

# Pre-K 4 SA and Gardendale Early Learning Program: Year 6 (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 is prepared by Westat under a professional services agreement for the program assessment of the Pre-K 4 San Antonio Program with the San Antonio Early Childhood Education Municipal Development Corporation, a Texas Municipal Development district (City of San Antonio). The content of the publication does not necessarily reflect the views or policies of the San Antonio Early Childhood Education Municipal Development Corporation, a Texas Municipal Development, or the City of San Antonio, nor does mention of trade names, commercial products, or organizations imply endorsement by the City of San Antonio.*

## Table of Contents

### Appendixes

Appendix A	Evaluation Methods	A-1
Appendix B	Additional Classroom Assessment Scoring System (CLASS) Results	B-1
Appendix C	Additional Growth, Observation, and Learning (GOLD) Results	C-1
Appendix D	Additional Woodcock-Johnson and Batería Results	D-1
Appendix E	Additional Receptive One-Word Picture Vocabulary Test (ROWPVT) and Expressive One-Word Picture Vocabulary Test (EOWPVT) Results	E-1
Appendix F	Additional Measures of Academic Progress (MAP) and mCLASS Results	F-1
Appendix G	Additional Social-Emotional Assessment Results	G-1
Appendix References		R-1

### Tables

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 Devereux Early Childhood Assessment (DECA) by parent and teacher ratings	A-5
Table B.1.	Overall average Year 6 CLASS scores ( $N = 13$ )	B-1
Table B.2.	Average Year 6 CLASS scores by grade level ( $N = 13$ )	B-2
Table C.1a.	Gardendale and normative sample pre-K comparison results for six GOLD outcomes across time	C-2
Table C.1b.	Gardendale and normative sample pre-K comparison results for six GOLD outcomes across time	C-3
Table C.2a.	Gardendale and normative sample kindergarten comparison results for six GOLD outcomes across time	C-4
Table C.2b.	Gardendale and normative sample kindergarten comparison results for six GOLD outcomes across time	C-5
Table C.3.	Growth results comparing fall 2024 and spring 2025 by GOLD outcome and age level	C-6
Table C.4.	Descriptives of kindergarten readiness by GOLD outcome	C-6
Table D.1.	Percentage analysis results comparing fall 2024 and spring 2025 by subtest and grade level	D-2
Table D.2.	Age equivalencies results comparing fall 2024 and spring 2025 by subtest and grade level	D-3
Table D.3.	Gap analysis age equivalency results comparing fall 2024 and spring 2025 by subtest and grade level	D-4

## Table of Contents (continued)

Table D.4.	Percentage analysis results comparing fall 2024 and spring 2025 for high-achieving children by grade level	D-5
Table D.5.	Age equivalencies results comparing fall 2024 and spring 2025 for high-achieving children by grade level	D-6
Table D.6.	Gap analysis age equivalency results comparing fall 2024 and spring 2025 for high-achieving children by grade level	D-7
Table E.1.	Descriptives of vocabulary performance levels by assessment time and grade level	E-2
Table E.2.	Vocabulary growth results comparing fall 2024 and spring 2025 by score type and grade level	E-4
Table F.1.	Gardendale and normative sample kindergarten comparison results for MAP mathematics across time	F-2
Table F.2.	Gardendale and normative sample first-grade comparison results for MAP mathematics across time	F-2
Table F.3.	Gardendale and normative sample second-grade comparison results for MAP mathematics across time	F-3
Table F.4.	Gardendale MAP mathematics growth results comparing fall 2024 and spring 2025 by grade level	F-3
Table F.5.	Gardendale and normative sample kindergarten comparison results for MAP reading across time	F-4
Table F.6.	Gardendale and normative sample first-grade comparison results for MAP reading across time	F-4
Table F.7.	Gardendale and normative sample second-grade comparison results for MAP reading across time	F-5
Table F.8.	Gardendale MAP reading growth results comparing fall 2024 and spring 2025 by grade level	F-5
Table F.9.	Gardendale and normative sample second-grade comparison results for MAP science across time	F-6
Table F.10.	Gardendale MAP science growth results comparing fall 2024 and spring 2025 by grade level	F-6
Table F.11.	mCLASS English results comparing score levels by grade over time by grade level	F-7
Table F.12.	mCLASS Spanish results comparing score levels by grade over time by grade level	F-8
Table F.13.	Gardendale mCLASS growth results comparing fall 2024 and spring 2025 by language of assessment and grade level	F-9
Table G.1.	Social-emotional (DECA) results for pre-K children comparing levels in fall 2024 and spring 2025 by outcome	G-2
Table G.2.	Social-emotional (DECA) growth results for pre-K children comparing fall 2024 and spring 2025 by outcome	G-3

## Table of Contents (continued)

Table G.3.	Social-emotional (mini-DESSA) results for kindergarten and first-grade children comparing levels in fall 2024 and spring 2025	G-3
Table G.4.	Social-emotional (mini-DESSA) growth results for kindergarten and first-grade children comparing levels in fall 2024 and spring 2025	G-4
Table G.5.	Social-emotional (mini-DESSA) results for second-grade children comparing levels in spring 2024 and spring 2025	G-5
Table G.6.	Social-emotional (mini-DESSA) growth results for second-grade children comparing levels in spring 2024 and spring 2025	G-6

# 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),<sup>1</sup> followed by 10-minute coding periods. Observers assign scores during various classroom activities. Westat averaged across all cycles for an overall quality score.

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

---

<sup>1</sup> We observed 7 (8.75%) irregular cycle lengths. For the cycles below 20 minutes, 5 were 15 minutes and 2 were 18 minutes. There were no cycles above 20 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) 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.) Gardendale conducts the GOLD assessment three times (fall, winter, and spring) throughout the school year. We used the revised norms for birth through third grade as a comparison to Gardendale children.

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

The 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). Numerous large-scale preschool studies have used this assessment (e.g., Early et al., 2007; Wong et al., 2008). For children requiring assessment in Spanish, we 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 Spanish for Letter-Word had a higher maximum level (18 years, 1 month) than children assessed in English (9 years, 2 months).

Subtest	English		Spanish	
	Minimum	Maximum	Minimum	Maximum
Letter-Word	2 years, 4 months	9 years, 2 months	5 years, 3 months	18 years, 1 month
Applied Problems	2 years, 0 months	9 years, 4 months	2 years, 0 months	8 years, 4 months

## 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 another assessment: the Wechsler Intelligence Scale for Children, fourth edition, Verbal Comprehension Index (Wechsler, 2003). There are some task differences between the two assessments, which may contribute to the lower correlation; however, both assessments assess similar abilities. Research demonstrates these measures to be 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 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 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 asks the child to name what the image shows. 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.

	Receptive		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) and Devereux Student Strengths Assessment (DESSA)

The DECA, second edition (LeBuffe & Naglieri, 1999), and the DESSA, second edition (Robitaille et al., 2024) 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<sup>2</sup> and six scales in the DESSA.<sup>3</sup> 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 =

<sup>2</sup> The two scales are Total Protective Factors and Behavioral Concerns. The three subscales are Initiative, Self-control, and Attachment.

<sup>3</sup> The six scales are Optimistic Thinking, Relationship Skills, Responsible Decision-Making, Self-Awareness, Self-Management, and Social Awareness.

never, 1 = rarely, 2 = occasionally, 3 = frequently, and 4 = very frequently). In a standardization sample, the DECA reported to have good reliability evidence (LeBuffe & Naglieri, 1999; Center for Resilient Children, 2013). 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 teacher ratings are substantial (McHugh, 2012).

	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 MAP is a norm-referenced assessment of mathematics, reading, and science achievement (NWEA, 2025). Gardendale administers it electronically on an iPad. Gardendale assessed children 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 allowed all grade levels 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. Gardendale assessed children in English and Spanish three times (fall, winter, and spring) throughout the school year. Teachers gave children in kindergarten an online assessment while teachers provided children in first and second grade a reading booklet, and their teacher entered their responses into the data system (Biancarosa et al., 2021).

### Analytic Approach

We addressed research questions 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, we mask all results based on less than 10 children or percentages that round to 0 percent or 100 percent. Where needed, we may mask additional results. Pre-K 4 SA submitted data collected by Gardendale and Pre-K 4 SA to Westat. We descriptively analyzed 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 6?), Westat descriptively analyzed CLASS observation data.

