

# Pre-K 4 SA and Gardendale Early Learning Program: Year 6 (2024–2025) Technical Evaluation Report

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, nor the City of San Antonio. Mention of trade names, commercial products, or organizations does not imply endorsement by the City of San Antonio.*

## Table of Contents

|   |                                     |
|---|-------------------------------------|
| <b>Executive Summary</b>  | <b>iii</b>                          |
| <b>Introduction</b>   | <b>1</b>                            |
| <b>Research Questions</b>   | <b>2</b>                            |
| <b>Evaluation Sample and Methods</b>  | <b>3</b>                            |
| Sample  | 3                                   |
| Methods   | 4                                   |
| <b>Evaluation Results</b>   | <b>6</b>                            |
| Teacher–Child Interaction Quality   | 6                                   |
| Kindergarten and First-Grade Readiness: Growth, Observation, and Learning (GOLD) Results                            | 7                                   |
| Pre-K (Kindergarten Readiness) Results  | 7                                   |
| Kindergarten (First-Grade Readiness) Results  | 8                                   |
| Pre-K and Kindergarten Growth Results   | 9                                   |
| Kindergarten Readiness Results  | 9                                   |
| Direct Child Assessments  | 11                                  |
| Woodcock-Johnson (WJ IV) and Batería  | 11                                  |
| Letter-Word High-Achieving Subgroup   | 13                                  |
| <b>Receptive One-Word Picture Vocabulary Test (ROWPVT) and Expressive One-Word Picture Vocabulary Test (EOWPVT)</b> | <b>15</b>                           |
| Measures of Academic Progress (MAP)   | 16                                  |
| mCLASS  | 18                                  |
| Social-Emotional Assessment   | 19                                  |
| Devereux Early Childhood Assessment (DECA)  | 19                                  |
| Devereux Student Strengths Assessment (DESSA)   | 20                                  |
| <b>Conclusions and Looking Ahead</b>  | <b>21</b>                           |
| Overview of Findings  | 21                                  |
| Comparing Assessment Results Across Measures  | 22                                  |
| Directions for Future Research  | <b>Error! Bookmark not defined.</b> |
| <b>Limitations and Recommendations</b>  | <b>26</b>                           |

## Table of Contents (continued)

### References

R-1

### Tables

|          |   |    |
|----------|---|----|
| Table 1. | Demographics of the overall sample                      | 3  |
| Table 2. | Summary of early literacy findings across assessments   | 24 |
| Table 3. | Summary of early numeracy findings across assessments   | 25 |
| Table 4. | Summary of social-emotional findings across assessments | 26 |

### Figures

|            |   |    |
|------------|---|----|
| Figure 1.  | Year 6 average classroom quality scores for Gardendale by CLASS domain  | 7  |
| Figure 2.  | Percentage of pre-K Gardendale children demonstrating kindergarten readiness in spring 2025 by GOLD outcome               | 10 |
| Figure 3.  | Percentage of kindergarten Gardendale children demonstrating kindergarten readiness in fall 2024 by GOLD outcome          | 11 |
| Figure 4.  | Percentage of Gardendale children meeting age equivalency by subtest and assessment time point                            | 12 |
| Figure 5.  | Size of achievement gap (in months) between Gardendale and normative sample by assessment and time                        | 13 |
| Figure 6.  | Size of difference in months between high-achieving Gardendale subgroup and normative sample by assessment and time point | 14 |
| Figure 7.  | Percentage of children within each performance level for receptive vocabulary by assessment time                          | 15 |
| Figure 8.  | Percentage of children within each performance level for receptive vocabulary by assessment time                          | 16 |
| Figure 9.  | Percentage of children within each performance level based on mCLASS literacy (English) by assessment time point          | 18 |
| Figure 10. | Percentage of children within each performance level based on mCLASS literacy (Spanish) by assessment time point          | 19 |

## Executive Summary

Pre-K 4 SA and Edgewood Independent School District have engaged in a partnership by opening, in 2019, a Pre-K 4 SA replication site at Gardendale Elementary School serving pre-kindergarten (pre-K) through second grade. The San Antonio Early Childhood Education Municipal Development Corporation contracted with Westat, a large employee-owned global research firm, to conduct an annual independent evaluation. Westat analyzed data collected by Pre-K 4 SA, Gardendale, and a team of observers who conducted classroom quality observations that included direct observations, teacher report assessments, and direct child achievement assessments.

