Pre-K 4 SA and Gardendale Early Learning Program:
Year 5 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

Table of Contents

Table of Contents (continued)

Appendix A

Evaluation Methods

Appendix A
Evaluation Methods

Appendix A provides additional information on the measures used in the Gardendale evaluation and details on the analytic approach used to answer the research questions.

Measures

Classroom Assessment Scoring System (CLASS)

CLASS, second edition, (Teachstone, 2023) is an observational system that assesses classroom practices by measuring the interactions between children and adults. Observations in the Gardendale 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 through 10 different dimensions (see Table A-1 for descriptions of each CLASS dimension) divided into 3 larger domains. The Emotional Support domain is measured through four dimensions: positive climate, negative climate, educator sensitivity, and regard for child perspectives. The Classroom Organization domain is measured through three dimensions: productivity, behavior management, and instructional learning formats. Finally, the Instructional Support domain is measured through 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 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 16 (20.12%) irregular cycle lengths observed. For the cycles less than 20 minutes, 7 were 15 minutes and 1 was 17 minutes. For the cycles longer than 20 minutes, 2 were 21 minutes, 3 were 22 minutes,1 was 23 minutes, 1 was 25 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) Assessment

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. (Other categories are available to be tailored to specific programs.) The GOLD assessment is conducted three times (fall, winter, and spring) throughout the school year. The new updated norms for birth through third grade were used as a comparison to Gardendale children.

The Woodcock-Johnson Achievement Test

The Woodcock-Johnson Tests of Achievement (WJ IV; Schrank et al., 2014) is an individually administered norm-referenced test to assess reading, oral language, mathematics, written language, and academic knowledge. We used two subtests in the current evaluation: the Letter-Word subtest and the Applied Problem subtest. Both subtests demonstrate excellent reliability (.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 Spanish for Letter-Word had a lower maximum level (10 years, 3 months) than children assessed in English (22 years, 0 months).

Table A-2.	Range of age equivalencies by assessment and language			
Subtest	**English**		**Spanish**	
	Minimum	**Maximum**	**Minimum**	**Maximum**
Letter-Word	3 years, 7 months	22 years, 0 months	4 years, 10 months	10 years, 3 months
Applied Problems	2 years, 2 months	8 years, 3 months	2 years, 0 months	8 years, 6 months

Peabody Picture Vocabulary Test (PPVT)

The Peabody Picture Vocabulary Test, fifth edition (Dunn & Dunn, 2019) is a norm-referenced 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. This measure has 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; therefore, 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]

Devereux Early Childhood Assessment (DECA) and Devereux Student Strengths Assessment (DESSA)

The Devereux Early Childhood Assessment (DECA; LeBuffe & Naglieri, 1999) and the Devereux Student Strengths Assessment (DESSA; LeBuffe et al., 2014) use teacher reports to measure children's social-emotional competencies in pre-K (DECA) and kindergarten through second grade (DESSA). Teachers report on the frequency of children's behavior on items constituting two main scales and three subscales in the DECA[3] and eight scales in the DESSA.[4] Both the DECA and DESSA are completed using a 5-point Likert-type scale to indicate how often within the past 4 weeks a child exhibited behaviors described by assessment items (0 = never, 1 = rarely, 2 = occasionally, 3 = frequently, and 4 = very frequently). In a standardization sample, the DECA was reported to have good reliability evidence (LeBuffe & Naglieri, 1999; Center for Resilient Children, 2013). 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 teacher ratings are substantial (McHugh, 2012).

Table A-3.	Reliability evidence for the Devereux Early Childhood Assessment (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

Measures of Academic Progress (MAP)

The Measures of Academic Progress (MAP) is a norm-referenced assessment of mathematics, reading, and science achievement (NWEA, 2023). It is administered electronically on an iPad.

---

[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.

[3] The two scales are Total Protective Factors and Behavioral Concerns. The three subscales are Initiative, Self-control, and Attachment.

[4] The eight scales are Self-Awareness, Social Awareness, Self-Management, Goal-Directed Behavior, Relationship Skills, Personal Responsibility, Decision Making, and Optimistic Thinking.

Children in Gardendale were assessed in English and Spanish three times (fall, winter, and spring) throughout the school year. For the mathematics assessment, children in second grade used text-to-speech, and all grade levels were allowed to use math manipulatives.

mCLASS

The mCLASS is an assessment of early literacy based on the Science of Reading. It assesses oral reading fluency based on a one-on-one observational model and measures phonemic awareness, phonics, fluency, vocabulary, and comprehension. Children in Gardendale were assessed in English and Spanish three times (fall, winter, and spring) throughout the school year. Children in kindergarten were given an online assessment while children in first and second grade were administered a reading booklet, and their teacher entered their responses into the data system (Biancarosa et al., 2021).

Analytic Approach

Research questions were addressed through an analysis of existing Pre-K 4 SA and Edgewood Independent School District databases and classroom observations. 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. Data collected by Gardendale and Pre-K 4 SA were submitted to Westat and descriptively analyzed to describe the sample of children in the evaluation. To address the first research question (What was the overall observed teacher–child interaction quality in Gardendale classrooms across Year 5?), CLASS observation data were descriptively analyses.

To address the second set of research questions (2A: How did pre-K and kindergarten Gardendale children compare to the normative sample on the GOLD outcomes?; 2B: Did pre-K and kindergarten Gardendale children demonstrate significant improvement on GOLD outcomes?; and 2C: What percentage of Gardendale children demonstrated kindergarten readiness as measured by GOLD outcomes?), data collected by Gardendale teachers were submitted to Westat. Data across the three assessment times (fall, winter, and spring) were combined for descriptive and inferential analyses. To answer question 2A, an independent-samples t test between Gardendale and the normative sample was conducted in the fall, winter, and spring for each outcome. To answer question 2B, a dependent-sample t test was conducted between the fall and spring for each outcome. The Benjamini-Hochberg (1995) technique was applied to correct for multiple hypothesis testing. To answer question 2C, descriptive statistics were conducted for the spring to determine the percentage of 4-year-old children who were ready for kindergarten at the end of the year, and for the fall to determine the percentage of kindergarten children who started the year ready for kindergarten.

As children were not randomly sampled, demographic tests of differences were conducted to determine if the sample of children included in and excluded from GOLD analyses were similar. No significant differences were found between children included in and excluded from analyses for grade level ($\chi^2(1) = 0.265$, $p = .607$), gender ($\chi^2(1) = 0.033$, $p = .856$), race ($\chi^2(2) = 0.941$, $p = .625$), Hispanic ethnicity ($\chi^2(1) = 0.769$, $p = .381$), receiving English Learner services ($\chi^2(1) = 0.824$, $p = .364$), or receiving special-education services $\chi^2(1) = 3.324$, $p = .068$). A significant difference was found between children included in and excluded from analyses for economically disadvantaged status ($\chi^2(1) = 7.440$, $p = .006$): Children with economically disadvantaged status were more likely to be included in analyses.

To address the third set of research questions (3A: What percentage of a random sample of Gardendale children performed at or above their age level in early literacy and early numeracy, and to what extent did the percentage change?; 3B: Did a random sample of Gardendale children demonstrate significant improvement in early literacy and early numeracy?; and 3C: Did a random sample of Gardendale 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 to determine and compare children's early literacy and early numeracy levels. For any research question in the third 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).[5] 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 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, only age equivalencies were used. A binary indicator was created for each outcome to determine if a child's assessed age level was either below or at/above their actual age level. To answer the first part of research question 3A, descriptive analyses were conducted for each outcome. To answer the second part, the percentages below, at, and above actual age level were computed and analyzed for each outcome. 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 were significant increases in the percentages over time. For research question 3B, dependent t tests were conducted for each outcome between the fall and spring to determine if there were significant increases over time. For research question 3C, 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 and denoted how many months they were above, below, or on par with the normative sample in their understanding of early literacy and early numeracy. A difference of differences analysis was conducted to determine if there was a significant reduction of the achievement gap from fall to spring. The Benjamini-Hochberg (1995) technique was applied to correct for multiple hypothesis testing.