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?), Pre-K 4 SA submitted data collected by Gardendale teachers to Westat. Westat combined data across the three assessment times (fall, winter, and spring) for descriptive and inferential analyses. To answer question 2A, Westat conducted an independent-samples *t* test between Gardendale and the normative sample in the fall, winter, and spring for each outcome. To answer question 2B, Westat conducted a dependent sample *t* test between the fall and spring for

each outcome. We applied the Benjamini-Hochberg (1995) technique to correct for multiple hypothesis testing. To answer question 2C, Westat computed descriptive statistics for the spring assessment point to determine the percentage of 4-year-old children who were ready for kindergarten at the end of the year, and for the fall assessment point to determine the percentage of kindergarten children who started the year ready for kindergarten.

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 in and excluded from analyses for grade level ( $\chi^2(1, N = 138) = 0.032, p = .858$ ), gender ( $\chi^2(1, N = 138) = 0.310, p = .578$ ), race<sup>4</sup> ( $\chi^2(4, N = 138) = 6.485, p = .166$ ), or Hispanic ethnicity<sup>5</sup> ( $\chi^2(1, N = 138) = 0.095, p = .758$ ). 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 economically disadvantaged status ( $\chi^2(1, N = 138) = 36.469, p = <.0001$ ), receiving special education services<sup>6</sup> ( $\chi^2(1, N = 138) = 6.497, p = .012$ ), and receiving English Learner services ( $\chi^2(1, N = 138) = 7.078, p = .008$ ). Children with economically disadvantaged status were more likely to be included in analyses. Children receiving special education services were more likely to be included in analyses. Children receiving English Learner services were more likely to be included in analyses. This implies the GOLD findings are not representative of all Gardendale children.

To address the third set of research questions (3A: What percentage 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 Gardendale children demonstrate significant improvement in early literacy and early numeracy?; and 3C: Did Gardendale 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>7</sup> to Westat. We conducted descriptive and inferential analyses to determine and compare children's early literacy and early numeracy levels. For any research question in the third set, we included children with data in both the fall and spring for each outcome analysis. We converted raw scores 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).<sup>8</sup> For age level, we converted the number of months 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. Because no finer levels (e.g., pre-K and 2 months) were available, and it would not be possible to determine growth over time, we only used age equivalencies. We created a binary indicator 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, we conducted descriptive analyses for each outcome. To answer the second part, we computed the percentages below, and at and above actual age level and analyzed each outcome. We conducted analyses based on the binary indicators: we conducted a McNemar

---

<sup>4</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 no statistically significant difference based on race.

<sup>5</sup> Due to low cell counts, we conducted a Fisher's exact test for Hispanic ethnicity in addition to the chi-square test. Both the chi-square and Fisher's exact test found no statistically significant difference based on Hispanic ethnicity.

<sup>6</sup> Due to low cell counts, we conducted a Fisher's exact test for special education status in addition to the chi-square test. Both the chi-square and Fisher's exact test found a statistically significant difference based on special education status.

<sup>7</sup> We found no significant differences between the random sample and the population for the following variables: gender, classroom type, grade level, and strata.

<sup>8</sup> The age norms differ depending on whether a child was assessed in English or Spanish.

test between the fall and spring for each outcome to determine if there were significant increases in the percentages over time. For research question 3B, we conducted dependent *t* tests 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. We conducted a difference of differences analysis to determine if there was a significant reduction of the achievement gap from fall to spring. We applied the Benjamini-Hochberg (1995) technique to correct for multiple hypothesis testing.

To address the fourth set of research questions (4A: What are the receptive and expressive vocabulary performance levels of children, and to what extent do the performance levels change over the year?; and 4B: Did children demonstrate significant growth in receptive and expressive vocabulary?), Pre-K 4 SA collected data from a random sample<sup>9</sup> and submitted to Westat for descriptive and inferential analysis. 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 4A, 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 4A, 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 4B, 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.

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?), Pre-K 4 SA submitted data collected by Gardendale to Westat. Westat combined data across the three assessment times (fall, winter, and spring) for descriptive and inferential analyses. We conducted analyses separately by grade level as norms vary based on grade level. We included children with data in the fall, winter, and spring in analyses. To answer research question 5A, we conducted an independent-samples *t* test between Gardendale and the normative sample<sup>10</sup> in the fall, winter, and spring by grade level. To answer research question 5B, we conducted a dependent-sample *t* test between the fall and spring to determine if there was significant growth. To answer research question 5C, we conducted an independent-samples *t* test between Gardendale and the normative sample<sup>11</sup> in the fall, winter, and spring for second-grade children. To answer research question 5D, we conducted a dependent-sample *t* test between the fall and spring to determine if there was significant growth. We applied the Benjamini-Hochberg (1995) technique to correct for multiple hypothesis testing.

---

<sup>9</sup> We found no significant differences between the random sample and the population for the following variables: gender, classroom type, grade level, and strata.

<sup>10</sup> Norms were provided in the technical manual (NWEA, 2025).

<sup>11</sup> Norms were provided in the technical manual (NWEA, 2025).

We conducted demographic tests of differences to determine if the sample of children included in and excluded from MAP analyses were similar. In the MAP mathematics assessment, we found no significant differences between children included in and excluded from analyses for grade level ( $\chi^2(2, N = 216) = 0.340, p = .844$ ), gender ( $\chi^2(1, N = 216) = 1.842, p = .175$ ) and Hispanic ethnicity<sup>12</sup> ( $\chi^2(1, N = 216) = 0.020, p = .888$ ). There were, however, significant differences between children included in and excluded from analyses for other characteristics. We found a significant difference between children included in and excluded from analyses for race (which is distinct from Hispanic ethnicity)<sup>13</sup> ( $\chi^2(3, N = 216) = 12.588, p = .006$ ), economically disadvantaged status ( $\chi^2(1, N = 216) = 85.155, p < .001$ ), receiving special education services ( $\chi^2(1, N = 216) = 9.303, p = .002$ ), and receiving English Learner services ( $\chi^2(1, N = 216) = 17.917, p < .001$ ). Children included in analyses were more likely to be American Indian or Alaska Native, White, or multiracial. It is important to note, the racial differences are an additional analysis to accompany the analyses of Hispanic ethnicity as the majority of children are Hispanic. Children included in analyses were more likely to have economically disadvantaged status. Children included in analyses were more likely to receive special education services. Children included in analyses were more likely to be receiving English Learner services. This implies the MAP mathematics findings are not representative of all Gardendale children.

In the MAP reading assessment, we found no significant differences between children included in and excluded from analyses for grade level ( $\chi^2(2, N = 216) = 1.083, p = .582$ ), gender ( $\chi^2(1, N = 216) = 0.621, p = .431$ ), race<sup>14</sup> ( $\chi^2(3, N = 216) = 5.325, p = .150$ ), and Hispanic ethnicity<sup>15</sup> ( $\chi^2(1, N = 216) = 2.271, p = .132$ ). There were, however, significant differences between children included in and excluded from analyses for other characteristics. We found a significant difference between children included in and excluded from analyses for economically disadvantaged status ( $\chi^2(1, N = 216) = 29.266, p < .001$ ): Children included in analyses were more likely to be economically disadvantaged status. We found a significant difference between children included in and excluded from analyses for receiving special education services ( $\chi^2(1, N = 216) = 6.581, p = .010$ ): Children included in analyses were more likely to be receiving special education services. We found a significant difference between children included in and excluded from analyses for receiving English Learner services ( $\chi^2(1, N = 216) = 72.386, p < .001$ ): Children included in analyses were more likely to not be receiving English Learner services. This implies the MAP reading findings are not representative of all Gardendale children.

In the MAP science assessment, we found no significant differences between children included in and excluded from analyses for gender ( $\chi^2(1, N = 65) = 0.466, p = .495$ ), race<sup>16</sup> ( $\chi^2(2, N = 65) = 1.348, p = .510$ ), Hispanic ethnicity<sup>17</sup> ( $\chi^2(1, N = 65) = 0.062, p = .803$ ), and receiving special

---

<sup>12</sup> Due to low cell counts, we conducted a Fisher's exact test for ethnicity in addition to the chi-square test. Both the chi-square and Fisher's exact test found no significant difference based on ethnicity.

