93.22	0.4761	Not significant	No difference
Negative climate	97	6.94	6.98	0.04	-1.795	78.45	0.0765	Not significant	No difference
Educator sensitivity	97	6.22	6.37	0.15	-1.081	89.84	0.2827	Not significant	No difference
Regard for child perspectives	97	6.46	6.53	0.07	-0.611	89.64	0.5428	Not significant	No difference
<b>Classroom Organization</b>	97	6.41	6.41	0.00	0.015	94.09	0.9880	Not significant	No difference
Behavior management	97	6.46	6.49	0.03	-0.230	89.98	0.8190	Not significant	No difference
Productivity	97	6.61	6.62	0.01	-0.158	94.95	0.8751	Not significant	No difference
Instructional learning formats	97	6.17	6.13	-0.04	0.362	94.89	0.7185	Not significant	No difference
<b>Instructional Support</b>	97	3.85	3.98	0.13	-0.767	91.20	0.4451	Not significant	No difference
Concept development	97	3.22	3.34	0.12	-0.559	94.84	0.5777	Not significant	No difference
Quality of feedback	97	3.97	4.08	0.11	-0.535	93.91	0.5936	Not significant	No difference
Language modeling	97	4.37	4.52	0.15	-0.920	94.46	0.3600	Not significant	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 hundredth.

<sup>a</sup> If there was a statically significant difference, this column lists the group with the greater score (i.e., the “favored” group). If there was no statistically significant difference, this column states there was “no difference.”

# Appendix C

Additional Growth, Observation, and  
Learning (GOLD) Results

## Appendix C

### Additional Growth, Observation, and Learning (GOLD) Results

Appendix C provides additional GOLD results that address the third set of research questions. Tables C.1 through C.4 present findings for the 2024–25 school year. Tables C.5 through C.7 present findings for children who attended for two school years (2023–24 and 2024–25) when they were 3- and 4-year-olds.

Tables C.1, C.2, and C.5 address research question 3A (How did Pre-K 4 SA children compare to the normative sample on GOLD outcomes?). We present the results separately for 3-year-old children and 4-year-old children as the norms vary depending on age level. Tables C.1 and C.2 display the findings based on the 2024–25 school year. Table C.5 displays the findings for children who attended for two school years (2023–24 and 2024–25) when they were 3- and 4-year-olds. Because of the complexity of the data for each GOLD outcome, these tables are split into two related subtables for Tables C.1 and C.2 (i.e., Tables C.1a, C.1b, C.2a, and C.2b) and three related subtables for Table C.5 (Table C.5a, C.5b, and C.5c).

Tables C.3 and C.6 address research question 4B (Did Pre-K 4 SA children demonstrate significant improvement on GOLD outcomes?). Table C.3 displays the findings based on the 2024–25 school year. Table C.6 displays the findings for children who attended for two school years (2023–24 and 2024–25) when they were 3- and 4-year-olds.

Tables C.4 and C.7 address research question 3C (What percentage of Pre-K 4 SA children demonstrated kindergarten readiness as measured by GOLD outcomes?). Table C.4 displays the findings based on the 2024–25 school year. Table C.7 displays the findings for children who attended for two school years (2023–24 and 2024–25) when they were 3- and 4-year-olds.

**Table C.1a. Pre-K 4 SA (3-year-old) and normative sample comparison results for Cognitive, Literacy, and Mathematics 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 <sup>a</sup>	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normative sample)
Cognitive	Fall	378.76	374.62	4.14	2.413	1983.359	0.0159	Significant	Pre-K 4 SA	
	Winter	413.43	415.58	-2.15	-1.275	1910.984	0.2026	Not significant	No difference	
	Spring	436.87	444.42	-7.55	-4.202	1949.772	<.0001	Significant	Normed	
Literacy	Fall	435.54	428.30	7.24	4.095	1646.453	<.0001	Significant	Pre-K 4 SA	
	Winter	459.88	461.74	-1.86	-0.939	1212.629	0.3479	Not significant	No difference	
	Spring	476.36	481.08	-4.72	-2.267	1179.412	0.0236	Significant	Normed	
Mathematics	Fall	299.50	290.08	9.42	5.692	1985.071	<.0001	Significant	Pre-K 4 SA	
	Winter	334.53	327.62	6.91	4.695	1942.759	<.0001	Significant	Pre-K 4 SA	
	Spring	357.24	352.05	5.19	3.548	1966.066	0.0004	Significant	Pre-K 4 SA	

**Note:** We present group mean information in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a statically significant difference, this column lists the group with the greater score (i.e., the “favored” group). If there was no statistically significant difference, this column states there was “no difference.”

**Table C.1b. Pre-K 4 SA (3-year-old) and normative sample comparison results for Oral Language, Physical, and Social-Emotional 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 <sup>a</sup>	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normative sample)
Oral Language	Fall	403.50	402.14	1.36	0.521	1349.870	0.6025	Not significant	No difference	
	Winter	437.73	442.82	-5.09	-1.853	1264.268	0.0640	Not significant	No difference	
	Spring	461.77	474.07	-12.30	-4.704	1150.859	<.0001	Significant	Normed	
Physical	Fall	495.38	496.07	-0.69	-0.357	1924.264	0.7212	Not significant	No difference	
	Winter	530.69	537.37	-6.68	-3.704	1859.219	0.0002	Significant	Normed	
	Spring	551.96	567.59	-15.63	-8.021	1841.065	<.0001	Significant	Normed	
Social-Emotional	Fall	376.89	379.88	-2.99	-1.893	1917.322	0.0584	Not significant	No difference	
	Winter	411.11	416.85	-5.74	-3.800	1904.131	0.0001	Significant	Normed	
	Spring	432.29	442.81	-10.52	-6.419	1972.120	<.0001	Significant	Normed	

**Note:** We present group mean information in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a statically significant difference, this column lists the group with the greater score (i.e., the “favored” group). If there was no statistically significant difference, this column states there was “no difference.”