To address the fourth set of research questions (4A: What were the receptive vocabulary performance levels of a random sample of Gardendale children's vocabulary?; 4B: Did a random sample of Gardendale children demonstrate significant improvement in receptive vocabulary?; and 4C: What types of receptive vocabulary improvement did a random sample of Gardendale children demonstrate?), data collected by Pre-K 4 SA were submitted to Westat and descriptively and inferentially analyzed. Children had to have data in both the fall and spring 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). To address research question 4A, descriptive analyses were conducted for the performance level description for each assessment time. To address research question 4B, 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 4C and aid the interpretation of changes in standard scores and growth scale values over time, score patterns were descriptively

---

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

analyzed according to the patterns detailed in the technical manual.[6] The Benjamini-Hochberg (1995) technique was applied to correct for multiple hypothesis testing.

To address the fifth set of research questions (5A: How did Gardendale children in kindergarten through second grade compare to the normative sample on MAP mathematics and reading?; 5B: Did Gardendale children in kindergarten through second grade demonstrate significant improvement on MAP mathematics and reading?; 5C. How did second-grade Gardendale children compare to the normative sample on MAP science?; and 5D: Did second-grade Gardendale children demonstrate significant improvement on MAP science?), data collected by Gardendale were submitted to Westat. Data across the three assessment times (fall, winter, and spring) were combined for descriptive and inferential analyses. Analyses were conducted separately by grade level as norms vary based on grade level. Children had to have data in the fall, winter, and spring to be included in analyses. To answer research question 5A, an independent-samples t test between Gardendale and the normative sample[7] was conducted in the fall, winter, and spring by grade level. To answer research question 5B, a dependent-sample t test was conducted between the fall and spring to determine if there was significant growth. To answer research question 5C, an independent-samples t test between Gardendale and the normative sample[8] was conducted in the fall, winter, and spring for second-grade children. To answer research question 5D, a dependent-sample t test was conducted between the fall and spring to determine if there was significant growth. 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 samples of children included in and excluded from analyses were similar. In the MAP math assessment, no significant differences were found between children included in and excluded from analyses for grade level ($\chi^2(2) = 3.059$, $p = .217$), gender ($\chi^2(1) = 0.052$, $p = .819$), race ($\chi^2(2) = 2.644$, $p = .267$), Hispanic ethnicity ($\chi^2(1) = 0.032$ $p = .859$), receiving English Learner services ($\chi^2(1) = 0.230$, $p = .632$), and receiving special-education services ($\chi^2(1) = 3.085$, $p = .079$). A significant difference was found between children included in and excluded from analyses for economically disadvantaged status ($\chi^2(1) = 20.205$, $p < .001$): Children with economically disadvantaged status were more likely to be included in analyses.

In the MAP reading assessment, no significant differences were found between children included in and excluded from analyses for gender ($\chi^2(1) = 0.007$, $p = .936$), race ($\chi^2(2) = 1.418$, $p = .492$), and Hispanic ethnicity ($\chi^2(1) = 0.411$, $p = .522$). A significant difference was found between children included in and excluded from analyses for grade level ($\chi^2(2) = 6.420$, $p = .040$): Children in first grade and second grade were more likely to be included in analyses. A significant difference was found between children included in and excluded from analyses for economically disadvantaged status ($\chi^2(1) = 15.091$, $p < .001$): Children with economically disadvantaged status were more likely to be included in analyses. A significant difference was found between children included in and excluded from analyses for receiving special-education services ($\chi^2(1) = 8.511$, $p = .004$): Children receiving special-education services were more likely to be included in analyses. A significant difference was found between children included in and excluded from analyses for receiving English

---

[6] For the category "Standard score does not change, growth scale value increases very little," a value of 6 was used to quantify very little increase for the growth scale value increase as no exact numerical value was provided in the technical manual.

[7] Norms were provided in the technical manual (Thum & Kuhfeld, 2020).

[8] Norms were provided in the technical manual (Thum & Kuhfeld, 2020).

Learner services ($\chi^2(1)$ = 90.796, $p < .001$): Children who were not receiving English Learner services were more likely to be included in analyses.

In the MAP science assessment, no significant differences were found between children included in and excluded from analyses for gender ($\chi^2(1)$ = 0.036, $p = .851$), race ($\chi^2(2)$ = 2.919, $p = .232$), Hispanic ethnicity ($\chi^2(2)$ = 1.137, $p = .286$), economically disadvantaged status ($\chi^2(1)$ = 0.394, $p = .530$), and receiving special-education services ($\chi^2(1)$ = 1.249, $p = .264$). A significant difference was found between children included in and excluded from analyses for receiving English Learner services ($\chi^2(1)$ = 20.661, $p < .001$): Children receiving English Learner services were more likely to be excluded from analyses.

To address the sixth set of research questions (6A: What were the performance levels of Gardendale children in kindergarten through second grade in mCLASS literacy?; and 6B: Did Gardendale children in kindergarten through second grade demonstrate significant improvement in mCLASS literacy?), data collected by Gardendale were submitted to Westat. Data across the three assessment times (fall, winter, and spring) were combined for descriptive and inferential analyses. Analyses were conducted separately by language of assessment as the norms vary based on language. Children had to have data in the fall, winter, and spring to be included in an outcome analysis. To address research question 6A, a Wilcoxon signed-rank test was conducted between the fall and spring based on language of assessment to determine if there were significant differences in the categories over time. To answer research question 6B, a dependent-sample *t* test was conducted between the fall and spring based on language of assessment and grade level to determine if there was significant improvement over the year. 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 samples of children included in and excluded from analyses were similar. No significant differences were found between children included in and excluded from analyses for grade level ($\chi^2(2)$ = 3.058, $p = .217$), gender ($\chi^2(1)$ = 0.002, $p = .965$), race ($\chi^2(2)$ = 4.572, $p = .102$), receiving English Learner services ($\chi^2(1)$ = 1.162, $p = .281$), and Hispanic ethnicity ($\chi^2(2)$ = 0.342, $p = .559$). A significant difference was found between children included in and excluded in analyses for economically disadvantaged status ($\chi^2(1)$ = 36.311, $p = .001$): Children with economically disadvantaged status were more likely to be included in analyses. A significant difference was found between children included and excluded in analyses for receiving special-education services ($\chi^2(1)$ = 4.535, $p = .033$): Children who received special-education services were more likely to be included in analyses.

To address the seventh set of research questions (7A: What were the levels of Gardendale children's social-emotional competence, and to what extent did the levels change?; and 7B: Did Gardendale children demonstrate significant improvement in social-emotional learning?), data collected by Gardendale were submitted to Westat and descriptively and inferentially analyzed. Two different assessments were administered because they are developmentally appropriate based on grade level: the DECA for pre-K children, and the DESSA for children in kindergarten through second grade. In the fall, data were not collected for second graders. Therefore, we used children's spring 2023 scores from the 2022–23 school year when they were first graders in place of 2023 fall scores. Because the data collection times differ, we conducted the analyses separately by age group: children in kindergarten through first grade were combined into a group, and second-grade children were a separate group. Children had to have data in both the fall and spring to be included in an outcome analysis of any research question in the seventh set. *T* scores were converted into three categorical levels—Needs Instruction, Typical, and Strengths—based on the technical manual (LeBuffe &Naglieri, 1999; LeBuffe et al., 2014). To address research question 7A, descriptive

analyses were conducted for the categorical levels for both fall and spring assessments. 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 categories over time. As Behavioral Concerns is binary, a McNemar test was conducted between the fall and spring to determine if there were significant differences in Behavioral Concerns over time. To address research question 7B, dependent t tests were conducted for the T scores of each outcome between the fall and spring to determine if there were significant increases over time. 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 samples of children included in and excluded from analyses were similar. For the DECA analyses, no significant differences were found between children included in and excluded from analyses for gender ($\chi^2(1)$ = 0.016, p = .900), economically disadvantaged status ($\chi^2(1)$ = 0.012, p = .912), receiving special-education services ($\chi^2(1)$ = 0. 847, p = .358), race ($\chi^2(1)$ = 0.191, p = .662), and receiving English Learner services ($\chi^2(1)$ = 0.413, p = .520). All children included in analyses were of Hispanic ethnicity; hence, a chi-square demographic test of differences could not be performed. This implies the findings were representative of all pre-K Gardendale children.