<sup>13</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.

<sup>14</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 no significant difference based on race.

<sup>15</sup> Due to low cell counts, we conducted a Fisher's exact test for ethnicity in addition to the chi-square test. Both the chi-square and Fisher's exact test found no significant difference based on ethnicity.

<sup>16</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 no significant difference based on race.

<sup>17</sup> Due to low cell counts, we conducted a Fisher's exact test for ethnicity in addition to the chi-square test. Both the chi-square and Fisher's exact test found no significant difference based on ethnicity.

education services ( $\chi^2(1, N = 65) = 3.385, p = .066$ ). 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 economically disadvantaged status ( $\chi^2(1, N = 65) = 26.220, p < .001$ ): Children included in analyses were more likely to have economically disadvantaged status. We found a significant difference between children included in and excluded from analyses for receiving English Learner services ( $\chi^2(1, N = 65) = 5.826, p = 0.016$ ): Children included in analyses were more likely to be receiving English Learner services. This implies the MAP science findings are not representative of all Gardendale children.

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?), Pre-K 4 SA submitted data collected by Gardendale to Westat. Westat combined data across the three assessment times (fall, winter, and spring) for descriptive and inferential analyses. We conducted analyses separately by language of assessment as the norms vary based on language. We included children with data in the fall, winter, and spring in each outcome analysis. To address research question 6A, we conducted a Wilcoxon signed-rank test 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, we conducted a dependent-sample *t* test between the fall and spring based on language of assessment and grade level to determine if there was significant improvement over the year. 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 mCLASS analyses were similar. We found no significant differences between children included in and excluded from analyses for grade level ( $\chi^2(2, N = 216) = 0.738, p = .692$ ), gender ( $\chi^2(1, N = 216) = 2.553, p = .110$ ), and Hispanic ethnicity<sup>18</sup> ( $\chi^2(1, N = 216) = 0.005, p = .945$ ). 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 race<sup>19</sup> ( $\chi^2(3, N = 216) = 11.984, p = .007$ ): Children included in analyses were more likely to be American Indian or Alaska Native, White, or multiracial. It is important to note, the racial differences are an additional analysis to accompany the analyses of Hispanic ethnicity as the majority of children are Hispanic. We found a significant difference between children included in and excluded in analyses for economically disadvantaged status ( $\chi^2(1, N = 216) = 78.921, p < .001$ ): Children included in analyses were more likely to have economically disadvantaged status. We found a significant difference between children included and excluded in analyses for receiving special-education services ( $\chi^2(1, N = 216) = 9.964, p = .002$ ): Children who included in analyses were more likely to receive special education services. We found a significant difference between children included and excluded in analyses for receiving English Learner services ( $\chi^2(1, N = 216) = 19.099, p < .001$ ): Children included in analyses were more likely to be receive English Learner services. This implies the mCLASS findings are not representative of all Gardendale children.

---

<sup>18</sup> Due to low cell counts, in addition to the chi-square test we also conducted a Fisher's exact test for ethnicity. Both the chi-square and Fisher's exact test found no significant difference based on ethnicity.

<sup>19</sup> Due to low cell counts, in addition to the chi-square test we also conducted a Fisher's exact test for race. Both the chi-square and Fisher's exact test found a significant difference based on race.

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?), Pre-K 4 SA submitted data collected by Gardendale to Westat for descriptive and inferential analysis. Teachers administered two different assessments 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. We included children with data in both the fall and spring in each outcome analysis of any research question in the seventh set. DECA converted *T* scores into three categorical levels—Needs Instruction, Typical, and Strengths—based on the technical manual (LeBuffe & Naglieri, 1999; Robitaille et al., 2024). To address the first part of research question 7A, we conducted descriptive analyses for the categorical levels for both fall and spring assessments. 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 categories over time. As Behavioral Concerns is binary, we conducted a McNemar test between the fall and spring to determine if there were significant differences in Behavioral Concerns over time. To address research question 7B, we conducted dependent *t* tests for the *T* scores of each outcome 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.

We conducted demographic tests of differences to determine if the sample of children included in and excluded from DECA analyses were similar. For the DECA analyses, we found no significant differences between children included in and excluded from analyses for gender ( $\chi^2(1, N = 67) = 0.467, p = .495$ ), receiving special education services ( $\chi^2(1, N = 67) = 3.391, p = .066$ ), race<sup>20</sup> ( $\chi^2(4, N = 67) = 5.547, p = .236$ ), and ethnicity<sup>21</sup> ( $\chi^2(1, N = 67) = 0.906, p = .341$ ). There were, however, significant differences between children included in and excluded from analyses for other demographic characteristics. We found significant differences between children included in and excluded from analyses for economically disadvantaged status ( $\chi^2(1, N = 67) = 12.238, p < .001$ ): Children included in analyses were more likely to have economically disadvantaged status than children excluded from analyses. This implies the DECA findings are not representative of all Gardendale children.

We conducted demographic tests of differences to determine if the sample of children included in and excluded from DESSA analyses were similar. For the DESSA analyses, we found no significant differences between children included in and excluded from analyses for gender ( $\chi^2(1, N = 216) = 3.243, p = .072$ ), receiving special-education services ( $\chi^2(1, N = 216) = 0.498, p = .480$ ), race<sup>22</sup> ( $\chi^2(3, N = 216) = 8.861, p = .073$ ), ethnicity<sup>23</sup> ( $\chi^2(1, N = 216) = 0.193, p = .661$ ), and grade level ( $\chi^2(2, N = 216) = 0.417, p = .812$ ). There were, however, significant differences between children included in and excluded from analyses for other characteristics. We did find significant differences between children included in and excluded from analyses for economically disadvantaged status ( $\chi^2(1, N = 216) = 48.699, p < .001$ ). Children included in analyses were more likely to have

---

<sup>20</sup> Due to low cell counts, we also 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 no significant difference based on race.

<sup>21</sup> All children included in analyses were of Hispanic ethnicity; hence, it was not possible to perform a chi-square demographic test of differences. Therefore, we conducted a Fisher’s exact test for ethnicity.

<sup>22</sup> Due to low cell counts, we conducted Fisher’s exact test for race in addition to the chi-square test. While the chi-square test suggested a significant difference, the Fisher’s exact test did not. Given the small cell sizes, results from Fisher’s exact test are considered more reliable in this context to interpret significant difference based on race.

<sup>23</sup> Due to low cell counts, we conducted a Fisher’s exact test for ethnicity in addition to the chi-square test. Both the chi-square and Fisher’s exact test found no significant difference based on ethnicity.

economically disadvantaged status than children excluded from analyses. This implies the DESSA 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 6?). 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 6 CLASS scores (N = 16)**

CLASS outcome by domain and dimension	M (SD)	Total range observed
<b>Emotional Support domain</b>	6.39 (0.27)	6.05–7.00
Positive climate	6.44 (0.38)	6.00–7.00
Negative climate <sup>a</sup>	6.99 (0.05)	6.80–7.00
Educator sensitivity	6.08 (0.38)	5.50–7.00
Regard for child perspectives	6.05 (0.52)	5.20–7.00
<b>Classroom Organization domain</b>	6.23 (0.46)	5.27–6.87
Behavior management	6.46 (0.47)	5.60–7.00
Productivity	6.49 (0.44)	5.40–7.00
Instructional learning formats	5.75 (0.55)	4.80–6.60
<b>Instructional Support domain</b>	4.05 (0.69)	2.67–5.00
Concept development	3.18 (0.80)	1.60–4.80
Quality of feedback	4.71 (0.91)	3.60–6.20
Language modeling	4.26 (0.84)	2.60–5.60

**Note:** Three classrooms (18.8 percent of the total sample) were prekindergarten, four classrooms (25.0 percent) were kindergarten, five classrooms (31.3 percent) were first grade, and four classrooms (25.0 percent) were second grade.