The evaluation investigated the demographic characteristics of the children served, classroom quality, and children's learning during the 2024–25 school year in Gardendale. During its sixth year, the partnership served more than 280 children. Of those children served, 58.3 percent were boys and 41.7 percent were girls, and a large majority (95.8 percent) were Hispanic.

Taken together, the results from the Year 6 evaluation suggest children served by the partnership experienced positive classroom environments and are benefiting from participation in Gardendale.

Specifically, results indicated the following key conclusions:

- Partnership classrooms provided high levels of Emotional Support and Classroom Organization, and midrange levels of Instructional Support.
- Gardendale children in pre-K and kindergarten demonstrated significant improvement in six readiness outcomes: Cognitive, Literacy, Mathematics, Oral Language, Physical, and Social-Emotional.
- Gardendale children showed significant improvement in indicators of early literacy and early numeracy; however, most children performed below their age level compared to a normative sample. We observed significant accelerated learning in early literacy, and significant accelerated learning in early numeracy for a small subgroup of children.
- Gardendale children showed significant improvement in indicators of receptive and expressive vocabulary, mathematics, and science. Gardendale children in pre-k showed significant improvement in social-emotional competency over time.

## Introduction

The benefit of providing high-quality early childhood education has received, and will continue to receive, considerable attention throughout the United States (Barnett, 2011; Campbell et al., 2002; Coogle et al., 2021; Heckman et al., 2010; Hill et al., 2015; Reynolds et al., 2011; Rolnick & Grunewald, 2003). Yet children who would benefit from high-quality education experiences do not have the opportunity to receive them. For example, children from racially marginalized communities, children from economically disadvantaged backgrounds, and children whose primary language is not English are exposed to lower quality instruction and learning environments across the United States more often than other groups of children (Bassok & Galdo, 2016; Valentino, 2018). Moreover, providing high-quality learning environments is vital to improving children's social-emotional, behavior, and achievement outcomes (Burchinal et al., 2010; Lippard et al., 2018; Martinsone et al., 2022; Perlman et al., 2016).

Because of limited public funding from federal and state governments, municipal governments are increasingly using funding in creative ways to provide more children with access to high-quality early childhood education and care. For example, in Texas, some districts will increase their funding by engaging with selected partners to improve child outcomes. In 2019, Pre-K 4 SA and Edgewood Independent School District engaged in a partnership (hereafter referred to as Gardendale) focusing on the expertise of Pre-K 4 SA to provide innovative early learning environments and support to children from kindergarten through second grade attending Gardendale. Pre-K 4 SA has used their learning model, with demonstrated positive results, to train the teaching staff on how to provide high-quality, evidence-based programming to Gardendale children (Decker-Woodrow et al., 2018; Decker-Woodrow et al., 2017; Decker-Woodrow et al., 2019; Decker-Woodrow & Price, 2016; Diaz et al., 2024; Diaz et al., 2023; Diaz et al., 2022; Edvance Research, 2015; Edvance Research, 2014; Villareal, 2019).

Because of the timing of the partnership, most children in Gardendale who attended during the 2024–25 school year were either born or had started their education journey during the COVID-19 pandemic. Research indicates that during the pandemic, young children nationwide experienced instruction losses and decreases in their social-emotional development and well-being compared to the years before the pandemic (Jung & Barnett, 2021; Weiland et al., 2021). Based on the National Assessment of Education Progress test scores, students on average experienced one-half (–0.494) of a grade level of learning loss in math and almost a third (–0.309) of a grade level of learning loss in reading between 2019 and 2022 (Fahle et al., 2023). When comparing this to student learning trends prior to the pandemic, students recovered 20–30 percent of learning loss in the 1st year but did not make any further recovery in the subsequent 3–4 years (Center for Education Policy Research, 2023). Therefore, given these disruptions and setbacks, there is a need for accelerated learning and additional education supports as many children have experienced educational achievement delays in their understanding (Socol, 2022). As of spring 2024, students are on average roughly half a grade level behind pre-pandemic achievement in mathematics and reading. Compared to 2022, students are even further behind in reading (Dewey et al., 2025). Moreover, the average student educated during the pandemic is less than halfway to a full academic recovery (CRPE, 2024).