For the DESSA analyses, no significant differences were found between children included in and excluded from analyses for gender ($\chi^2(1)$ = 0.087, p = .768), receiving special-education services ($\chi^2(1)$ = 0.91, p = .341), race ($\chi^2(2)$ = 3.551, p = .169), ethnicity ($\chi^2(1)$ = 0.841, p = .359), and receiving English Learner services ($\chi^2(1)$ = 0.144, p = .705). Significant differences were found between children included in and excluded from analyses for grade level ($\chi^2(2)$ = 44.451, p < .001) and economically disadvantaged status ($\chi^2(1)$ = 4.066, p = .044). Children included in analyses were more likely to be in kindergarten than children excluded from analyses, and children included in analyses were more likely to be economically disadvantaged than children excluded from analyses. This implies the findings are not representative of all Gardendale children.

Appendix B

Additional Classroom Assessment Scoring System (CLASS) Results

Appendix B
Additional Classroom Assessment Scoring System (CLASS) Results

Appendix B provides additional CLASS results to address the first research question (What was the overall observed teacher–child interaction quality in Gardendale classrooms across Year 5?). There are two tables provided. Table B-1 provides overall descriptives for the domains and dimensions; Table B-2 provides overall descriptives for the domains and dimensions by grade level.

Table B-1. Overall average Year 5 CLASS scores (*N* = 13)		
CLASS outcome	**M (SD)**	**Total range observed**
Emotional Support domain	6.11 (0.52)	5.20–6.85
Positive climate	6.22 (0.71)	5.00–7.00
Negative climate[a]	6.75 (0.36)	6.20–7.00
Educator sensitivity	5.94 (0.65)	5.00–6.80
Regard for child perspectives	5.53 (0.68)	4.40–6.75
Classroom Organization domain	5.74 (0.69)	4.60–6.80
Behavior management	5.82 (0.95)	4.00–7.00
Productivity	5.95 (0.83)	4.40–7.00
Instructional learning formats	5.44 (0.87)	3.40–6.60
Instructional Support domain	3.71 (0.91)	2.73–5.40
Concept development	3.28 (1.14)	1.80–5.00
Quality of feedback	4.20 (0.90)	2.80–5.80
Language modeling	3.66 (0.95)	2.60–5.60

Note: Domains are in bold font, and dimensions are in plain font.

One classroom (7.7 percent of the total sample) was prekindergarten, five classrooms (38.5 percent) were kindergarten, four classrooms (30.8 percent) were first grade, and three classrooms (23.1 percent) were second grade.

M = mean; SD = standard deviation

[a] Negative climate scores are reversed so that a higher number represents a more emotionally supportive environment (or, a lower negative climate).

Table B-2.   Average Year 5 CLASS scores by grade level (*N* = 13)

CLASS outcome	Kindergarten		First grade		Second grade	
	M (SD)	Total range observed	M (SD)	Total range observed	M (SD)	Total range observed
Emotional Support domain	6.08 (0.52)	(5.40–6.75)	6.30 (0.47)	(5.70–6.85)	6.22 (0.50)	(5.70–6.70)
Positive climate	6.08 (0.74)	(5.00–7.00)	6.55 (0.64)	(5.60–7.00)	6.33 (0.70)	(5.60–7.00)
Negative climate[a]	6.68 (0.36)	(6.20–7.00)	6.80 (0.40)	(6.20–7.00)	7.00 (0.00)	(7.00–7.00)
Educator sensitivity	5.77 (0.55)	(5.00–6.25)	6.30 (0.38)	(6.00–6.80)	6.07 (0.95)	(5.00–6.80)
Regard for child perspectives	5.79 (0.69)	(5.00–6.75)	5.55 (0.72)	(5.00–6.60)	5.47 (0.46)	(5.20–6.00)
Classroom Organization domain	5.37 (0.59)	(4.60–6.17)	6.18 (0.47)	(5.73–6.80)	6.07 (0.70)	(5.33–6.73)
Behavior management	5.62 (1.04)	(4.00–6.50)	6.25 (0.90)	(5.00–7.00)	6.00 (0.87)	(5.40–7.00)
Productivity	5.62 (0.76)	(4.40–6.50)	6.25 (0.64)	(5.60–6.80)	6.47 (0.92)	(5.40–7.00)
Instructional learning formats	4.86 (0.89)	(3.40–5.50)	6.05 (0.68)	(5.20–6.60)	5.73 (0.64)	(5.00–6.20)
Instructional Support domain	3.48 (0.58)	(2.93–4.33)	4.12 (1.04)	(2.87–5.40)	3.58 (1.46)	(2.73–5.27)
Concept development	2.96 (0.93)	(1.80–4.00)	3.75 (1.39)	(1.80–5.00)	3.07 (1.55)	(1.80–4.80)
Quality of feedback	3.84 (0.74)	(3.00–5.00)	4.85 (0.72)	(4.20–5.80)	4.00 (1.31)	(2.80–5.40)
Language modeling	3.64 (0.43)	(3.00–4.00)	3.75 (1.20)	(2.60–5.40)	3.67 (1.68)	(2.60–5.60)

Note: Domains are in bold font, and dimensions are in plain font.

One classroom (7.7 percent of the total sample) was prekindergarten, five classrooms (38.5 percent) were kindergarten, four classrooms (30.8 percent) were first grade, and three classrooms (23.1 percent) were second grade. Results for pre-K are not provided to protect confidentiality.

M = mean; SD = standard deviation.

[a] Negative climate scores are reversed so that a higher number represents a more emotionally supportive environment (or, a lower negative climate).

Appendix C

Additional Growth, Observation, and Learning (GOLD) Results

Appendix C
Additional Growth, Observation, and Learning (GOLD) Results

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

Table C-1a.   Gardendale and normative sample pre-K comparison results for six GOLD outcomes across time

Outcome	Time point	Gardendale mean	Normed mean	Gap (Pre-K–normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored[a]	Graphic depiction of finding (Blue line = Gardendale; Orange line = normative sample)
Cognitive	Fall	394.07	427.94	−33.87	−2.918	27.325	.0070	Significant	Normed	
	Winter	474.07	472.29	1.78	0.335	28.361	.7400	Not significant	No difference	
	Spring	494.11	506.46	−12.35	−2.000	28.242	.0552	Not significant	No difference	
Literacy	Fall	410.70	474.47	−63.77	−2.663	26.047	.0131	Significant	Normed	
	Winter	456.07	506.52	−50.45	−2.78	26.050	.0099	Significant	Normed	
	Spring	456.48	527.01	−70.53	−3.261	26.040	.0031	Significant	Normed	
Mathematics	Fall	350.14	341.41	8.73	0.841	27.289	.4074	Not significant	No difference	
	Winter	390.43	379.14	11.29	2.094	27.851	.0454	Not significant	No difference	
	Spring	402.61	405.25	−2.64	−0.411	27.705	.6842	Not significant	No difference	

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

df = degrees of freedom.