M = mean; SD = standard deviation

<sup>a</sup> 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 6 CLASS scores by grade level (N = 16)**

CLASS outcome by domain and dimension	Pre-K		Kindergarten		First grade		Second grade	
	M (SD)	Total range observed	M (SD)	Total range observed	M (SD)	Total range observed	M (SD)	Total range observed
<b>Emotional Support domain</b>	6.43 (0.11)	(6.30–6.53)	6.26 (0.19)	(6.08–6.48)	6.39 (0.26)	(6.18–6.83)	6.49 (0.44)	(6.05–7.00)
Positive climate	6.67 (0.31)	(6.40–7.00)	6.20 (0.16)	(6.00–6.40)	6.36 (0.41)	(6.00–7.00)	6.60 (0.49)	(6.00–7.00)
Negative climate <sup>a</sup>	7.00 (0.00)	(7.00–7.00)	7.00 (0.00)	(7.00–7.00)	7.00 (0.00)	(7.00–7.00)	6.95 (0.10)	(6.80–7.00)
Educator sensitivity	6.17 (0.29)	(6.00–6.50)	5.94 (0.43)	(5.50–6.50)	6.00 (0.35)	(5.50–6.50)	6.25 (0.50)	(6.00–7.00)
Regard for child perspectives	5.87 (0.58)	(5.20–6.20)	5.90 (0.26)	(5.60–6.20)	6.20 (0.37)	(5.80–6.80)	6.15 (0.89)	(5.20–7.00)
<b>Classroom Organization domain</b>	6.29 (0.08)	(6.20–6.33)	5.82 (0.48)	(5.27–6.33)	6.35 (0.33)	(6.00–6.67)	6.47 (0.58)	(5.60–6.87)
Behavior management	6.53 (0.23)	(6.40–6.80)	6.00 (0.37)	(5.60–6.40)	6.56 (0.46)	(6.00–7.00)	6.75 (0.50)	(6.00–7.00)
Productivity	6.53 (0.12)	(6.40–6.60)	6.15 (0.53)	(5.40–6.60)	6.60 (0.32)	(6.20–7.00)	6.65 (0.57)	(5.80–7.00)
Instructional learning formats	5.80 (0.53)	(5.20–6.20)	5.30 (0.60)	(4.80–6.00)	5.88 (0.30)	(5.40–6.20)	6.00 (0.69)	(5.00–6.60)
<b>Instructional Support domain</b>	4.49 (0.49)	(3.93–4.87)	3.53 (0.57)	(2.80–4.20)	4.27 (0.49)	(3.60–4.87)	3.97 (0.98)	(2.67–5.00)
Concept development	3.20 (0.72)	(2.60–4.00)	2.65 (0.57)	(2.00–3.40)	3.44 (0.38)	(3.00–4.00)	3.35 (1.34)	(1.60–4.80)
Quality of feedback	5.27 (1.27)	(3.80–6.00)	4.30 (0.62)	(3.60–5.00)	4.84 (1.11)	(3.60–6.20)	4.55 (0.70)	(3.80–5.40)
Language modeling	5.00 (0.87)	(4.00–5.60)	3.65 (0.62)	(2.80–4.20)	4.52 (0.61)	(3.80–5.40)	4.00 (0.97)	(2.60–4.80)

**Note:** Three classrooms (18.8 percent of the total sample) were prekindergarten, four classrooms (25.0 percent) were kindergarten, five classrooms (31.3 percent) were first grade, and four classrooms (25.0 percent) were second grade.

M = mean; SD = standard deviation.

<sup>a</sup> 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 second 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?). Because of the complexity of the data for each GOLD outcome, we split these tables into two related subtables (i.e., Tables C.1a, C.1b, C.2a, and C.2b). Additionally, we present the results for pre-K and kindergarten in separate tables because 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 Cognitive, Literacy, and Mathematics 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 <sup>a</sup>	Graphic depiction of finding (Blue line = Gardendale; Orange line = normative sample)
Cognitive	Fall	403.68	427.94	-24.26	-2.688	49.980	0.0097	Significant	Normed	
	Winter	456.10	472.29	-16.19	-2.565	50.732	0.0133	Significant	Normed	
	Spring	521.30	506.46	14.84	1.662	50.060	0.1027	Not significant	No difference	
Literacy	Fall	493.03	474.47	18.56	4.474	36.161	0.0001	Significant	Gardendale	
	Winter	524.54	506.52	18.02	4.427	35.322	0.0001	Significant	Gardendale	
	Spring	535.37	527.01	8.36	2.746	36.897	0.0093	Significant	Gardendale	
Mathematics	Fall	340.14	341.41	-1.27	-0.209	50.556	0.8352	Not significant	No difference	
	Winter	374.36	379.14	-4.78	-0.896	50.576	0.3746	Not significant	No difference	
	Spring	422.14	405.25	16.89	3.794	51.725	0.0004	Significant	Gardendale	

**Note:** We present group mean information in scaled scores. 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. Gardendale and normative sample pre-K comparison results for Oral Language, Physical, and Social-Emotional 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 <sup>a</sup>	Graphic depiction of finding (Blue line = Gardendale; Orange line = normative sample)
Oral Language	Fall	435.08	461.29	-26.21	-2.290	36.581	0.0279	Not significant	No difference	
	Winter	492.62	507.14	-14.52	-1.417	36.710	0.1650	Not significant	No difference	
	Spring	521.03	544.96	-23.93	-1.981	36.620	0.0551	Not significant	No difference	
Physical	Fall	512.71	547.91	-35.20	-4.714	52.663	<.0001	Significant	Normed	
	Winter	561.67	593.33	-31.66	-5.955	53.762	<.0001	Significant	Normed	
	Spring	620.38	628.50	-8.12	-0.917	52.119	0.3632	Not significant	No difference	
Social-Emotional	Fall	405.61	426.44	-20.83	-4.615	53.174	<.0001	Significant	Normed	
	Winter	448.61	466.51	-17.90	-4.261	53.484	0.0001	Significant	Normed	
	Spring	492.63	497.35	-4.72	-0.931	53.183	0.3560	Not significant	No difference	

**Note:** We present group mean information in scaled scores. 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. Gardendale and normative sample kindergarten comparison results for Cognitive, Literacy, and Mathematics 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 <sup>a</sup>	Graphic depiction of finding (Blue line = Gardendale; Orange line = normative sample)
Cognitive	Fall	479.77	480.12	-0.35	-0.047	56.739	0.9627	Not significant	No difference	
	Winter	530.48	536.18	-5.70	-0.709	56.451	0.4810	Not significant	No difference	
	Spring	562.64	575.75	-13.11	-1.166	55.866	0.2487	Not significant	No difference	
Literacy	Fall	508.88	525.74	-16.87	-2.792	39.888	0.0080	Significant	Normed	
	Winter	557.75	570.16	-12.41	-2.432	40.379	0.0196	Significant	Normed	
	Spring	578.63	602.01	-2.34	-3.591	40.096	0.0009	Significant	Normed	
Mathematics	Fall	393.64	385.08	8.56	1.502	57.314	0.1385	Not significant	No difference	
	Winter	449.41	440.51	8.90	1.831	57.714	0.0723	Not significant	No difference	
	Spring	485.30	481.27	3.58	0.512	56.449	0.6107	Not significant	No difference	

**Note:** We present group mean information in scaled scores. The Adjusted significance column indicates significance levels (*p* values) after adjustment for multiple hypothesis testing correction (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.2b. Gardendale and normative sample kindergarten comparison results for Oral Language, Physical, and Social-Emotional 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 <sup>a</sup>	Graphic depiction of finding (Blue line = Gardendale; Orange line = normative sample)
Oral Language	Fall	466.44	516.61	-50.17	-3.412	44.455	0.0014	Significant	Normed	
	Winter	493.91	570.84	-76.93	-4.289	44.280	0.0001	Significant	Normed	
	Spring	527.29	610.39	-83.10	-4.473	44.293	0.0001	Significant	Normed	
Physical	Fall	611.13	602.15	8.98	0.918	55.903	0.3628	Not significant	No difference	
	Winter	666.71	655.72	10.99	1.650	56.659	0.1045	Not significant	No difference	
	Spring	695.88	693.14	2.74	0.443	57.263	0.6595	Not significant	No difference	
Social-Emotional	Fall	464.41	473.23	-8.82	-1.156	56.144	0.2526	Not significant	No difference	
	Winter	506.39	522.66	-16.27	-3.686	58.691	0.0005	Significant	Normed	
	Spring	546.93	558.00	-11.07	-2.556	59.604	0.0131	Significant	Normed	