**Table C.2a. Pre-K 4 SA (new 4-year-old) and normative sample comparison results for Cognitive, Literacy, and Mathematics 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 <sup>a</sup>	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normative sample)
Cognitive	Fall	428.81	427.94	0.87	0.273	255.958	0.7853	Not significant	No difference	
	Winter	467.33	472.29	-4.96	-1.597	251.373	0.1116	Not significant	No difference	
	Spring	495.08	506.46	-11.38	-3.220	249.697	0.0015	Significant	Normed	
Literacy	Fall	475.56	474.47	1.09	0.442	221.070	0.6586	Not significant	No difference	
	Winter	498.31	506.52	-8.21	-2.416	195.441	0.0166	Significant	Normed	
	Spring	513.35	527.01	-13.66	-2.980	191.722	0.0033	Significant	Normed	
Mathematics	Fall	340.34	341.41	-1.07	-0.363	248.545	0.7170	Not significant	No difference	
	Winter	378.28	379.14	-0.86	-0.344	251.229	0.7312	Not significant	No difference	
	Spring	401.33	405.25	-3.92	-1.466	253.029	0.1438	Not significant	No difference	

**Note:** We present group mean information in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a significant difference after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995), this column lists the group whose score was greater (i.e., the “favored” group). If there was no statistically significant difference, this column states there was “no difference.”

**Table C.2b. Pre-K 4 SA (new 4-year-old) and normative sample comparison results for Oral Language, Physical, and Social-Emotional 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 <sup>a</sup>	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normative sample)
Oral Language	Fall	455.07	461.29	-6.22	-1.104	195.650	0.2710	Not significant	No difference	
	Winter	490.71	507.14	-16.43	-2.479	191.797	0.0140	Significant	Normed	
	Spring	521.77	544.96	-23.19	-3.615	194.542	0.0004	Significant	Normed	
Physical	Fall	542.28	547.91	-5.63	-1.557	249.597	0.1209	Not significant	No difference	
	Winter	585.65	593.33	-7.68	-2.345	250.075	0.0198	Significant	Normed	
	Spring	612.24	628.50	-16.26	-5.021	256.615	0.0000	Significant	Normed	
Social-Emotional	Fall	417.45	426.44	-9.00	-2.825	246.027	0.0051	Significant	Normed	
	Winter	454.97	466.51	-11.54	-3.982	250.652	0.0001	Significant	Normed	
	Spring	479.56	497.35	-17.80	-5.566	254.374	0.0000	Significant	Normed	

**Note:** We present group mean information in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a significant difference, this column lists the group whose score was greater (i.e., the “favored” group). If there was no statistically significant difference, this column states 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 2024 mean	Spring 2025 mean	Growth (spring 2025–fall 2024)	t-test statistic	df	Initial p value	Adjusted significance	Time favored <sup>a</sup>
3-year-old	Cognitive	951	378.76	436.87	58.11	-49.843	950	<.0001	Significant	Spring 2025
	Literacy	877	435.54	476.36	40.82	-28.483	876	<.0001	Significant	Spring 2025
	Mathematics	951	299.50	357.24	57.74	-54.555	950	<.0001	Significant	Spring 2025
	Oral language	872	403.50	461.77	58.27	-41.904	871	<.0001	Significant	Spring 2025
	Physical	952	495.38	551.96	56.58	-42.060	951	<.0001	Significant	Spring 2025
	Social-emotional	951	376.89	432.29	55.39	-51.232	950	<.0001	Significant	Spring 2025
New 4-year-old	Cognitive	218	428.81	495.08	66.27	-23.645	217	<.0001	Significant	Spring 2025
	Literacy	186	475.56	513.35	37.79	-9.460	185	<.0001	Significant	Spring 2025
	Mathematics	218	340.34	401.33	60.99	-31.424	217	<.0001	Significant	Spring 2025
	Oral language	184	455.07	521.77	66.70	-20.583	183	<.0001	Significant	Spring 2025
	Physical	218	542.28	612.24	69.96	-24.742	217	<.0001	Significant	Spring 2025
	Social-emotional	218	417.45	479.56	62.11	-25.141	217	<.0001	Significant	Spring 2025

**Note:** We present group mean information in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a significant difference after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995), this column lists the time with the higher score (i.e., the “favored” time). If there was no statistically significant difference, this column states there was “no difference.”