Previous external evaluation reports of Gardendale suggested most children were not performing at their age level in early literacy and early numeracy in recent years (Diaz & Decker-Woodrow, 2021; Diaz et al., 2023). This suggests that the pandemic-era challenges in early learning were relevant locally in Gardendale as well. Given these challenges, it is imperative to understand how early

childhood initiatives and collaborative partnerships are supporting children nationwide in the subsequent years moving beyond the pandemic.

The San Antonio Early Childhood Education Municipal Development Corporation contracted with Westat, a large employee-owned global research firm, to conduct an independent evaluation of Gardendale. This report marks the 6th year of the partnership (2024–25 school year) and complements the previous reports provided. This report begins with a summary of the evaluation sample and methods. Then it presents Year 6 evaluation findings for the program. The results section describes information on classroom quality, then presents outcome analysis results of (1) the Teaching Strategies’ Growth, Observation, and Learning (GOLD) assessment; (2) children’s early literacy, early numeracy, and receptive and expressive vocabulary; and (3) children’s social-emotional development. We then summarize, synthesize, and compare all the findings across assessments, followed by limitations. The final section of this report outlines directions for future research.

This report is one in a series of reports documenting results of Gardendale during the 2024–25 school year. Alongside this report, a research brief includes a high-level summary of the findings, and a supplement provides more detailed and technical information.

## Research Questions

The Year 6 (2024–25) evaluation of Gardendale addressed the following main research questions and sub questions:

1. What was the observed teacher–child interaction quality of Gardendale classrooms in Year 6?
2. A. How did pre-K and kindergarten Gardendale children compare to the normative sample on GOLD outcomes?  
B. Did pre-K and kindergarten Gardendale children demonstrate significant improvement on GOLD outcomes?  
C. What percentage of pre-K and kindergarten Gardendale children demonstrated kindergarten readiness as measured by GOLD outcomes?
3. A. 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?  
B. Did Gardendale children demonstrate significant improvement in early literacy and early numeracy?  
C. Did Gardendale children experience accelerated learning to help narrow achievement gaps in early literacy and early numeracy?
4. A. What were the receptive and expressive vocabulary performance levels of Gardendale children, and to what extent did the performance levels change over the year?  
B. Did Gardendale children demonstrate significant growth in receptive and expressive vocabulary?
5. A. How did Gardendale children in kindergarten through second grade compare to the normative sample on Measures of Academic Progress (MAP) mathematics and reading?  
B. Did Gardendale children in kindergarten through second grade demonstrate significant improvement on MAP mathematics and reading?

- C. How did second-grade Gardendale children compare to the normative sample on MAP science?
- D. Did second-grade Gardendale children demonstrate significant improvement on MAP science?
6. A. What were the performance levels of Gardendale children in kindergarten through second grade in mCLASS literacy?
- B. Did Gardendale children in kindergarten through second grade demonstrate significant improvement in mCLASS literacy?
7. A. What were the levels of Gardendale children’s social-emotional competence, and to what extent did the levels change?
- B. Did Gardendale children demonstrate significant improvement in social-emotional competence?

## Evaluation Sample and Methods

This section provides descriptive information about the classrooms and the demographic characteristics of the children served during the 2024–25 school year, as well as a brief discussion of the methods used.

### Gardendale Year 6 (2024-25) Sample Characteristics: Children

Westat analyzed data for 283 children attending Gardendale. Table 1 includes the demographics for the sample. Children were in pre-K (23.7 percent of total sample) through second grade (23.0 percent of total sample), with most children in the first grade (28.3 percent of total sample). There were more boys (58.3 percent) than girls (41.7 percent). Most children were Hispanic (95.8 percent), were economically disadvantaged (74.6 percent), did not receive special education services (82.7 percent), and were not receiving English Learner services (72.8 percent). Regarding the services children received, 17.3 percent received special education services and 27.2% received English Learner services.