**Note:** We present group mean information in scaled scores. The Adjusted significance column indicates significance levels (*p* values) after adjustment for multiple hypothesis testing correction (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.3. Growth results comparing fall 2024 and spring 2025 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>
Pre-K	Cognitive	50	403.68	521.30	117.62	-10.864	49	<.0001	Significant	Spring 2025
	Literacy	35	493.03	535.37	42.34	-13.665	34	<.0001	Significant	Spring 2025
	Mathematics	50	340.14	422.14	82.00	-10.236	49	<.0001	Significant	Spring 2025
	Oral Language	37	435.08	521.03	85.95	-11.824	36	<.0001	Significant	Spring 2025
	Physical	52	512.71	620.38	107.67	-11.782	51	<.0001	Significant	Spring 2025
	Social-Emotional	51	405.61	492.63	87.02	-12.082	50	<.0001	Significant	Spring 2025
Kindergarten	Cognitive	56	479.77	562.64	82.88	-7.319	55	<.0001	Significant	Spring 2025
	Literacy	40	508.88	578.63	69.75	-12.740	39	<.0001	Significant	Spring 2025
	Mathematics	56	393.64	485.30	91.66	-17.239	55	<.0001	Significant	Spring 2025
	Oral Language	45	466.44	527.29	60.84	-5.007	44	<.0001	Significant	Spring 2025
	Physical	56	611.13	695.88	84.75	-8.688	55	<.0001	Significant	Spring 2025
	Social-Emotional	56	464.41	546.93	82.52	-10.862	55	<.0001	Significant	Spring 2025

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

<sup>a</sup> If there was a statically significant difference, the 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**

Grade	Outcome	Sample size	Percentage ready for kindergarten
Pre-K spring 2025	Cognitive	50	94.0
	Literacy	35	97.1
	Mathematics	50	82.0
	Oral Language	37	78.4
	Physical	52	94.2
	Social-Emotional	51	94.1
Kindergarten fall 2024	Cognitive	56	58.9
	Literacy	40	52.5
	Mathematics	56	67.9
	Oral Language	45	51.1
	Physical	56	75.0
	Social-Emotional	56	58.9

# Appendix D

Additional Woodcock-Johnson and  
Batería Results

## Appendix D

### Additional Woodcock-Johnson and Bateria Results

Appendix D provides additional direct child assessment results using the Woodcock-Johnson and Bateria to address the third set of research questions. There are six tables provided. Based on analyses, the results indicated there was a subset of high-achieving children whose early literacy and numeracy performance was different from the majority of children; therefore, we present their results 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 the findings for the subset of high-achieving children. Table D.1 addresses both parts of research question 3A (What percentage 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 Gardendale children demonstrate significant improvement in early literacy and early numeracy?). Table D.3 addresses research question 3C (Did 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 2024 and spring 2025 by subtest and grade level**

Outcome	Grade	Sample size	Fall 2024 percentage	Spring 2025 percentage	Difference (spring 2025–fall 2024)	X <sup>2</sup> statistic	df	Initial p value	Adjusted significance	Time favored <sup>a</sup>
Letter-Word	Pre-K	14	35.71	28.57	-7.14	1.00	1	0.3173	Not significant	No difference
	Kindergarten	18	27.78	38.89	11.11	2.00	1	0.1573	Not significant	No difference
	First	13	<33.00	<43.00	<10.00	Because of limited variation, it was not possible to conduct analyses.				
	Second	14	21.43	42.86	21.43	3.00	1	0.0833	Not significant	No difference
	Total	59	<38.00	<55.00	<17.00	2.67	1	0.1025	Not significant	No difference
Applied Problems	Pre-K	11	9.09	27.27	18.18	2.00	1	0.1573	Not significant	No difference
	Kindergarten	18	33.33	50.00	16.67	1.29	1	0.2568	Not significant	No difference
	First	19	36.84	26.32	-10.53	0.67	1	0.4142	Not significant	No difference
	Second	15	6.67	13.33	6.67	1.00	1	0.3173	Not significant	No difference
	Total	63	23.81	30.16	6.35	1.00	1	0.3173	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. The Adjusted significance column indicates significance levels (*p* values) after adjustment for multiple hypothesis testing correction (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a statistically 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 D.2. Age equivalencies results comparing fall 2024 and spring 2025 by subtest and grade level**

Outcome	Grade	Sample size	Fall 2024 percentage	Spring 2025 percentage	Growth (spring 2025–fall 2024)	t-test statistic	df	Initial p value	Adjusted significance	Time favored <sup>a</sup>
Letter-Word	Pre-K	14	4 years, 6 months	4 years, 10 months	5 months <sup>b</sup>	6.3	13	<.0001	Significant	Spring 2025
	Kindergarten	18	5 years, 5 months	6 years, 2 months	9 months	8.6	17	<.0001	Significant	Spring 2025
	First	13	6 years, 3 months	7 years	9 months	6.1	12	0.0001	Significant	Spring 2025
	Second	14	7 years, 5 months	8 years, 3 months	11 months <sup>b</sup>	4.8	13	0.0004	Significant	Spring 2025
	Total	59	5 years, 10 months	6 years, 6 months	8 months	11.1	58	<.0001	Significant	Spring 2025
Applied Problems	Pre-K	11	3 years, 10 months	4 years, 4 months	6 months	2.7	10	0.0242	Significant	Spring 2025
	Kindergarten	18	5 years, 4 months	5 years, 11 months	6 months <sup>b</sup>	3.7	17	0.0020	Significant	Spring 2025
	First	19	6 years, 2 months	6 years, 9 months	7 months	4.4	18	0.0004	Significant	Spring 2025
	Second	15	6 years, 7 months	7 years, 3 months	8 months	5.0	14	0.0002	Significant	Spring 2025
	Total	63	5 years, 8 months	6 years, 2 months	7 months <sup>b</sup>	7.9	62	<.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 for multiple hypothesis testing correction (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a statistically 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> 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 grade level**

Outcome	Grade	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	Time favored <sup>a</sup>
Letter-Word	Pre-K	14	- 2 months	- 3 months	- 1 month	-1.8	13	0.0936	Not significant	No difference
	Kindergarten	18	- 4 months	- 1 month	3 months	2.7	17	0.0139	Significant	Spring 2025
	First	13	- 3 months	0 months	3 months	1.9	12	0.0797	Not significant	No difference
	Second	14	- 4 months	0 months	4 months	2.0	13	0.0680	Not significant	No difference
	Total	59	- 3 months	- 1 month	2 months	3.0	58	0.0045	Significant	Spring 2025
Applied Problems	Pre-K	11	- 9 months	- 10 months	0 months <sup>b</sup>	-0.1	10	0.9204	Not significant	No difference
	Kindergarten	18	- 4 months	- 4 months	0 months	0.0	17	0.9938	Not significant	No difference
	First	19	- 4 months	- 3 months	0 months <sup>b</sup>	0.3	18	0.7785	Not significant	No difference
	Second	15	-1 year, 2 months	-1 year	2 months	1.1	14	0.2943	Not significant	No difference
	Total	63	- 7 months	- 7 months	1 month <sup>b</sup>	0.6	62	0.5460	Not significant	No difference

**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 for multiple hypothesis testing correction (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

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

<sup>b</sup> If there was a statistically 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 there was “no difference.”

<sup>c</sup> Because of rounding, the gap is not the exact difference between fall 2024 and spring 2025.

**Table D.4. Percentage analysis results comparing fall 2024 and spring 2025 for high-achieving children by grade level**

Outcome	Grade	Sample size	Fall 2024 percentage	Spring 2025 percentage	Difference (spring 2025–fall 2024)	X <sup>2</sup> statistic	df	Initial p value	Adjusted significance	Time favored <sup>a</sup>
Letter-Word	Pre-K	*	*	*	*	1.0	1.0	0.3173	Not significant	No difference
	Kindergarten	*	*	*	*					
	First	*	*	*	*					
	Second	*	*	*	*					
	Total	*	80.00	90.00	10.00					
Applied Problems	Pre-K	*	*	*	*	Because of limited variation, it was not possible to conduct analyses.				
	Kindergarten	*	*	*	*					
	First	*	*	*	*					
	Second	*	*	*	*					
	Total	*	*	*	*					

**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. The Adjusted significance column indicates significance levels (*p* values) after adjustment for multiple hypothesis testing correction (Benjamini-Hochberg, 1995).