**Table C.4. Descriptives of kindergarten readiness by GOLD outcome and 4-year-old children**

Outcome	New 4-year-old children		All 4-year-old children	
	Sample size	Percentage ready for kindergarten	Sample size	Percentage ready for kindergarten
Cognitive	218	79.82	857	86.81
Literacy	186	64.52	772	73.83
Mathematics	218	69.27	857	76.66
Oral Language	184	78.80	768	84.11
Physical	218	86.24	857	92.07
Social-Emotional	218	80.28	858	87.18

**Table C.5a. Pre-K 4 SA and normative sample comparison results for Cognitive and Literacy GOLD outcomes across time for children attending as both 3-year-olds and 4-year-olds**

Outcome	Age	Assessment time	Pre-K 4 SA mean	Normed mean	Gap (Pre-K 4 SA–normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored <sup>a</sup>	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normative sample)
Cognitive	3-year-old	Fall	371.68	374.62	-2.94	-1.375	959.543	0.1695	Not significant	No difference	
		Winter	413.24	415.58	-2.34	-1.118	939.250	0.2639	Not significant	No difference	
		Spring	440.17	444.42	-4.25	-1.998	1000.246	0.0460	Not significant	No difference	
	4-year-old	Fall	451.58	427.94	23.64	12.590	1007.100	<.0001	Significant	Pre-K 4 SA	
		Winter	487.19	472.29	14.90	7.897	925.336	<.0001	Significant	Pre-K 4 SA	
		Spring	514.75	506.46	8.29	4.094	956.705	<.0001	Significant	Pre-K 4 SA	
Literacy	3-year-old	Fall	428.59	428.30	0.29	0.137	816.389	0.8910	Not significant	No difference	
		Winter	466.00	461.74	4.26	2.649	902.947	0.0082	Significant	Pre-K 4 SA	
		Spring	482.78	481.08	1.70	1.066	922.444	0.2865	Not significant	No difference	
	4-year-old	Fall	493.22	474.47	18.75	14.380	1099.939	<.0001	Significant	Pre-K 4 SA	
		Winter	512.13	506.52	5.61	4.420	825.413	<.0001	Significant	Pre-K 4 SA	
		Spring	526.52	527.01	-0.49	-0.358	830.075	0.7207	Not significant	No difference	

**Note:** We present group mean information in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995). *df* = degrees of freedom.

<sup>a</sup> If there was a statically significant difference, this column lists the group with the greater score (i.e., the “favored” group). If there was no statistically significant difference, this column states there was “no difference.”

**Table C.5b. Pre-K 4 SA and normative sample comparison results for Mathematics and Oral Language GOLD outcomes across time for children attending as both 3-year-olds and 4-year-olds**

Outcome	Age	Assessment time	Pre-K 4 SA mean	Normed mean	Gap (Pre-K 4 SA–normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored <sup>a</sup>	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normative sample)
Mathematics	3-year-old	Fall	296.32	290.08	6.24	3.019	958.407	0.0026	Significant	Pre-K 4 SA	
		Winter	334.94	327.62	7.32	4.115	978.022	<.0001	Significant	Pre-K 4 SA	
		Spring	359.94	352.05	7.89	4.559	1009.427	<.0001	Significant	Pre-K 4 SA	
	4-year-old	Fall	363.32	341.41	21.91	14.266	1046.581	<.0001	Significant	Pre-K 4 SA	
		Winter	393.77	379.14	14.63	10.466	1011.857	<.0001	Significant	Pre-K 4 SA	
		Spring	413.17	405.25	7.92	5.420	1060.335	<.0001	Significant	Pre-K 4 SA	
Oral Language	3-year-old	Fall	395.93	402.14	-6.21	-2.019	746.257	0.0439	Not significant	No difference	
		Winter	435.91	442.82	-6.91	-2.319	749.281	0.0207	Significant	Normed	
		Spring	464.61	474.07	-9.46	-2.942	655.758	0.0034	Significant	Normed	
	4-year-old	Fall	485.95	461.29	24.66	9.912	787.911	<.0001	Significant	Pre-K 4 SA	
		Winter	516.13	507.14	8.99	2.734	666.225	0.0064	Significant	Pre-K 4 SA	
		Spring	547.85	544.96	2.89	0.985	743.792	0.3248	Not significant	No difference	

**Note:** We present group mean information in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995). *df* = degrees of freedom.

<sup>a</sup> If there was a statically significant difference, this column lists the group with the greater score (i.e., the “favored” group). If there was no statistically significant difference, this column states there was “no difference.”

**Table C.5c. Pre-K 4 SA and normative sample comparison results for Physical and Social-Emotional GOLD outcomes across time for children attending as both 3-year-olds and 4-year-olds**

Outcome	Age	Assessment time	Pre-K 4 SA mean	Normed mean	Gap (Pre-K 4 SA–normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored <sup>a</sup>	Graphic depiction of finding (Blue line = Pre-K 4 SA; Orange line = normative sample)
Physical	3-year-old	Fall	483.90	496.07	-12.17	-5.113	958.770	<.0001	Significant	Normed	
		Winter	526.88	537.37	-10.49	-4.471	891.995	<.0001	Significant	Normed	
		Spring	556.96	567.59	-10.63	-4.616	968.118	<.0001	Significant	Normed	
	4-year-old	Fall	565.90	547.91	17.99	8.354	926.399	<.0001	Significant	Pre-K 4 SA	
		Winter	603.26	593.33	9.93	5.437	994.110	<.0001	Significant	Pre-K 4 SA	
		Spring	630.33	628.50	1.83	0.957	1020.703	0.3390	Not significant	No difference	
Social-Emotional	3-year-old	Fall	366.91	379.88	-12.97	-6.259	899.757	<.0001	Significant	Normed	
		Winter	407.07	416.85	-9.78	-4.774	874.710	<.0001	Significant	Normed	
		Spring	432.03	442.81	-10.78	-5.598	1022.660	<.0001	Significant	Normed	
	4-year-old	Fall	440.83	426.44	14.39	7.889	914.152	<.0001	Significant	Pre-K 4 SA	
		Winter	474.92	466.51	8.41	5.026	966.244	<.0001	Significant	Pre-K 4 SA	
		Spring	497.88	497.35	0.53	0.289	1020.668	0.7724	Not significant	No difference	

**Note:** We present group mean information in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995). *df* = degrees of freedom.