[a] If a statically significant difference was found, 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. Gardendale and normative sample pre-K comparison results for six GOLD outcomes across time

Outcome	Time point	Gardendale mean	Normed mean	Gap (Pre-K–normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored[a]	Graphic depiction of finding (Blue line = Gardendale; Orange line = normative sample)
Oral Language	Fall	442.37	461.29	−18.92	−2.045	18.447	.0554	Not significant	No difference	
Oral Language	Winter	497.58	507.14	−9.56	−1.073	18.471	.2973	Not significant	No difference	
Oral Language	Spring	526.47	544.96	−18.49	−1.844	18.453	.0813	Not significant	No difference	
Physical	Fall	536.43	547.91	−11.48	−1.190	27.518	.2462	Not significant	No difference	
Physical	Winter	560.54	593.33	−32.79	−12.310	33.599	<.0001	Significant	Normed	
Physical	Spring	576.04	628.50	−52.46	−15.690	31.597	<.0001	Significant	Normed	
Social-Emotional	Fall	434.29	426.44	7.85	0.720	27.280	.4797	Not significant	No difference	
Social-Emotional	Winter	460.21	466.51	−6.30	−1.070	27.932	.2946	Not significant	No difference	
Social-Emotional	Spring	479.68	497.35	−17.67	−4.100	29.421	.0003	Significant	Normed	

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

df = degrees of freedom.

[a] If a statically significant difference was found, 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.   Gardendale and normative sample kindergarten comparison results for six GOLD outcomes across time

Outcome	Time point	Gardendale mean	Normed mean	Gap (Kinder–normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored[a]	Graphic depiction of finding (Blue line = Gardendale; Orange line = normative sample)
Cognitive	Fall	484.96	480.12	4.84	0.526	76.562	.6008	Not significant	No difference	
	Winter	532.76	536.18	−3.42	−0.524	78.027	.6021	Not significant	No difference	
	Spring	577.68	575.75	1.93	0.281	78.218	.7794	Not significant	No difference	
Literacy	Fall	449.68	525.74	−76.06	−6.550	74.450	<.0001	Significant	Normed	
	Winter	441.13	570.16	−129.03	−7.676	74.236	<.0001	Significant	Normed	
	Spring	508.52	602.01	−93.49	−6.993	74.490	<.0001	Significant	Normed	
Mathematics	Fall	408.70	385.08	23.62	4.237	78.302	.0001	Significant	Gardendale	
	Winter	435.00	440.51	−5.51	−1.133	78.700	.2606	Not significant	No difference	
	Spring	474.51	481.27	−6.76	−1.148	77.819	.2544	Not significant	No difference	

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

df = degrees of freedom.

[a] If a statically significant difference was found, 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."

Westat®

Table C-2b.   Gardendale and normative sample kindergarten comparison results for six GOLD outcomes across time

Outcome	Time point	Gardendale mean	Normed mean	Gap (Kinder– normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored[a]	Graphic depiction of finding (Blue line = Gardendale; Orange line = normative sample)
Oral Language	Fall	417.23	516.61	−99.38	−5.412	65.429	<.0001	Significant	Normed	
	Winter	490.45	570.84	−80.39	−4.738	65.463	<.0001	Significant	Normed	
	Spring	511.26	610.39	−99.13	−5.495	65.459	<.0001	Significant	Normed	
Physical	Fall	594.01	602.15	−8.14	−1.12	77.253	.2663	Not significant	No difference	
	Winter	645.24	655.72	−10.48	−2.36	80.252	.0206	Significant	Normed	
	Spring	676.45	693.14	−16.69	−3.37	79.881	.0012	Significant	Normed	
Social-Emotional	Fall	464.68	473.23	−8.55	−1.256	76.968	.2131	Not significant	No difference	
	Winter	512.78	522.66	−9.88	−1.935	78.709	.0566	Not significant	No difference	
	Spring	546.16	558.00	−11.84	−2.062	78.477	.0425	Not significant	No difference	

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

df = degrees of freedom.

[a] If a statically significant difference was found, 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]
Pre-K	Cognitive	28	394.07	494.11	100.04	−9.628	27	<.001	Significant	Spring 2024
	Literacy	27	410.70	456.48	45.78	−9.270	26	<.001	Significant	Spring 2024
	Mathematics	28	350.14	402.61	52.47	−6.633	27	<.001	Significant	Spring 2024
	Oral language	19	442.37	526.47	84.10	−14.370	18	<.001	Significant	Spring 2024
	Physical	28	536.43	576.04	39.61	−3.717	27	.001	Significant	Spring 2024
	Social-emotional	28	434.29	479.68	45.39	−5.053	27	<.001	Significant	Spring 2024
Kindergarten	Cognitive	76	484.96	577.68	92.72	−9.906	75	<.001	Significant	Spring 2024
	Literacy	75	449.68	508.52	58.84	−16.541	74	<.001	Significant	Spring 2024
	Mathematics	76	408.70	474.51	65.81	−11.383	75	<.001	Significant	Spring 2024
	Oral language	66	417.23	511.26	94.03	−10.642	65	<.001	Significant	Spring 2024
	Physical	76	594.01	676.45	82.44	−10.933	75	<.001	Significant	Spring 2024
	Social-emotional	76	464.68	546.16	81.48	−14.664	75	<.001	Significant	Spring 2024

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

[a] If a statically significant difference was found, 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

Grade	Outcome	Sample size	Percentage ready for kindergarten
Pre-K spring 2024	Cognitive	28	85.7
	Literacy	27	59.3
	Mathematics	28	67.9
	Oral language	19	89.5
	Physical	28	85.7
	Social-emotional	28	89.3
Kindergarten fall 2023	Cognitive	76	64.5
	Literacy	75	41.3
	Mathematics	76	65.8
	Oral Language	66	36.4
	Physical	76	56.6
	Social-emotional	76	64.3

Appendix D

Additional Woodcock-Johnson and Batería Results

Appendix D
Additional Woodcock-Johnson and Batería Results

Appendix D provides additional direct child assessment results using the Woodcock-Johnson and Batería to address the second set of research questions. There are six tables provided. Based on Letter-Word analyses, the results indicated there was a subset of high-achieving children whose early literacy performance was different from the majority of children; therefore, their results are presented separately. Tables D-1, D-2, and D-3 present findings for the majority of children for both outcomes, and Tables D-4, D-5, and D-6 present Letter-Word findings for the subset of high-achieving children. Table D-1 addresses both parts of research question 3A (What percentage of a random sample of Gardendale 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 3B (Did a random sample of Gardendale children demonstrate significant improvement in early literacy and early numeracy?). Table D-3 addresses research question 3C (Did a random sample of Gardendale children experience accelerated learning to help narrow achievement gaps in early literacy and early numeracy?). Table D-4 addressed both parts of research question 3A only for the high-achieving subgroup; Table D-5 addresses research question 3B only for the high-achieving subgroup; and Table D-6 addresses research question 3C only for the high-achieving subgroup.

Table D-1.	Percentage analysis results comparing fall 2023 and spring 2024 by subtest and grade level									
Outcome	Grade	Sample size	Fall 2023 percentage	Spring 2024 percentage	Difference (spring 2024–fall 2023)	X^2 statistic	df	Initial p value	Adjusted significance	Time favored[a]
Letter-Word	Pre-K	*	50.00	50.00	0.00	Because of limited variation, it was not possible to conduct analyses.				
	Kindergarten	<20	47.37	47.37	0.00					
	First	21	14.29	14.29	0.00					
	Second	<20	21.43	21.43	0.00					
	Total	58	29.31	29.31	0.00					
Applied Problems	Pre-K	*	16.70	16.70	0.00	Because of limited variation, it was not possible to conduct analyses.				
	Kindergarten	23	26.10	30.44	4.34	0.00	1	1	Not Significant	No difference
	First	22	22.73	13.63	−9.10	0.50	1	.480	Not Significant	No difference
	Second	<20	5.90	5.90	0.00	Because of limited variation, it was not possible to conduct analyses.				
	Total	68	19.12	17.64	−1.48	0.00	1	1	Not Significant	No difference

Note: Because of rounding, decimals may not agree to the nearest hundredths, and percentages may not sum to 100 percent. Letter-Word measures early literacy skills, and Applied Problems measures early numeracy skills.

df = degrees of freedom. * = Masked because of small sample size.