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

<sup>a</sup> If there was a statistically 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 D.5. Age equivalencies results comparing fall 2024 and spring 2025 for high-achieving children by grade level**

Outcome	Grade	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	Pre-K	*	2 years, 4 months	5 years, 0 months	2 years, 8 months	Because of limited sample size, it was not possible to conduct analyses.				
	Kindergarten	*	6 years, 5 months	10 years, 9 months	4 years, 4 months	Because of limited sample size, it was not possible to conduct analyses.				
	First	*	8 years, 0 months	14 years, 8 months	6 years, 9 months <sup>b</sup>	7.06	*	0.0009	Significant	Spring 2025
	Second	*	11 years, 9 months	15 years, 3 months	3 years, 6 months	6.38	*	0.0989	Not significant	No difference
	Total	*	8 years, 0 months	13 years, 5 months	5 years, 5 months	6.97	*	0.0001	Significant	Spring 2025
Applied Problems	Pre-K	*	2 years, 7 months	5 years, 3 months	2 years, 9 months <sup>b</sup>	13.99	*	0.0008	Significant	Spring 2025
	Kindergarten	*	3 years, 0 months	6 years, 8 months	3 years, 8 months	Because of limited sample size, it was not possible to conduct analyses.				
	First	*	N/A	N/A	N/A					
	Second	*	7 years, 3 months	9 years, 4 months	2 years, 1 month					
	Total	*	3 years, 5 months	6 years, 2 months	2 years, 9 months	11.48	*	0.0001	Significant	Spring 2025

**Note:** Letter-Word measures early literacy skills, and Applied Problems measures early numeracy skills. N/A = There were no first grade children included in analyses. The Adjusted significance column indicates significance levels (p values) after adjustment for multiple hypothesis testing correction (Benjamini-Hochberg, 1995).

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

<sup>a</sup> If there was a 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> Because of rounding, the gap is not the exact difference between fall 2024 and spring 2025.

**Table D.6. Gap analysis age equivalency results comparing fall 2024 and spring 2025 for high-achieving children by grade level**

Outcome	Grade	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	Time favored <sup>b</sup>
Letter-Word	Pre-K	*	-2 years, 3 months	- 1 month	2 years, 2 months	Because of limited sample size, it was not possible to conduct analyses.				
	Kindergarten	*	7 months	4 years, 4 months	3 years, 10 months <sup>c</sup>	Because of limited sample size, it was not possible to conduct analyses.				
	First	*	1 year, 5 months	7 years, 8 months	6 years, 3 months	6.5	*	0.0013	Significant	Spring 2025
	Second	*	4 years, 1 month	7 years, 1 month	3 years	5.5	*	0.1143	Not significant	No difference
	Total	*	1 year, 6 months	6 years, 5 months	4 years, 11 months	6.3	*	0.0001	Significant	Spring 2025
Applied Problems	Pre-K	*	-2 years	2 months	2 years, 2 months	11.4	*	0.0014	Significant	Spring 2025
	Kindergarten	*	-2 years, 10 months	4 months	3 years, 2 months	Because of limited sample size, it was not possible to conduct analyses.				
	First	*	NA	NA	NA					
	Second	*	- 7 months	11 months	1 year, 7 months <sup>c</sup>					
	Total	*	-1 year, 11 months	4 months	2 years, 3 months	9.3	*	0.0002	Significant	Spring 2025

**Note:** Letter-Word measures early literacy skills, and Applied Problems measures early numeracy skills. N/A = There were no first grade children included in analyses. The Adjusted significance column indicates significance levels (*p* values) after adjustment for multiple hypothesis testing correction (Benjamini-Hochberg, 1995).

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

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

<sup>b</sup> If there was a 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 there was “no difference.”

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

# Appendix E

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

## Appendix E

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

Appendix E provides additional direct child assessment results for the ROWPVT (Receptive) and EOWPVT (Expressive) to address the fourth set of research questions. There are two tables provided. Table E.1 addresses research question 4A (What were the receptive and expressive vocabulary performance levels of Gardendale children?). Table E.2 addresses research question 4B (Did Gardendale children demonstrate significant improvement in receptive and expressive vocabulary?).

**Table E.1. Descriptives of vocabulary performance levels by assessment time and grade level**

Exam	Level	Grade 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>
Receptive	Well below expected	Pre-K	14	<10.0	<10.0	<10.0	3	0.2330	Not significant	No difference
	Below expected			<33.0	<10.0	<-22.0				
	Expected			64.3	92.9	28.6				
	Above expected			<25.0	<10.0	<-15.0				
	Well above expected			<10.0	<10.0	<10.0				
	Well below expected	Kindergarten	12	<10.0	<10.0	<10.0	12	0.2330	Not significant	No difference
	Below expected			8.3	25.0	16.7				
	Expected			75.0	58.3	-16.7				
	Above expected			8.3	16.7	8.3				
	Well above expected			<10.0	<10.0	<10.0				
	Well below expected	First	20	<7.0	<7.0	<7.0	7	0.4840	Not significant	No difference
	Below expected			15.0	10.0	-5.0				
	Expected			<90.0	<83.0	<7.0				
	Above expected			5.0	10.0	5.0				
	Well above expected			<7.0	<7.0	<7.0				
	Well below expected	Second	15	<10.0	<10.0	<10.0	1	1.0000	Not significant	No difference
	Below expected			<10.0	<10.0	<10.0				
	Expected			80.0	86.7	6.7				
	Above expected			<10.0	<10.0	<-10.0				
	Well above expected			<10.0	<10.0	<10.0				
Well below expected	Total	61	<2.0	<2.0	<2.0	72	0.8293	Not significant	No difference	
Below expected			13.1	9.8	-3.3					
Expected			75.4	80.3	4.9					
Above expected			8.2	6.6	-1.6					
Well above expected			<2.0	<2.0	<2.0					

Exam	Level	Grade 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>
Expressive	Well below expected	Pre-K	13	<10.0	<10.0	<10.0	3	0.2330	Not significant	No difference
	Below expected			30.8	7.7	-23.1				
	Expected			53.8	69.2	15.4				
	Above expected			<9.0	<25.0	<16.0				
	Well above expected			15.4	7.7	-7.7				
	Well below expected	Kindergarten	13	<10.0	<10.0	<10.0	0	0.1489	Not significant	No difference
	Below expected			<35.0	<25.0	<10.0				
	Expected			69.2	46.2	-23.1				
	Above expected			7.7	30.8	23.1				
	Well above expected			<10.0	<10.0	<10.0				
	Well below expected	First	20	<19.0	<12.0	<7.0	4	0.8501	Not significant	No difference
	Below expected			<24.0	<17.0	<7.0				
	Expected			<89.0	<82.0	<7.0				
	Above expected			10.0	5.0	-5.0				
	Well above expected			<31.0	<19.0	<12.0				
	Well below expected	Second	15	<10.0	<10.0	<10.0	1.5	1.000	Not significant	No difference
	Below expected			<40.0	<30.0	<10.0				
	Expected			<80.0	<70.0	<10.0				
	Above expected			<34.0	<24.0	<10.0				
	Well above expected			<10.0	<10.0	<10.0				
Well below expected	Total	61	<10.0	<8.0	<2.0	28	0.0963	Not significant	No difference	
Below expected			18.0	13.1	-4.9					
Expected			65.6	63.9	-1.6					
Above expected			8.2	14.8	6.6					
Well above expected			<10.0	<8.0	<2.0					

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

<sup>a</sup> If there was a 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 E.2. Vocabulary growth results comparing fall 2024 and spring 2025 by score type and grade level**

Outcome	Grade level	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	Pre-K	14	97.93	104.14	6.21	2.192	13	0.0472	Significant	Spring 2025
	Kindergarten	12	100.17	97.25	-2.92	-0.696	11	0.5008	Not significant	No difference
	First	20	97.60	101.00	3.40	1.874	19	0.0764	Not significant	No difference
	Second	15	94.20	99.67	5.47	2.389	14	0.0315	Significant	Spring 2025
	Total	61	97.34	100.66	3.31	2.432	60	0.0180	Significant	Spring 2025
Expressive	Pre-K	13	97.15	102.69	5.54	1.895	12	0.0825	Not significant	No difference
	Kindergarten	13	102.54	104.85	2.31	1.123	12	0.2834	Not significant	No difference
	First	20	95.85	100.00	4.15	2.210	19	0.0396	Significant	Spring 2025
	Second	15	92.20	94.67	2.47	1.302	14	0.2139	Not significant	No difference
	Total	61	96.66	100.30	3.64	3.405	60	0.0012	Significant	Spring 2025

**Note:** Because of rounding, decimals may not agree to the nearest hundredth, and percentages may not sum to 100 percent. The Adjusted significance column indicates significance levels (*p* values) after adjustment for multiple hypothesis testing correction (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a 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.”