<sup>a</sup> If there was a statically significant difference, this column lists the group with the greater score (i.e., the “favored” group). If there was no statistically significant difference, this column states there was “no difference.”

**Table C.6. Growth results comparing fall 2023 and spring 2025 by GOLD outcome for children attending as both 3-year-olds and 4-year-olds**

Outcome	Sample size	3-year-old		4-year-old		Growth (spring 2025–fall 2023)	t-test statistic	df	Initial p value	Adjusted significance	Time favored <sup>a</sup>
		Fall 2023 mean	Spring 2024 mean	Fall 2024 mean	Spring 2025 mean						
Cognitive	608	371.68	440.17	451.58	514.75	143.07	-69.061	607	0.000	Significant	Spring 2025
Literacy	540	428.59	482.78	493.22	526.52	97.92	-56.552	539	0.000	Significant	Spring 2025
Mathematics	609	296.32	359.94	363.32	413.17	116.85	-67.003	608	0.000	Significant	Spring 2025
Oral Language	548	395.93	464.61	485.95	547.85	151.92	-56.996	547	0.000	Significant	Spring 2025
Physical	611	483.90	556.96	565.90	630.33	146.43	-65.694	610	0.000	Significant	Spring 2025
Social-Emotional	611	366.91	432.03	440.83	497.88	130.97	-69.671	610	0.000	Significant	Spring 2025

**Note:** We present group mean information in scaled scores. Because of rounding, decimals may not agree to the nearest hundredth. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a statically significant difference, this column lists the time with the higher score (i.e., the “favored” time). If there was no statistically significant difference, this column states there was “no difference.”

**Table C.7. Descriptives of kindergarten readiness by GOLD outcome for children attending as both 3-year-olds and 4-year-olds**

Outcome	Sample size	Fall 2023	Winter 2023	Spring 2024	Fall 2024	Winter 2024	Spring 2025
Cognitive	608	<1.00	<1.00	1.97	38.49	73.36	89.14
Literacy	540	<1.00	<1.00	1.30	22.22	59.07	77.59
Mathematics	609	<1.00	<1.00	1.64	16.26	61.58	78.98
Oral language	548	<1.00	1.46	2.37	46.53	72.45	86.13
Physical	611	<1.00	2.29	3.76	56.30	85.60	93.78
Social-emotional	611	<1.00	<1.00	1.80	42.06	77.91	89.53

**Note:** Because of masking, we provide a range for some results.

# Appendix D

Additional Woodcock-Johnson and  
Batería Results

## Appendix D

### Additional Woodcock-Johnson and Bateria Results

Appendix D provides additional direct child assessment results from the WJ and Bateria 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 Pre-K 4 SA children demonstrate significant improvement in early literacy and early numeracy?). Table D-3 addresses research question 4C (Did 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 2024 and spring 2025 by subtest and years with Pre-K 4 SA**

Outcome	Years with Pre-K 4 SA	Sample size	Fall 2024 percentage	Spring 2025 percentage	Difference (spring 2025–fall 2024)	$\chi^2$ statistic	df	Initial <i>p</i> value	Significance	Time favored <sup>a</sup>
Letter-Word	New children	39 <sup>b</sup>	46.2	38.5	-7.7	3.00	1	0.0833	Not significant	No difference
	Returning children	24	41.7	29.2	-12.5	3.00	1	0.0833	Not significant	No difference
	Total	63	44.4	34.9	-9.5	6.00	1	0.0143	Significant	Fall 2024
Applied Problems	New children	39	23.1	35.9	12.8	5.00	1	0.0253	Significant	Spring 2025
	Returning children	25	24.0	20.0	-4.0	0.33	1	0.5637	Not significant	No difference
	Total	64	23.4	29.7	6.2	2.00	1	0.1573	Not significant	No difference

**Note:** Because of rounding, decimals may not agree to the nearest tenth. Letter-Word measures early literacy skills, and Applied Problems measures early numeracy skills. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a statically significant difference, this column lists the time with the higher score (i.e., the “favored” time). If there was no statistically significant difference, this column states there was “no difference.”

<sup>b</sup> This sample included 34 3-year-old and 5 4-year-old children.