[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 grade level

Outcome	Grade	Sample size	Fall 2023 percentage	Spring 2024 percentage	Growth (spring 2024– fall 2023)	t-test statistic	df	Initial p value	Adjusted significance	Time favored[a]
Letter-Word	Pre-K	*	4 years, 11 months	5 years, 3 months	4 months	2.00	*	.140	Not Significant	No difference
	Kindergarten	<20	5 years, 3 months	5 years, 10 months	7 months	7.47	<19	<.001	Significant	Spring 2024
	First	21	6 years, 0 months	6 years, 7 months	7 months	7.38	20	<.001	Significant	Spring 2024
	Second	<20	6 years, 10 months	7 years, 4 months	5 months[b]	2.26	<19	.042	Not Significant	No difference
	Total	58	5 years, 10 months	6 years, 5 months	6 months[b]	8.67	57	<.001	Significant	Spring 2024
Applied Problems	Pre-K	*	3 years, 5 months	4 years, 4 months	11 months	3.77	*	.013	Significant	Spring 2024
	Kindergarten	23	4 years, 10 months	5 years, 5 months	7 months	4.41	22	<.001	Significant	Spring 2024
	First	22	5 years, 11 months	6 years, 6 months	7 months	8.14	21	<.001	Significant	Spring 2024
	Second	<20	6 years, 7 months	7 years, 0 months	6 months[b]	3.23	<19	.005	Significant	Spring 2024
	Total	68	5 years, 6 months	6 years, 1 month	7 months	8.82	67	<.001	Significant	Spring 2024

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

df = degrees of freedom. * = Masked because of small sample size.

[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 grade level

Outcome	Grade	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	Time favored[a]
Letter-Word	Pre-K	*	4 months	2 months	−2 months	−0.77	*	.496	Not Significant	No difference
	Kindergarten	<20	−4 months	−4 months	1 month[b]	0.99	<19	.336	Not Significant	No difference
	First	21	−10 months	−9 months	1 month	0.93	20	.363	Not Significant	No difference
	Second	<20	−9 months	−10 months	−1 month	−0.42	<19	.681	Not Significant	No difference
	Total	58	−7 months	−7 months	0 months	0.36	57	.723	Not Significant	No difference
Applied Problems	Pre-K	*	−1 year, 3 months	−10 months	5 months	1.68	*	.154	Not Significant	No difference
	Kindergarten	23	−9 months	−8 months	1 month	0.62	23	.543	Not Significant	No difference
	First	22	−11 months	−10 months	1 month	0.71	21	.484	Not Significant	No difference
	Second	<20	−1 year	−1 year, 1 month	−1 month	−0.43	<19	.680	Not Significant	No difference
	Total	68	−11 months	−10 months	1 month	0.99	67	.330	Not Significant	No difference

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

df = degrees of freedom. * = Masked because of small sample size.

[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. 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 2023 and spring 2024.

Westat®

Table D-4. Percentage analysis results comparing fall 2023 and spring 2024 for high-achieving children by grade level

Outcome	Grade	Sample size	Fall 2023 percentage	Spring 2024 percentage	Difference (spring 2024– fall 2023)	X^2 statistic	df	Initial p value	Adjusted significance	Time favored[a]
Letter-Word	Pre-K	N/A	N/A	N/A	N/A	Because of limited variation, it was not possible to conduct analyses.				
	Kindergarten	*	*	*	*					
	First	*	*	*	*	1	*	.32	Not Significant	No difference
	Second	*	*	*	*	Because of limited variation, it was not possible to conduct analyses.				
	Total	*	*	*	*	1	*	.32	Not Significant	No difference

Note: Letter-Word measures early literacy skills. N/A = No pre-K children were included in analyses.

df = degrees of freedom. * = Masked because of small sample size.

[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-5.   Age equivalencies results comparing fall 2023 and spring 2024 for high-achieving children by grade level

Outcome	Grade	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]
	Pre-K	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Letter-Word	Kindergarten	*	6 years, 6 months	10 years, 2 months	3 years, 8 months	3.63	*	.171	Not Significant	No difference
	First	*	6 years, 5 months	11 years, 3 months	4 years, 10 months	3.19	*	.171	Not Significant	No difference
	Second	*	10 years, 7 months	14 years, 5 months	3 years, 10 months	3.63	*	.086	Not Significant	No difference
	Total	*	8 years, 0 months	12 years, 0 months	4 years, 0 months	5.28	*	.001	Significant	Spring 2024

Note: Letter-Word measures early literacy skills. N/A = No pre-K children were included in analyses.

df = degrees of freedom. * = Masked because of small sample size.

[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."

Westat

Table D-6.	Gap analysis age equivalency results comparing fall 2023 and spring 2024 for high-achieving children by grade level									
Outcome	Grade	Sample size	Average gap between age equivalency and actual age, fall 2023	Average gap between age equivalency and actual age, spring 2023	Gap (spring 2024– fall 2023)	*t*-test statistic	*df*	Initial *p* value	Adjusted significance	Time favored[a]
Letter-Word	Pre-K	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Kindergarten	*	10 months	4 years	3 years, 2 months	1.80	*	.214	Not Significant	No difference
	First	*	0 months	4 years, 4 months	4 years, 4 months	3.16	*	.195	Not Significant	No difference
	Second	*	3 years	6 years, 3 months	3 years, 4 months[b]	2.75	*	.110	Not Significant	No difference
	Total	*	1 year, 5 months	4 years, 11 months	3 years, 6 months	4.59	*	.003	Significant	Spring 2024

Note: Letter-Word measures early literacy skills. N/A = No pre-K children were included in analyses.

df = degrees of freedom. * = Masked because of small sample size.

[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. 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.

Westat®

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 for the Peabody Picture Vocabulary Test to address the fourth set of research questions. There are three tables provided. Table E-1 addresses research question 4A (What were the receptive vocabulary performance levels of a random sample of Gardendale children?). Table E-2 addresses research question 4B (Did a random sample of Gardendale children demonstrate significant improvement in receptive vocabulary?). Table E-3 addresses research question 4C (What types of receptive vocabulary improvement did a random sample of Gardendale children demonstrate?).

Table E-1.	Descriptives of vocabulary performance levels by assessment time and grade level				
Level	Grade level	Sample size	Fall 2023 percentage	Spring 2024 percentage	Growth (spring 2024–fall 2023)
Well below expected	Pre-K	*	*	*	*
Below expected			≤30.00	11.11	≤30.00
Expected			≤80.00	77.78	≤10.00
Above expected			*	*	*
Well above expected			*	*	*
Well below expected	Kindergarten	<20	≤28.00	≤12.00	−15.79
Below expected			31.58	42.11	10.53
Expected			47.37	47.37	0.00
Above expected			≤7.00	≤7.00	5.26
Well above expected			≤7.00	≤7.00	0.00
Well below expected	First	<20	≤24.00	≤29.00	5.55
Below expected			38.89	27.78	−11.11
Expected			44.44	50.00	5.56
Above expected			≤7.00	≤7.00	0.00
Well above expected			≤7.00	≤7.00	0.00
Well below expected	Second	<20	≤18.00	≤18.00	0.00
Below expected			15.79	21.05	5.26
Expected			73.68	68.42	−5.26
Above expected			≤7.00	≤7.00	0.00
Well above expected			≤7.00	≤7.00	0.00
Well below expected	Total	65	≤16.00	≤14.00	−1.54
Below expected			27.69	27.69	0.00
Expected			58.46	58.46	0.00
Above expected			≤2.00	≤2.00	1.54
Well above expected			≤2.00	≤2.00	0.00

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

* = Masked because of small sample size.