# 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?) in English and Spanish, respectively. 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 <sup>a</sup>
Mathematics	Fall	139.36	141.20	-1.81	-1.44	54.007	0.1564	Not significant	No difference
	Winter	147.16	150.70	-3.49	-2.31	54.005	0.0245	Significant	Normed
	Spring	158.58	157.80	0.81	0.56	54.006	0.5780	Not significant	No difference

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

*df* = degrees of freedom.

<sup>a</sup> If there was a 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 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 <sup>a</sup>
Mathematics	Fall	154.52	159.30	-4.77	-3.05	64.005	0.0033	Significant	Normed
	Winter	166.12	168.30	-2.14	-1.34	64.005	0.1834	Not significant	No difference
	Spring	172.95	175.00	-2.04	-1.23	64.005	0.2219	Not significant	No difference

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

*df* = degrees of freedom.

<sup>a</sup> If there was a 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 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 <sup>a</sup>
Mathematics	Fall	162.40	172.90	-10.47	-5.56	51.003	<.0001	Significant	Normed
	Winter	174.96	181.20	-6.24	-3.75	51.005	.0005	Significant	Normed
	Spring	181.90	187.50	-5.56	-3.47	51.005	.0011	Significant	Normed

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

*df* = degrees of freedom.

<sup>a</sup> If there was a 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 F.4. Gardendale MAP mathematics growth results comparing fall 2024 and spring 2025 by grade level**

Grade level	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>
Kindergarten	55	139.36	158.58	19.22	-20.31	54	<.0001	Significant	Spring 2025
First	65	154.52	172.95	18.43	-17.39	64	<.0001	Significant	Spring 2025
Second	52	162.40	181.90	19.50	-16.05	51	<.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 for multiple hypothesis testing correction (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a 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 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 <sup>a</sup>
Reading	Fall	139.59	138.10	1.52	1.01	38.003	0.3169	Not significant	No difference
	Winter	146.26	146.00	0.25	0.14	38.002	0.8928	Not significant	No difference
	Spring	155.41	152.00	3.45	1.68	38.002	0.1005	Not significant	No difference

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

*df* = degrees of freedom.

<sup>a</sup> If there was a 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 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 <sup>a</sup>
Reading	Fall	146.38	155.40	-8.99	-4.53	38.002	<.0001	Significant	Normed
	Winter	162.03	162.50	-0.48	-0.26	38.003	0.8001	Not significant	No difference
	Spring	163.23	167.90	-4.64	-2.37	38.003	0.0231	Significant	Normed

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

*df* = degrees of freedom.

<sup>a</sup> If there was a 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 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 <sup>a</sup>
Reading	Fall	161.70	170.10	-8.36	-3.99	36.003	0.0003	Significant	Normed
	Winter	167.84	176.70	-8.86	-3.71	36.002	0.0007	Significant	Normed
	Spring	178.11	181.70	-3.57	-2.02	36.004	0.0513	Not significant	No difference

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

*df* = degrees of freedom.

<sup>a</sup> If there was a 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 F.8. Gardendale MAP reading growth results comparing fall 2024 and spring 2025 by grade level**

Grade level	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>
Kindergarten	39	139.59	155.41	15.82	-11.61	38	<.0001	Significant	Spring 2025
First	39	146.38	163.23	16.85	-10.82	38	<.0001	Significant	Spring 2025
Second	37	161.70	178.11	16.41	-10.15	36	<.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 for multiple hypothesis testing correction (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a 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 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 <sup>a</sup>
Science	Fall	167.96	176.20	-8.26	-5.13	48.098	<.0001	Significant	Normed
	Winter	177.69	181.40	-3.72	-2.54	48.105	0.0145	Significant	Normed
	Spring	181.06	185.30	-4.25	-2.90	48.100	0.0056	Significant	Normed

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

*df* = degrees of freedom.

<sup>a</sup> If there was a 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 F.10. Gardendale MAP science growth results comparing fall 2024 and spring 2025 by grade level**

Grade level	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>
Second	49	167.96	181.06	13.10	-9.56	48	<.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 for multiple hypothesis testing correction (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a 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 F.11. Gardendale mCLASS English results comparing score levels by grade over time by grade level**

Score level	Grade level	Sample size	Fall 2024 percentage	Winter 2024 percentage	Spring 2025 percentage	Growth (spring 2025–fall 2024)	W statistic	Initial <i>p</i> value	Adjusted significance	Time favored <sup>a</sup>
Well below benchmark	Kindergarten	40	42.50	32.50	20.00	-22.50	22.00	0.0009	Significant	Spring 2025
Below benchmark			7.50	12.50	5.00	-2.50				
At benchmark			27.50	25.00	35.00	7.50				
Above benchmark			22.50	30.00	40.00	17.50				
Well below benchmark	First	39	46.15	35.90	28.21	-17.95	0.00	0.0001	Significant	Spring 2025
Below benchmark			30.77	30.77	20.51	-10.26				
At benchmark			7.69	20.51	25.64	17.95				
Above benchmark			15.38	12.82	25.64	10.26				
Well below benchmark	Second	35	62.86	68.57	60.00	-2.86	20.00	0.7897	Not significant	No difference
Below benchmark			17.14	14.29	20.00	2.86				
At benchmark			<21.00	<18.00	<21.00	<3.0				
Above benchmark			<6.00	<9.00	<6.00	<3.0				
Well below benchmark	Total	114	50.00	44.74	35.09	-14.91	124.50	<.0001	Significant	Spring 2025
Below benchmark			18.42	19.30	14.91	-3.51				
At benchmark			17.54	19.30	26.32	8.77				
Above benchmark			14.04	16.67	23.68	9.65				

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

<sup>a</sup> If there was a 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 F.12. Gardendale mCLASS Spanish results comparing score levels by grade over time by grade level**

Score level	Grade level	Sample size	Fall 2024 percentage	Winter 2024 percentage	Spring 2025 percentage	Growth (spring 2025–fall 2024)	W statistic	Initial <i>p</i> value	Adjusted significance	Time favored <sup>a</sup>
Well below benchmark	Kindergarten	16	12.50	18.75	18.75	6.25	24.00	0.4070	Not significant	No difference
Below benchmark			<31.00	<26.00	<38.00	<7.00				
At benchmark			<59.00	<57.00	<52.00	<7.00				
Above benchmark			18.75	12.50	12.50	-6.25				
Well below benchmark	First	26	26.92	23.08	11.54	-15.38	0.00	0.0002	Significant	Spring 2025
Below benchmark			19.23	7.69	7.69	-11.54				
At benchmark			42.31	42.31	30.77	-11.54				
Above benchmark			11.54	26.92	50.00	38.46				
Well below benchmark	Second	14	<-36.00	<7.00	<7.00	<-29.00	0.00	0.0051	Significant	Spring 2025
Below benchmark			<-36.00	<7.00	<7.00	<-29.00				
At benchmark			35.71	50.00	21.43	-14.29				
Above benchmark			21.43	50.00	78.57	57.14				
Well below benchmark	Total	56	21.43	16.07	10.71	-10.71	75.00	<.0001	Significant	Spring 2025
Below benchmark			21.43	8.93	10.71	-10.71				
At benchmark			41.07	46.43	32.14	-8.93				
Above benchmark			16.07	28.57	46.43	30.36				

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

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

**Table F.13. Gardendale mCLASS growth results comparing fall 2024 and spring 2025 by language of assessment and grade level**

Language of assessment	Grade level	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>
English	Kindergarten	40	294.65	433.88	139.23	-30.68	39	<.0001	Significant	Spring 2025
	First	39	328.82	451.82	123.00	-26.06	38	<.0001	Significant	Spring 2025
	Second	35	311.86	418.17	106.31	-46.40	34	<.0001	Significant	Spring 2025
	Total	114	311.62	435.19	123.57	-46.34	113	<.0001	Significant	Spring 2025
Spanish	Kindergarten	16	281.56	397.44	115.88	-21.70	15	<.0001	Significant	Spring 2025
	First	26	361.08	471.69	110.62	-19.36	25	<.0001	Significant	Spring 2025
	Second	14	358.57	502.14	143.57	-25.66	13	<.0001	Significant	Spring 2025
	Total	56	337.73	458.09	120.36	-31.81	55	<.0001	Significant	Spring 2025
English & Spanish	Total	170	320.22	442.74	122.51	-56.29	169	<.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 for multiple hypothesis testing correction (Benjamini-Hochberg, 1995).