**Table D.2. Age equivalencies results comparing fall 2024 and spring 2025 by subtest and years with Pre-K 4 SA**

Outcome	Years with Pre-K 4 SA	Sample size	Fall 2024 mean	Spring 2025 mean	Growth (spring 2025–fall 2024)	t-test statistic	df	Initial p value	Adjusted significance	Time favored <sup>a</sup>
Letter-Word	New children	39 <sup>b</sup>	3 years, 9 months	4 years, 2 months	5 months	5.32	38	<.0001	Significant	Spring 2025
	Returning children	24	4 years, 4 months	4 years, 7 months	3 months	2.26	23	0.0333	Significant	Spring 2025
	Total	63	4 years, 0 months	4 years, 4 months	4 months	5.44	62	<.0001	Significant	Spring 2025
Applied Problems	New children	39	3 years, 3 months	3 years, 10 months	7 months	5.66	38	<.0001	Significant	Spring 2025
	Returning children	25	4 years, 0 months	4 years, 8 months	8 months	5.38	24	<.0001	Significant	Spring 2025
	Total	64	3 years, 6 months	4 years, 2 months	7 months <sup>c</sup>	7.81	63	<.0001	Significant	Spring 2025

**Note:** Letter-Word measures early literacy skills, and Applied Problems measures early numeracy skills. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a statically significant difference, this column lists the time with the higher score (i.e., the “favored” time). If there was no statistically significant difference, this column states there was “no difference.”

<sup>b</sup> This sample included 34 3-year-old and 5 4-year-old children.

<sup>c</sup> Because of rounding, the gap is not the exact difference between fall and spring.

**Table D.3. Gap analysis age equivalency results comparing fall 2024 and spring 2025 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 2024	Average gap between age equivalency and actual age, spring 2025	Difference <sup>a</sup> (spring 2025–fall 2024)	t-test statistic	df	Initial p value	Adjusted significance	Gap reduced <sup>b</sup>
Letter-Word	New children	39 <sup>c</sup>	-1 month	-1 month	-1 month <sup>d</sup>	-1.02	38	0.3121	Not significant	No change
	Returning children	24	-4 months	-6 months	-3 months <sup>d</sup>	-2.00	23	0.0578	Not significant	No change
	Total	63	-2 months	-3 months	-2 months <sup>d</sup>	-2.10	62	0.0396	Significant	Fall 2024
Applied Problems	New children	39	-7 months	-6 months	1 month	0.81	38	0.4255	Not significant	No change
	Returning children	25	-8 months	-6 months	2 months	1.37	24	0.1834	Not significant	No change
	Total	64	-7 months	-6 months	1 month	1.48	63	0.1429	Not significant	No change

**Note:** Letter-Word measures early literacy skills, and Applied Problems measures early numeracy skills. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> Negative differences indicate a decrease in student performance of education center children and positive differences indicate an improvement in student performance of education center children.

<sup>b</sup> If there was a statically significant difference, this indicates if the gap was reduced and the column lists the time with the smaller gap (i.e., the “favored” time). If there was no statistically significant difference, this column states that there was “no change.”

<sup>c</sup> This sample included 34 3-year-old and 5 4-year-old children.

<sup>d</sup> Because of rounding, the gap is not the exact difference between fall and spring.

# Appendix E

Additional Receptive and Expressive One-Word  
Picture Vocabulary Test (ROWPVT and EOWPVT)  
Results

## Appendix E

### Additional Receptive and Expressive One-Word Picture Vocabulary Test (ROWPVT and EOWPVT) Results

Appendix E provides additional direct child assessment results from the ROWPVT (Receptive) and EOWPVT (Expressive) to address the fifth set of research questions. There are two tables provided. Table E.1 addresses both parts of research question 5A (What were the receptive and expressive vocabulary performance levels of children, and to what extent did the performance levels change over the year?). Table E-2 addresses research question 5B (Did children demonstrate significant growth in receptive and expressive vocabulary?).

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

Outcome	Level	Years with Pre-K 4 SA	Sample size	Fall 2024 percentage	Spring 2025 percentage	Growth (spring 2024–fall 2025)	W statistic	Initial p value	Adjusted significance	Time favored <sup>a</sup>
Receptive	Well below expected	New children	42 <sup>b</sup>	<16.0	<8.0	<-8.0	4	0.0146	Significant	Spring 2025
	Below expected			<8.0	<11.0	<3.0				
	Expected			83.3	76.2	-7.1				
	Above expected			7.1	14.3	7.1				
	Well above expected			<3.0	<8.0	<5.0				
	Well below expected	Returning children	26	<10.0	<3.0	<-7.0	0	0.0719	Not significant	No difference
	Below expected			3.8	7.7	3.8				
	Expected			84.6	73.1	-11.5				
	Above expected			3.8	15.4	11.5				
	Well above expected			<6.0	<9.0	<3.0				
	Well below expected	Total	68	<11.0	<3.0	<-8.0	6	0.0022	Significant	Spring 2025
	Below expected			<10.0	<15.0	<5.0				
	Expected			83.8	75.0	-8.8				
	Above expected			5.9	14.7	8.8				
	Well above expected			1.5	4.4	2.9				

Outcome	Level	Years with Pre-K 4 SA	Sample size	Fall 2024 percentage	Spring 2025 percentage	Growth (spring 2024–fall 2025)	W statistic	Initial <i>p</i> value	Adjusted significance	Time favored <sup>a</sup>
Expressive	Well below expected	New children	41 <sup>c</sup>	<9.0	<3.0	<-6.0	16	0.0050	Significant	Spring 2025
	Below expected			12.2	7.3	-4.9				
	Expected			70.7	65.9	-4.9				
	Above expected			12.2	19.5	7.3				
	Well above expected			<6.0	<14.0	<8.0				
	Well below expected	Returning children	27	<8.0	<4.0	<-4.0	8	0.2986	Not significant	No difference
	Below expected			<11.0	<15.0	<4.0				
	Expected			<74.0	<78.0	<4.0				
	Above expected			14.8	18.5	3.7				
	Well above expected			<4.0	<8.0	<4.0				
	Well below expected	Total	68	<5.0	<2.0	<-3.0	46	0.0030	Significant	Spring 2025
	Below expected			10.3	7.4	-2.9				
	Expected			70.6	67.6	-2.9				
	Above expected			13.2	19.1	5.9				
	Well above expected			<6.0	<12.0	<6.0				