Table E-2.   Vocabulary growth results comparing fall 2023 and spring 2024 by score type and grade level

Score type	Grade level	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	Pre-K	*	90.22	91.00	0.78	0.46	*	.659	Not Significant	None
	Kindergarten	<20	83.89	86.58	2.69	1.40	<19	.177	Not Significant	None
	First	<20	84.89	84.78	−0.11	−0.07	<19	.942	Not Significant	None
	Second	<20	87.00	87.74	0.74	0.46	<19	.626	Not Significant	None
	Total	65	85.95	87.03	1.08	1.27	64	.208	Not Significant	None
Growth scale value	Pre-K	*	465.89	470.56	4.67	4.54	*	.002	Significant	Spring 2024
	Kindergarten	<20	468.69	473.89	5.20	3.74	<19	.002	Significant	Spring 2024
	First	<20	474.61	476.94	2.33	2.25	<19	.038	Significant	Spring 2024
	Second	<20	481.79	484.94	3.15	3.25	<19	.004	Significant	Spring 2024
	Total	65	473.77	477.51	3.74	6.26	64	<.001	Significant	Spring 2024

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

df = degrees of freedom. * = Masked because of small sample size.

[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 change comparing fall 2023 and spring 2024 by grade level		
Score change	**Grade**	**N**	**Percentage**
Stagnant/losing and widened gap	Pre-K	*	≤15.00
Learning but widened gap		*	33.33
Learning slowed but still on par		*	*
Learning		*	44.44
Learning and narrowed gap		*	≤15.00
Stagnant/losing and widened gap	Kindergarten	*	≤31.00
Learning but widened gap		*	≤15.00
Learning slowed but still on par		*	*
Learning		*	26.30
Learning and narrowed gap		*	47.40
Stagnant/losing and widened gap	First	*	33.33
Learning but widened gap		*	≤21.00
Learning slowed but still on par		*	*
Learning		*	≤32.00
Learning and narrowed gap		*	33.33
Stagnant/losing and widened gap	Second	*	≤20.00
Learning but widened gap		*	≤20.00
Learning slowed but still on par		*	≤20.00
Learning		*	31.60
Learning and narrowed gap		*	21.10
Stagnant/losing and widened gap	Total	*	21.53
Learning but widened gap		*	≤16.00
Learning slowed but still on par		*	≤7.00
Learning		*	29.20
Learning and narrowed gap		*	30.80

Note: Because of rounding, decimals may not agree to the nearest hundredths, and percentages may not sum to 100 percent. * = Masked because of small sample size.
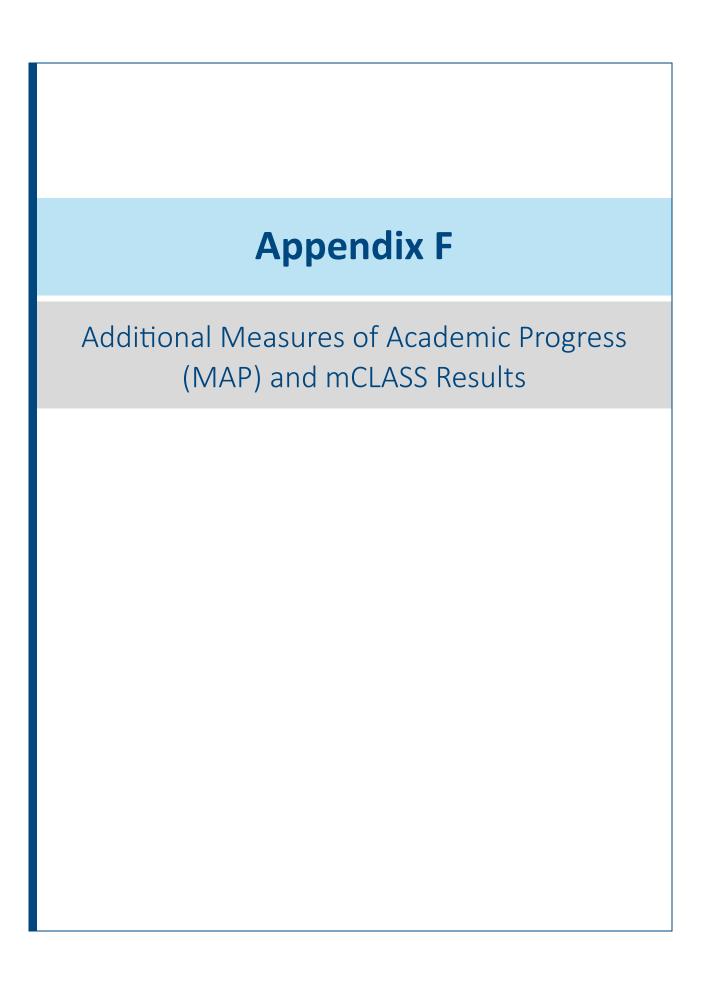
Appendix F

Additional Measures of Academic Progress (MAP) and mCLASS Results

Appendix F
Additional Measures of Academic Progress (MAP) and mCLASS Results

Appendix F provides additional MAP and mCLASS results. There are thirteen tables provided. Tables F-1 through F-10 present MAP findings, and Tables F-11 through F-13 present mCLASS findings. Tables F-1 through F-3 address research question 5A (How did Gardendale children in kindergarten through second grade compare to the normative sample on MAP mathematics and reading?) for mathematics results. Table F-4 addresses research question 5B (Did Gardendale children in kindergarten through second grade demonstrate significant improvement on MAP mathematics and reading?) for mathematics results. Tables F-5 through F-7 address research question 5A for reading results, and Table F-8 addresses research question 5B for reading results. Table F-9 addresses research question 5C (How did second-grade Gardendale children compare to the normative sample on MAP science?). Table F-10 addresses research question 5D (Did second-grade Gardendale children demonstrate significant improvement on MAP science?).

Tables F-11 and F-12 address research question 6A (What were the performance levels of Gardendale children in kindergarten through second grade in mCLASS literacy?). Table F-13 addresses research question 6B (Did Gardendale children in kindergarten through second grade demonstrate significant improvement in mCLASS literacy?).

Table F-1.   Gardendale and normative sample kindergarten comparison results for MAP mathematics across time

Outcome	Time point	Gardendale mean	Normed mean	Gap (Gardendale–normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored[a]
Mathematics	Fall	137.2	139.6	−2.36	−2.31	68.153	.0237	Not significant	No difference
	Winter	149.7	150.1	−0.45	−0.32	68.076	.7477	Not significant	No difference
	Spring	155.8	157.1	−1.36	−0.92	68.068	.3598	Not significant	No difference

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 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 F-2.   Gardendale and normative sample first-grade comparison results for MAP mathematics across time

Outcome	Time point	Gardendale mean	Normed mean	Gap (Gardendale–normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored[a]
Mathematics	Fall	149.9	160.1	−10.16	−6.09	56.044	<.0001	Significant	Normed
	Winter	162.5	170.2	−7.65	−3.87	56.032	.0003	Significant	Normed
	Spring	183.0	176.4	6.64	2.77	56.024	.0076	Significant	Gardendale

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 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 F-3. Gardendale and normative sample second-grade comparison results for MAP mathematics across time

Outcome	Time point	Gardendale mean	Normed mean	Gap (Gardendale–normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored[a]
Mathematics	Fall	163.1	175.0	−11.90	−7.79	50.052	<.0001	Significant	Normed
	Winter	174.6	184.1	−9.44	−5.06	50.035	<.0001	Significant	Normed
	Spring	178.7	189.4	−10.73	−5.79	50.038	<.0001	Significant	Normed