| Demographic characteristic                  | N (percentage)      |             |
|---|---------------------|-------------|
| <b>Gender</b>                               | Male                | 165 (58.3%) |
|   | Female              | 118 (41.7%) |
| <b>Grade level</b>                          | Pre-K               | 67 (23.7%)  |
|   | Kindergarten        | 71 (25.1%)  |
|   | First               | 80 (28.3%)  |
|   | Second              | 65 (23.0%)  |
| <b>Race/ethnicity</b>                       | Hispanic/Latino     | 271 (95.8%) |
|   | Not Hispanic/Latino | 12 (4.2%)   |
| <b>Economically disadvantaged</b>           | Yes                 | 211 (74.6%) |
|   | No                  | 72 (25.4%)  |
| <b>Receiving special education services</b> | Yes                 | 49 (17.3%)  |
|   | No                  | 234 (82.7%) |
| <b>Receiving English Learner services</b>   | Yes                 | 77 (27.2%)  |
|   | No                  | 206 (72.8%) |

**Note:** Because of rounding decimals to the nearest tenths, totals for any given characteristic may not sum to 100 percent.

## Gardendale Year 6 (2024-25) Sample Characteristics: Classrooms

---

We included data from a total of 16 classrooms with observation data from the spring to answer the first research question pertaining to teacher-child interaction quality. Three of the classrooms were pre-K (18.8 percent of the total classrooms assessed), four were kindergarten (25.0 percent of the total classrooms assessed), five were first grade (31.3 percent of the total classrooms assessed), and four were second grade (25.0 percent of the total classrooms assessed).

### Methods

---

Westat addressed all research questions by analyzing existing Pre-K 4 SA and Gardendale databases and classroom observations. To answer the first question, Westat and its partners collected data collected and Westat analyzed descriptively from the Classroom Assessment Scoring System (CLASS), second edition, (Teachstone, 2023) for Gardendale classrooms. CLASS is an observational system that assesses classroom practices by measuring the interactions between children and adults. Observers assign scores during various classroom activities over the course of a morning. Westat averaged those across all observation periods (or cycles) for overall quality scores in three domains (which encompass 10 different dimensions in total). The CLASS domains and dimensions are

- Emotional Support domain (positive climate, negative climate, educator sensitivity, and regard for child perspectives dimensions)
- Classroom Organization domain (productivity, behavior management, and instructional learning formats)
- Instructional Support (concept development, quality of feedback, and language modeling)

CLASS uses a 7-point Likert-type scale, in which a score of 1 or 2 indicates low-range quality; a score of 3, 4, or 5 indicates midrange quality; and a score of 6 or 7 indicates high-range quality. Observers assign each dimension and domain a score during each of five 20-minute cycles. Observers also record the number of children and adults in the classroom during each of the five 20-minute cycles. (See Appendix A for more detailed information.)

To address the second set of research questions, Westat used descriptive and inferential analyses on the Growth, Observation, and Learning (GOLD) outcomes. GOLD is a teacher-reported measure that collects information on children's progress 3 times throughout the school year on 36 objectives across 6 main categories: Cognitive, Literacy, Oral Language, Mathematics, Physical, and Social-Emotional. It was not possible to collect data on a comparison or control group, we conducted comparisons using the nationally representative normed data based on age bands for the GOLD assessment (Lambert, 2020; see Appendix A for more detailed information).

To address the third set of questions, Westat descriptively and inferentially analyzed data submitted by Pre-K 4 SA. Pre-K 4 SA administered two direct assessments, early literacy (Letter-Word) and early numeracy (Applied Problems), to a random sample<sup>1</sup> of Gardendale children in the fall and

---

<sup>1</sup> Westat pulled a stratified random sample based on grade level and classroom language(s) of instruction (monolingual and bilingual) to representative of all Gardendale children and submitted to Pre-K 4 SA. Demographics tests of differences revealed no significant differences on a range of demographic variables between the sample and population. (For more information see the Evaluation Methods in the Supplemental Appendix.)

spring. These two assessments are subtests from the Woodcock-Johnson IV Tests of Achievement (WJ IV; Schrank et al., 2014) and matching subtests from the Bateria III Spanish assessment (Muñoz-Sandoval et al., 2005; see Appendix A for more detailed information). Pre-K 4 SA chose them based on a Westat recommendation because early childhood research uses these assessments and they complement the GOLD findings by providing additional insights from a different perspective: that of a trained assessor as compared to a teacher report (Bloom & Weiland, 2014; McCormick, 2022; Puma et al., 2010; Weiland, 2016). The GOLD findings provide an overall perspective and measure multiple aspects of early literacy (e.g., phonological awareness, phonics, and word recognition) and numeracy (e.g., number concepts and operations, spatial relationships and shapes, and knowledge of patterns). Letter-Word findings are more nuanced and measure symbolic learning and the identification of isolated letters and words, while Applied Problems measures a child's ability to apply simple number concepts and solve math problems.