**Note:** Because of rounding, decimals may not agree to the nearest tenth, and percentages may sum to more than 100 percent. Because of masking, for some results we provide a range. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a statically significant difference, this column lists the time with the higher score (i.e., the “favored” time). If there was no statistically significant difference, this column states there was “no difference.”

<sup>b</sup> The receptive vocabulary sample included 35 3-year-old and 7 4-year-old children.

<sup>c</sup> The expressive vocabulary sample included 34 3-year-old and 7 4-year-old children.

**Table E.2. Vocabulary growth results comparing fall 2024 and spring 2025 by assessment and years with Pre-K 4 SA**

Outcome	Years with Pre-K 4 SA	Sample size	Fall 2024 mean	Spring 2025 mean	Growth (spring 2025–fall 2024)	t-test statistic	df	Initial p value	Adjusted significance	Time favored <sup>a</sup>
Receptive	New children	42 <sup>b</sup>	97.95	104.05	6.10	3.78	41	0.0010	Significant	Spring 2025
	Returning children	26	97.42	102.73	5.31	3.03	25	0.0060	Significant	Spring 2025
	Total	68	97.75	103.54	5.79	4.86	67	<.0001	Significant	Spring 2025
Expressive	New children	41 <sup>c</sup>	100.88	107.20	6.32	3.86	40	<.0001	Significant	Spring 2025
	Returning children	27	103.15	106.04	2.89	2.00	26	0.0570	Not significant	No difference
	Total	68	101.78	106.74	4.96	4.30	67	<.0001	Significant	Spring 2025

**Note:** Because of rounding, decimals may not agree to the nearest hundredth. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a statically significant difference, this column lists the time with the higher score (i.e., the “favored” time). If there was no statistically significant difference, this column states there was “no difference.”

<sup>b</sup> The receptive vocabulary sample included 35 3-year-old and 7 4-year-old children.

<sup>c</sup> The expressive vocabulary sample included 34 3-year-old and 7 4-year-old children.

# 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 2024–25 school year. Tables F.3 and F.4 present findings for children who attended for two school years (2023–24 and 2024–25) 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 2024–25 school year. Table F.3 displays the findings for children who attended for two school years (2023–24 and 2024–25) 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 2024–25 school year. Table F.4 displays the findings for children who attended for two school years (2023–24 and 2024–25) when they were 3- and 4-year-olds.

**Table F.1. Social-emotional results comparing levels in fall 2024 and spring 2025 by outcome and years with Pre-K 4 SA**

Outcome	Years with Pre-K 4 SA	Level	Sample size	Fall 2024 percentage	Spring 2025 percentage	Growth (spring 2025–fall 2024)	W statistic	Initial p value	Adjusted significance	Time favored <sup>a</sup>
Initiative	New 3-year-old children	Needs Instruction	833	31.21	19.93	-11.28	8706	<.0001	Significant	Spring 2025
		Typical		64.59	72.15	7.56				
		Strengths		4.20	7.92	3.72				
	New 4-year-old children	Needs Instruction	149	19.46	8.72	-10.74	76	<.0001	Significant	Spring 2025
		Typical		73.15	75.17	2.01				
		Strengths		7.38	16.11	8.72				
	Returning 4-year-old children	Needs Instruction	600	7.50	5.67	-1.83	4347	<.0001	Significant	Spring 2025
		Typical		75.17	69.83	-5.33				
		Strengths		17.33	24.50	7.17				
	Total	Needs Instruction	1,582	21.11	13.46	-7.65	28542	<.0001	Significant	Spring 2025
		Typical		69.41	71.55	2.15				
		Strengths		9.48	14.98	5.50				
Self-Control	New 3-year-old children	Needs Instruction	833	20.17	17.29	-2.88	7448	0.0014	Significant	Spring 2025
		Typical		71.67	72.03	0.36				
		Strengths		8.16	10.68	2.52				
	New 4-year-old children	Needs Instruction	149	16.11	10.74	-5.37	240	0.0167	Significant	Spring 2025
		Typical		70.47	71.14	0.67				
		Strengths		13.42	18.12	4.70				
	Returning 4-year-old children	Needs Instruction	600	10.83	9.67	-1.17	5542	0.0298	Significant	Spring 2025
		Typical		73.50	71.17	-2.33				