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 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 F-4. Gardendale MAP mathematics growth results comparing fall 2023 and spring 2024 by grade level

Grade level	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]
Kindergarten	69	137.2	155.8	18.55	−16.04	68	<.0001	Significant	Spring 2024
First	57	149.9	183.0	33.14	−13.64	56	<.0001	Significant	Spring 2024
Second	51	163.1	178.7	15.55	−11.46	50	<.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 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 F-5.   Gardendale and normative sample kindergarten comparison results for MAP reading across time

Outcome	Time point	Gardendale mean	Normed mean	Gap (Gardendale–normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored[a]
Reading	Fall	135.6	136.7	−1.09	−0.72	40.040	.4739	Not significant	No difference
	Winter	143.0	146.3	−3.23	−2.01	40.033	.0517	Not significant	No difference
	Spring	147.3	153.1	−5.75	−3.32	40.030	.0019	Significant	Normed

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 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 F-6.   Gardendale and normative sample first-grade comparison results for MAP reading across time

Outcome	Time point	Gardendale mean	Normed mean	Gap (Gardendale–normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored[a]
Reading	Fall	146.6	155.9	−9.29	−6.65	46.0546	<.0001	Significant	Normed
	Winter	156.2	165.9	−9.64	−4.79	46.0286	<.0001	Significant	Normed
	Spring	174.9	171.4	3.47	1.14	46.0145	.2594	Not significant	No difference

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 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 F-7. Gardendale and normative sample second-grade comparison results for MAP reading across time

Outcome	Time point	Gardendale mean	Normed mean	Gap (Gardendale– normed)	*t*-test statistic	*df*	Initial *p* value	Adjusted significance	Group favored[a]
Reading	Fall	163.5	172.4	−8.82	−3.83	39.0247	.0004	Significant	Normed
	Winter	169.1	181.2	−12.10	−5.27	39.0244	<.0001	Significant	Normed
	Spring	175.2	185.6	−10.34	−4.14	39.0218	.0002	Significant	Normed

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 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 F-8. Gardendale MAP reading growth results comparing fall 2023 and spring 2024 by grade level

Grade level	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]
Kindergarten	41	135.6	147.3	11.78	−8.05	40	<.0001	Significant	Spring 2024
First	47	146.6	174.9	28.23	−11.00	46	<.0001	Significant	Spring 2024
Second	40	163.5	175.2	11.70	−6.32	39	<.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 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 F-9.   Gardendale and normative sample second-grade comparison results for MAP science across time

Outcome	Time point	Gardendale mean	Normed mean	Gap (Gardendale–normed)	t-test statistic	df	Initial p value	Adjusted significance	Group favored[a]
Science	Fall	170.0	177.7	−7.70	−4.23	25.030	.0003	Significant	Normed
	Winter	176.6	184.6	−7.97	−3.43	25.016	.0021	Significant	Normed
	Spring	180.3	187.8	−7.60	−3.12	25.015	.0045	Significant	Normed

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 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 F-10.   Gardendale MAP science growth results comparing fall 2023 and spring 2024 by grade level

Grade level	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]
Second	26	170.0	180.3	10.27	−4.09	25	.0004	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 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 F-11.   mCLASS English results comparing score levels by grade over time by grade level

Score level	Grade level	Sample size	Fall 2023 percentage	Winter 2023 percentage	Spring 2024 percentage	Growth (spring 2024– fall 2023)	S statistic	Initial p value	Adjusted significance	Time favored[a]
Well below benchmark	Kindergarten	46	41.30	45.65	41.30	0.00	140.00	.5312	Not significant	No difference
Below benchmark			21.74	13.04	4.35	−17.39				
At benchmark			13.04	17.39	41.30	28.26				
Above benchmark			23.91	23.91	13.04	−10.87				
Well below benchmark	First	50	48.00	44.00	32.00	−16.00	34.00	.0029	Significant	Spring 2024
Below benchmark			16.00	18.00	18.00	2.00				
At benchmark			26.00	26.00	32.00	6.00				
Above benchmark			10.00	12.00	18.00	8.00				
Well below benchmark	Second	42	69.05	57.14	57.14	−11.91	0.00	.0011	Significant	Spring 2024
Below benchmark			11.90	16.67	14.29	2.39				
At benchmark			16.67	21.43	21.43	4.76				
Above benchmark			2.38	4.76	7.14	4.76				
Well below benchmark	Total	138	52.17	48.55	42.75	−9.42	423.00	<.0001	Significant	Spring 2024
Below benchmark			16.67	15.94	12.32	−4.35				
At benchmark			18.84	21.74	31.88	13.04				
Above benchmark			12.32	13.77	13.04	0.72				

Note: 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."

Score level	Grade level	Sample size	Fall 2023 percentage	Winter 2023 percentage	Spring 2024 percentage	Growth (spring 2024– fall 2023)	S statistic	Initial p value	Adjusted significance	Time favored[a]
Table F-12.	mCLASS Spanish results comparing score levels by grade over time by grade level									
Well below benchmark			15.38	7.69	7.69	−7.69				
Below benchmark	Kindergarten	26	19.23	3.85	3.85	−15.38	13.50	.0229	Significant	Spring 2024
At benchmark			46.15	61.54	53.85	7.70				
Above benchmark			19.23	26.92	34.62	15.39				
Well below benchmark			41.67	16.67	16.67	−25.00				
Below benchmark	First	12	25.00	25.00	8.33	−16.67	0.00	.0533	Not significant	No difference
At benchmark			16.67	41.67	58.33	41.66				
Above benchmark			16.67	16.67	16.67	0.00				
Well below benchmark			16.67	≤10.00	≤10.00	≤-6.67				
Below benchmark	Second	12	25.00	≤27.00	25.00	0.00	3.00	.2330	Not significant	No difference
At benchmark			41.67	58.33	66.67	25.00				
Above benchmark			16.67	25.00	≤18.00	≤1.33				
Well below benchmark			22.00	8.00	8.00	−14.00				
Below benchmark	Total	50	22.00	12.00	10.00	−12.00	32.00	.0009	Significant	Spring 2024
At benchmark			38.00	56.00	58.00	20.00				
Above benchmark			18.00	24.00	24.00	6.00				

Note: Because of rounding, decimals may not agree to the nearest hundredths, and percentages may not sum to 100 percent..
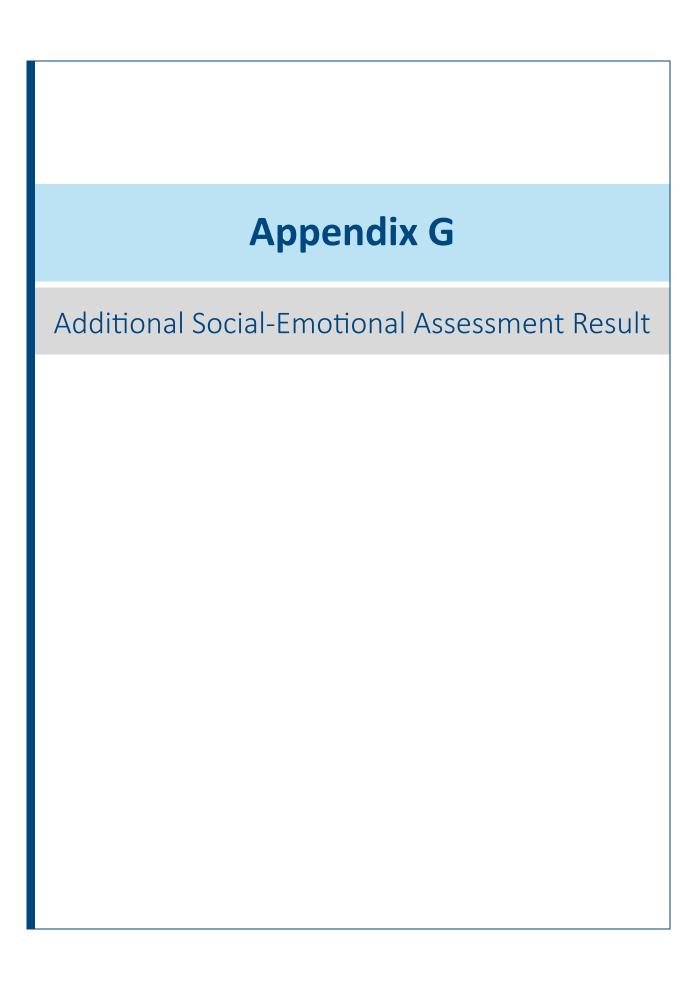
[a] If a statistically significant difference was found after adjustment to correct for multiple hypothesis testing using the Benjamini-Hochberg (1995) technique, the group whose score was greater (i.e., the "favored" group) is listed in this column. If there was no statistically significant difference, this column states that there was "no difference."