*df* = degrees of freedom.

<sup>a</sup> If there was a 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.”

# Appendix G

Additional Social-Emotional Assessment Results

## Appendix G

### Additional Social-Emotional Assessment Results

Appendix G provides additional social-emotional assessment results that address the seventh set of research questions. Teachers administered two different assessments depending on grade level. Teachers administered the Devereux Early Childhood Assessment (DECA) to pre-K children, and the Devereux Student Strengths Assessment (DESSA), mini and full<sup>24</sup> versions, to children in kindergarten through second grade. There are six tables provided. Tables G.1, G.3, and G.5 address both parts of 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.

---

<sup>24</sup> 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).

**Table G.1. Social-emotional (DECA) results for pre-K children comparing levels in fall 2024 and spring 2025 by outcome**

Outcome	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>
Initiative	Needs Instruction	52	19.23	5.77	-13.46	27.0	<.0001	Significant	Spring 2025
	Typical		75.00	51.92	-23.08				
	Strengths		5.77	42.31	36.54				
Self-Control	Needs Instruction	52	11.54	9.62	-1.92	65.0	0.0028	Significant	Spring 2025
	Typical		76.92	51.92	-25.00				
	Strengths		11.54	38.46	26.92				
Attachment	Needs Instruction	52	15.38	1.92	-13.46	26.0	<.0001	Significant	Spring 2025
	Typical		82.69	50.00	-32.69				
	Strengths		1.92	48.08	46.15				
Total Protective Factors	Needs Instruction	52	13.46	3.85	-9.62	0.0	<.0001	Significant	Spring 2025
	Typical		84.62	32.69	-51.92				
	Strengths		1.92	63.46	61.54				
Behavioral Concerns <sup>b</sup>	Typical	52	<95.00	<93.00	<2.00	0.0	1.0000	Not significant	No difference
	Area of Need		<14.00	<12.00	<2.00				

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

<sup>a</sup> If there was a 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. 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 2024 and spring 2025 by outcome**

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>
Initiative	52	47.50	57.62	10.12	9.825	51	<.0001	Significant	Spring 2025
Self-Control	52	49.06	56.90	7.85	6.438	51	<.0001	Significant	Spring 2025
Attachment	52	46.46	59.27	12.81	8.331	51	<.0001	Significant	Spring 2025
Total Protective Factors	52	47.37	59.12	11.75	11.432	51	<.0001	Significant	Spring 2025
Behavioral Concerns	52	45.35	44.63	-0.71	-0.809	51	0.4222	Not significant	No difference

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

*df* = degrees of freedom.

<sup>a</sup> If there was a 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 G.3. Social-emotional (mini-DESSA) results for kindergarten through second-grade children comparing levels in fall 2024 and spring 2025**

Outcome	Level	Sample size	Fall 2024 percentage	Spring 2025 percentage	Difference (spring 2025–fall 2024)	W statistic	Initial p value	Adjusted significance	Time favored <sup>a</sup>
Overall Total	Needs Instruction	154	<2.00	<5.00	<7.00	1,062.0	0.0667	Not significant	No difference
	Typical		<62.00	<64.00	<2.00				
	Strengths		40.26	35.71	-4.55				

**Note:** Because of rounding, decimals may not agree to the nearest hundredths. Because of masking, for some results we provide a range. The Adjusted significance column indicates significance levels (*p* values) after adjustment for multiple hypothesis testing correction (Benjamini-Hochberg, 1995).

<sup>a</sup> If there was a statistically 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 G.4. Social-emotional (mini-DESSA) growth results for kindergarten through second-grade children comparing levels in fall 2024 and spring 2025**

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>
<b>Overall Total</b>	154	56.28	55.92	-0.36	0.418	153	0.6766	Not significant	No difference

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

*df* = degrees of freedom.

<sup>a</sup> If there was a 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 G.5. Social-emotional (full DESSA) results for kindergarten through second-grade children comparing levels in fall 2024 and spring 2025**

Outcome	Level	Sample size	Fall 2024 percentage	Spring 2025 percentage	Difference (spring 2025–fall 2024)	W statistic	Initial <i>p</i> value	Adjusted significance	Time favored <sup>a</sup>
Self-Awareness	Needs Instruction	17	58.82	52.94	-5.88	20.0	0.7897	Not significant	No difference
	Typical		<48.00	<61.00	<13.00				
	Strengths		<14.00	<21.00	<7.00				
Self-Management	Needs Instruction	17	64.71	52.94	-11.76	8.0	0.2986	Not significant	No difference
	Typical		<43.00	<55.00	<12.00				
	Strengths		<7.00	<13.00	<6.00				
Social Awareness	Needs Instruction	17	<50.00	<31.00	<-19.00	0.0	0.0719	Not significant	No difference
	Typical		<72.00	<79.00	<7.00				
	Strengths		<7.00	<26.00	<-19.00				
Relationship Skills	Needs Instruction	17	<37.00	<50.00	<13.00	6.0	0.7656	Not significant	No difference
	Typical		70.59	52.94	-17.65				
	Strengths		<7.00	<12.00	<19.00				
Decision-Making	Needs Instruction	17	76.47	52.94	-23.53	4.0	0.0726	Not significant	No difference
	Typical		<31.00	<56.00	<25.00				
	Strengths		<7.00	<20.00	<13.00				
Optimistic Thinking	Needs Instruction	17	<54.00	<67.00	<13.00	9.0	0.7656	Not significant	No difference
	Typical		52.94	47.06	-5.88				
	Strengths		<7.00	<14.00	<7.00				
Overall Total	Needs Instruction	17	76.47	52.94	-23.53	10.0	0.1096	Not significant	No difference
	Typical		<31.00	<56.00	<25.00				
	Strengths		<7.00	<20.00	<13.00				

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

<sup>a</sup> If there was a 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 G.6. Social-emotional (full DESSA) growth results for kindergarten through second-grade children comparing levels in fall 2024 and spring 2025**

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>
Self-Awareness	17	40.06	41.18	1.12	0.461	16	0.6513	Not significant	No difference
Self-Management	17	38.12	40.65	2.53	1.203	16	0.2464	Not significant	No difference
Social Awareness	17	42.41	45.12	2.71	1.423	16	0.1739	Not significant	No difference
Relationship Skills	17	41.00	44.18	3.18	1.539	16	0.1433	Not significant	No difference
Decision-Making	17	37.88	41.29	3.41	1.460	16	0.1638	Not significant	No difference
Optimistic Thinking	17	37.94	39.71	1.76	0.664	16	0.5163	Not significant	No difference
<b>Overall Total</b>	17	37.53	40.18	2.65	1.216	16	0.2414	Not significant	No difference

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

*df* = degrees of freedom.

<sup>a</sup> If there was a significant difference, the column lists the time whose score was greater (i.e., the “favored” time) in this column. If there was no statistically significant difference, this column states 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., & 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>.
- 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](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.
- 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.
- 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. (2025). *MAP growth: 2025 detailed research study*. HMH Education Company. <https://www.nwea.org/resource-center/white-paper/88182/MAP-Growth-Norms-Technical-Manual.pdf/>
- Robitaille, J. L., Johnson, E. S., LeBuffe, P. A., Naglieri, J. A., & Shapiro, V. B. (2024). *The Devereux Student Strengths Assessment, Second Edition Technical Manual (DESSA 2)*. Riverside Insights.
- 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
- 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*, 27(1), 122–54.