Outcome	Years with Pre-K 4 SA	Level	Sample size	Fall 2024 percentage	Spring 2025 percentage	Growth (spring 2025–fall 2024)	W statistic	Initial p value	Adjusted significance	Time favored <sup>a</sup>
	Total	Strengths	1,582	15.67	19.17	3.50	30966	<.0001	Significant	Spring 2025
		Needs Instruction		16.25	13.78	-2.47				
		Typical		72.25	71.62	-0.63				
		Strengths		11.50	14.60	3.10				
Attachment	New 3-year-old children	Needs Instruction	833	22.93	16.09	-6.84	7128	<.0001	Significant	Spring 2025
		Typical		69.39	72.15	2.76				
		Strengths		7.68	11.76	4.08				
	New 4-year-old children	Needs Instruction	149	20.81	15.44	-5.37	252	0.0081	Significant	Spring 2025
		Typical		68.46	67.79	-0.67				
		Strengths		10.74	16.78	6.04				
	Returning 4-year-old children	Needs Instruction	600	11.33	8.50	-2.83	4759	<.0001	Significant	Spring 2025
		Typical		72.67	69.00	-3.67				
		Strengths		16.00	22.50	6.50				
	Total	Needs Instruction	1,582	18.33	13.15	-5.18	28580	<.0001	Significant	Spring 2025
		Typical		70.54	70.54	0.00				
		Strengths		11.13	16.31	5.18				
Total Protective Factors	New 3-year-old children	Needs Instruction	833	26.41	18.37	-8.04	7988	<.0001	Significant	Spring 2025
		Typical		65.67	71.19	5.52				
		Strengths		7.92	10.44	2.52				
	New 4-year-old children	Needs Instruction	149	20.81	12.75	-8.05	220	0.0005	Significant	Spring 2025
		Typical		66.44	67.11	0.67				
		Strengths		12.75	20.13	7.38				

Outcome	Years with Pre-K 4 SA	Level	Sample size	Fall 2024 percentage	Spring 2025 percentage	Growth (spring 2025–fall 2024)	W statistic	Initial <i>p</i> value	Adjusted significance	Time favored <sup>a</sup>
	Returning children	Needs Instruction	600	9.83	7.00	-2.83	5074	<.0001	Significant	Spring 2025
		Typical		69.50	65.67	-3.83				
		Strengths		20.67	27.33	6.67				
	Total	Needs Instruction	1,582	19.60	13.53	-6.07	30730	<.0001	Significant	Spring 2025
		Typical		67.19	68.71	1.52				
		Strengths		13.21	17.76	4.55				
Behavioral Concerns <sup>c</sup>	New 3-year-old children	Typical	833	83.07	83.79	0.72	0	0.6396	Not Significant	No difference
		Area of Need		16.93	16.21	-0.72				
	New 4-year-old children	Typical	149	89.93	87.25	-2.68	1	0.4533	Not Significant	No difference
		Area of Need		10.07	12.75	2.68				
	Returning 4-year-old children	Typical	600	89.33	86.67	-2.67	3	0.0648	Not Significant	No difference
		Area of Need		10.67	13.33	2.67				
	Total	Typical	1,582	86.09	85.21	-0.88	1	0.3531	Not Significant	No difference
		Area of Need		13.91	14.79	0.88				

**Note:** Because of rounding, decimals may not agree to the nearest hundredth, and percentages may sum to more than 100 percent. Because of masking, we provide a range for some results. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a significant difference after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995), this column lists the time with the higher score (i.e., the “favored” time). If there was no statistically significant difference, this column states there was “no difference.”

<sup>b</sup> The finding was almost statistically significant indicating a significant increase over time.

<sup>c</sup> We conducted a McNemar’s test 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 2024 and spring 2025 by outcome and years with Pre-K 4 SA**

Outcome	Years with Pre-K 4 SA	Sample size	Fall 2024 mean	Spring 2025 mean	Growth (spring 2025–fall 2024)	t-test statistic	df	Initial p value	Adjusted significance	Time favored <sup>a</sup>
Initiative	New 3-year-old children	833	45.12	48.33	3.21	12.606	832	<.0001	Significant	Spring 2025
	New 4-year-old children	149	48.99	52.70	3.71	6.302	148	<.0001	Significant	Spring 2025
	Returning 4-year-old children	600	52.79	55.17	2.38	8.314	599	<.0001	Significant	Spring 2025
	Total	1,582	48.39	51.34	2.94	16.214	1,581	<.0001	Significant	Spring 2025
Self-Control	New 3-year-old children	833	47.28	48.89	1.61	6.899	832	<.0001	Significant	Spring 2025
	New 4-year-old children	149	49.74	51.88	2.14	3.603	148	0.0004	Significant	Spring 2025
	Returning 4-year-old children	600	51.47	52.70	1.23	4.471	599	<.0001	Significant	Spring 2025
	Total	1,582	49.10	50.61	1.51	8.883	1,581	<.0001	Significant	Spring 2025
Attachment	New 3-year-old children	833	46.79	49.32	2.53	9.919	832	<.0001	Significant	Spring 2025
	New 4-year-old children	149	48.70	51.09	2.40	3.961	148	0.0001	Significant	Spring 2025
	Returning 4-year-old children	600	51.49	52.72	1.24	3.772	599	0.0002	Significant	Spring 2025
	Total	1,582	48.75	50.77	2.03	10.547	1,581	<.0001	Significant	Spring 2025
Total Protective Factors	New 3-year-old children	833	45.84	48.68	2.84	12.168	832	<.0001	Significant	Spring 2025
	New 4-year-old children	149	49.09	52.27	3.18	5.710	148	<.0001	Significant	Spring 2025
	Returning 4-year-old children	600	52.27	54.07	1.80	6.241	599	<.0001	Significant	Spring 2025
	Total	1,582	48.58	51.06	2.48	14.316	1,581	<.0001	Significant	Spring 2025
Behavioral Concerns <sup>b</sup>	New 3-year-old children	833	50.25	49.82	-0.42	-1.861	832	0.0631	Not Significant	No difference
	New 4-year-old children	149	47.29	46.87	-0.42	-0.785	148	0.4337	Not Significant	No difference
	Returning 4-year-old children	600	47.93	47.27	-0.66	-2.521	599	0.0120	Significant	Spring 2025
	Total	1,582	49.09	48.58	-0.51	-3.135	1,581	0.0017	Significant	Spring 2025