Table F-13.   Gardendale mCLASS growth results comparing fall 2023 and spring 2024 by language of assessment and grade level

Language of assessment	Grade level	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]
English	Kindergarten	46	294.6	417.2	122.6	−30.2	45	<.0001	Significant	Spring 2024
	First	50	326.9	448.5	121.6	−36.0	49	<.0001	Significant	Spring 2024
	Second	42	311.0	417.0	106.0	−41.8	41	<.0001	Significant	Spring 2024
Spanish	Kindergarten	26	293.9	427.7	133.8	−17.1	25	<.0001	Significant	Spring 2024
	First	12	351.3	453.2	101.9	−21.7	11	<.0001	Significant	Spring 2024
	Second	12	358.3	449.2	90.9	−11.7	11	<.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 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 G

Additional Social-Emotional Assessment Result

Appendix G
Additional Social-Emotional Assessment Results

Appendix G provides additional social-emotional assessment results which address the seventh set of research questions. Two different assessments were administered depending on grade level. The Devereux Early Childhood Assessment (DECA) was administered to pre-K children, and the Devereux Student Strengths Assessment (DESSA) mini[9] version was administered to children in kindergarten through second grade. There are six tables provided. Tables G-1, G-3, and G-5 address research question 7A (What were the levels of Gardendale children's social-emotional competence, and to what extent did the levels change?). Tables G-2, G-4, and G-6 address research question 7B (Did Gardendale children demonstrate significant improvement in social-emotional learning?). Tables G-1 and G-2 present DECA results, and Tables G-3 through G-6 present DESSA results.

---

[9] Teachers were instructed to administer a mini-DESSA to all children and a full DESSA to any child with a mini score in the lowest category (Needs Instruction). In the 2023–24 school year, no children received a full DESSA in the fall and spring. Therefore, no DESSA analyses for the full version were conducted.

Table G-1. Social-emotional (DECA) results for pre-K children comparing levels in fall 2023 and spring 2024 by outcome

Outcome	Level	Sample size	Fall 2023 percentage	Spring 2024 percentage	Growth (spring 2024–fall 2023)	X^2 statistic	Initial *p* value	Adjusted significance	Time favored[a]
Initiative	Needs Instruction	27	14.81	≤4.00	≤-10.81	4.87	.027	Significant	Spring 2024
	Typical		81.48	≤34.00	≤-47.48				
	Strengths		3.70	70.37	66.67				
Self-control	Needs Instruction	27	18.52	3.70	−14.82	14.66	<.001	Significant	Spring 2024
	Typical		44.44	77.77	33.33				
	Strengths		37.03	18.52	−18.51				
Attachment	Needs Instruction	27	≤4.00	≤4.00	≤4.00	8.33	.004	Significant	Spring 2024
	Typical		≥96.00	≤26.00	≤-70.00				
	Strengths		≤4.00	77.77	≤77.77				
Total Protective Factors	Needs Instruction	27	7.41	≤4.00	≤-3.41	4.50	.034	Significant	Spring 2024
	Typical		74.07	≤37.00	≤-37.07				
	Strengths		18.52	66.66	48.14				
Behavioral Concerns	Typical	27	85.19	85.19	0.00	0.00[b]	1.000	Not significant	No difference
	Area of Need		14.81	14.81	0.00				

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

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

Table G-2. Social-emotional (DECA) growth results for pre-K children comparing fall 2023 and spring 2024 by outcome

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]
Initiative	27	49.63	61.96	12.33	8.75	26	<.001	Significant	Spring 2024
Self-control	27	53.59	54.41	0.81	0.43	26	.670	Not significant	No difference
Attachment	27	53.89	62.70	8.81	9.89	26	<.001	Significant	Spring 2024
Total Protective Factors	27	53.00	61.30	8.30	7.67	26	<.001	Significant	Spring 2024
Behavioral Concerns	27	46.59	46.56	−0.04	−0.04	26	.972	Not significant	No difference

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 G-3. Social-emotional (mini-DESSA) results for kindergarten and first-grade children comparing levels in fall 2023 and spring 2024

Outcome	Level	Sample size	Fall 2023 percentage	Spring 2024 percentage	Difference (spring 2024– fall 2023)	X^2 statistic	Initial *p* value	Adjusted significance	Time favored[a]
Overall Total	Needs Instruction	82	17.07	14.63	−2.44	31.14	<.001	Significant	Spring 2024
	Typical		64.63	64.63	0.00				
	Strengths		18.30	20.73	2.43				

Note: 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 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 G-4. Social-emotional (mini-DESSA) growth results for kindergarten and first-grade children comparing levels in fall 2023 and spring 2024

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]
Overall Total	82	49.23	52.05	2.82	2.60	81	.011	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 G-5. Social-emotional (mini-DESSA) results for second-grade children comparing levels in spring 2023 and spring 2024

Outcome	Level	Sample size	Spring 2023 percentage[a]	Spring 2024 percentage	Difference (spring 2024– spring 2023)	X^2 statistic	Initial p value	Adjusted significance	Time favored[b]
Overall Total	Needs Instruction	25	≤4.00	20.00	≤20.00	5.47	.065	Not significant	No difference
	Typical		64.00	40.00	−24.00				
	Strengths		≤40.00	40.00	≤40.00				

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

[a] As no fall 2023 data was available, we used the most recent assessment time point (spring 2023). It should be noted children were in first grade during spring 2023.

[b] 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 G-6.	Social-emotional (mini-DESSA) growth results for second-grade children comparing levels in spring 2023 and spring 2024								
Outcome	Sample size	Spring 2023 mean[a]	Spring 2024 mean	Growth (spring 2024– spring 2023)	t-test statistic	df	Initial p value	Adjusted significance	Time favored[b]
Overall Total	25	54.60	55.88	1.28	0.54	24	.703	Not significant	No difference

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

df = degrees of freedom.

[a] As no fall 2023 data was available, we used the most recent assessment time point (spring 2023). It should be noted children were in first grade during spring 2023.

[b] 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."

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.

Biancarosa, G., Kennedy, P. C., Park, S., and Otterstedt, J. (2021). *mCLASS DIBELS 8th Edition: Administration and scoring guide*. Center on Teaching & Learning.

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.

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., Shapiro, V. B., & Naglieri, J. A. (2014). *The Devereux Student Strengths Assessment (DESSA)*. Apperson.

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

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.

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

NWEA. (2023). *MAP growth K-2 reading and mathematics content*. Houghton Mifflin Harcourt. https://www.nwea.org/resource-center/fact-sheet/48194/MAP-Growth-K-2-reading-and-mathematics-content_NWEA_onesheet.pdf/.

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.

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

Thum, Y. M., & Kuhfeld, M. (2020). *NWEA 2020 MAP growth achievement status and growth norms for students and schools*. NWEA.

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*, *27*(1), 122–54.