**Note:** Because of rounding, decimals may not agree to the nearest hundredth. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom

<sup>a</sup> If there was a significant difference after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995), this column lists the time with the higher score (i.e., the “favored” time). If there was no statistically significant difference, this column states there was “no difference.”

<sup>b</sup> For the Behavioral Concerns outcome, negative growth means that behavioral issues decreased in spring 2025, indicating a positive finding.

**Table F.3. Social-emotional results comparing levels from fall 2023 to spring 2025 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 2025–fall 2023)	W statistic	Initial <i>p</i> value	Adjusted significance	Time favored <sup>a</sup>
			Fall 2023 percentage	Spring 2024 percentage	Fall 2024 percentage	Spring 2025 percentage					
Initiative	Needs Instruction	526	40.11	14.64	6.84	5.89	-34.22	1,295.0	<.0001	Significant	Spring 2025
	Typical		56.84	72.24	75.29	68.25	11.41				
	Strengths		3.04	13.12	17.87	25.86	22.81				
Self-Control	Needs Instruction	526	26.43	17.49	11.03	9.51	-16.92	3,045.0	<.0001	Significant	Spring 2025
	Typical		66.54	67.30	73.38	71.10	4.56				
	Strengths		7.03	15.21	15.59	19.39	12.36				
Attachment	Needs Instruction	526	26.05	7.79	11.22	8.37	-17.68	3,630.0	<.0001	Significant	Spring 2025
	Typical		<69.00	<74.00	<74.00	<71.00	<2.00				
	Strengths		<6.00	<20.00	<17.00	<24.00	<18.00				
Total Protective Factors	Needs Instruction	526	33.08	12.74	9.51	7.22	-25.86	2,590.0	<.0001	Significant	Spring 2025
	Typical		63.12	68.82	69.96	64.26	1.14				
	Strengths		3.80	18.44	20.53	28.52	24.71				
Behavioral Concerns <sup>b</sup>	Typical	526	79.47	81.56	89.54	87.07	7.60	20.0	<.0001	Significant	Spring 2025
	Area of Need		20.53	18.44	10.46	12.93	-7.60				

**Note:** Because of rounding, decimals may not agree to the nearest hundredth, and percentages may sum to more than 100 percent. Because of masking, for some results we provide a range. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a statically significant difference, this column lists the time with the higher score (i.e., the “favored” time). If there was no statistically significant difference, this column states there was “no difference.”

<sup>b</sup> We conducted a McNemar’s test 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 2023 to spring 2025 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 (spring 2025–fall 2023)	t-test statistic	df	Initial p value	Adjusted significance	Time favored <sup>a</sup>
		Fall 2023 mean	Spring 2024 mean	Fall 2024 mean	Spring 2025 mean						
Initiative	526	43.40	49.93	52.96	55.42	12.02	29.786	525	<.0001	Significant	Spring 2025
Self-Control	526	45.57	49.54	51.43	52.83	7.27	18.922	525	<.0001	Significant	Spring 2025
Attachment	526	46.24	52.16	51.66	53.01	6.77	15.258	525	<.0001	Significant	Spring 2025
Total Protective Factors	526	44.30	50.67	52.39	54.33	10.03	25.419	525	<.0001	Significant	Spring 2025
Behavioral Concerns <sup>b</sup>	526	51.32	49.81	47.97	47.18	-4.14	-11.010	525	<.0001	Significant	Spring 2025

**Note:** Because of rounding, decimals may not agree to the nearest hundredth. The Adjusted significance column indicates significance levels (*p* values) after adjustment to correct for multiple hypothesis testing (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a statically significant difference, this column lists the time with the higher score (i.e., the “favored” time). If there was no statistically significant difference, this column states there was “no difference.”

<sup>b</sup> For the Behavioral Concerns outcome, negative growth means that behavioral issues decreased in spring 2025, 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.
- 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* (2<sup>nd</sup> ed.). Center for Educational Measurement and Evaluation, University of North Carolina. [https://teachingstrategies.com/wp-content/uploads/2020/10/2020-Tech-Manual\\_GOLD.pdf](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 Manual (Spanish-Bilingual Edition)*. Academic Therapy Publications, Inc.
- Martin, N. A. (2013b). *Receptive One-Word Picture Vocabulary Test Manual (Spanish-Bilingual Edition)*. Academic Therapy Publications, Inc.
- Martin, N. A., & Brownell, R. (2011a). *Expressive One-Word Picture Vocabulary Test Manual*. Academic Therapy Publications, Inc.
- Martin, N. A., & Brownell, R. (2011b). *Receptive One-Word Picture Vocabulary Test Manual*. 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.
- 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.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). The Psychological Corporation.
- Wong, V. C., Cook, T. D., Barnett, W. S., & 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.