To address the fourth set of questions, Pre-K 4 SA submitted data to Westat for descriptive and inferential analyses. Two direct assessments of vocabulary were administered to a random sample<sup>2</sup> of Gardendale children in fall and spring : the Receptive One-Word Picture Vocabulary Test and the Expressive One-Word Picture Vocabulary Test, in English and Spanish-Bilingual editions (ROWPVT and EOWPVT; Martin, 2013a, 2013b; Martin & Brownell; 2011a; 2011b; see Appendix A for more detailed information). For the ROWPVT, the assessor presents the children with pictorial images of words and asks them to select the image that matches the word said by the examiner. For the EOWPVT, the assessor presents children with one image at a time and asks them to name what is displayed in the picture. As with the WJ, Pre-K 4 SA chose these assessments based on a Westat recommendation because early childhood research uses them, and they complement the GOLD findings by providing additional insights from a trained assessor as compared to a teacher report (Ingvalson et al., 2023; Vance et al., 1989). While the GOLD findings provide an overall perspective and measure multiple aspects of early literacy and numeracy, the ROWPVT and EOWPVT findings are more nuanced and measure receptive and expressive vocabulary knowledge and understanding.

To address the fifth set of questions, Pre-K 4 SA submitted data collected by Gardendale to Westat for descriptive and inferential analysis. Gardendale administered an electronic assessment of mathematics, reading, and science, the Measures of Academic Progress (MAP; NWEA, 2025) to children in kindergarten through second grade three times throughout the school year. Participating children entered their responses on an iPad. We included children in analyses<sup>3</sup> if they had outcome data for all three assessment times. It was not possible to collect collected data from a comparison or control group because of resource constraints so we conducted comparisons using the nationally representative normed data (NWEA, 2025; see Appendix A for more detailed information).<sup>4</sup>

To address the sixth set of questions, Pre-K 4 SA submitted data collected by Gardendale to Westat for descriptive and inferential analysis. Gardendale administered an assessment of early literacy, the mCLASS, to children in kindergarten through second grade three times throughout the school year. It is based on the Science of Reading, uses a one-on-one observational model, and measures

---

<sup>2</sup> Westat pulled a second stratified random sample based on grade level and classroom language(s) of instruction (monolingual and bilingual) and submitted to Pre-K 4 SA. (For more information see the Evaluation Methods in the Supplemental Appendix.)

<sup>3</sup> We conducted demographic tests of differences to determine if the sample of children included in and excluded from analyses were similar. (See Appendix A, Analytic Approach for more detailed information.)

<sup>4</sup> Gardendale children assessed in English and in Spanish for mathematics were compared to the normative sample based on grade level (NWEA, 2025). Gardendale children assessed in Spanish for reading were not compared to the normative sample because the Spanish normative sample for reading is not nationally representative.

phonemic awareness, phonics, fluency, vocabulary, and comprehension. Gardendale administered an online assessment to children in kindergarten, and for first and second grade, they administered a reading booklet to children, and their teacher entered their responses into the data system (Biancarosa et al., 2021). We conducted analyses for children with outcome data for all three assessment times.<sup>5</sup> (See Appendix A for more detailed information.)

To address the seventh set of questions, Pre-K 4 SA submitted data collected by Gardendale to Westat for descriptive and inferential analysis. Gardendale administered a teacher-reported assessment of social-emotional competence, the Devereux Early Childhood Assessment, second edition (DECA; LeBuffe & Naglieri, 2012), to children in pre-K in the fall and spring, and the Devereux Student Strengths Assessment, second edition (DESSA; Robitaille et al., 2024), to children in kindergarten through second grade in the fall and spring. The DECA and DESSA use a strengths-based approach to assessment, and as such, they focus on building children’s social-emotional strengths. They also emphasize the importance of promoting children’s social-emotional competency because that contributes to building their resilience to overcome adversity. (See Appendix A for more detailed information.)

## Evaluation Results

### Teacher–Child Interaction Quality

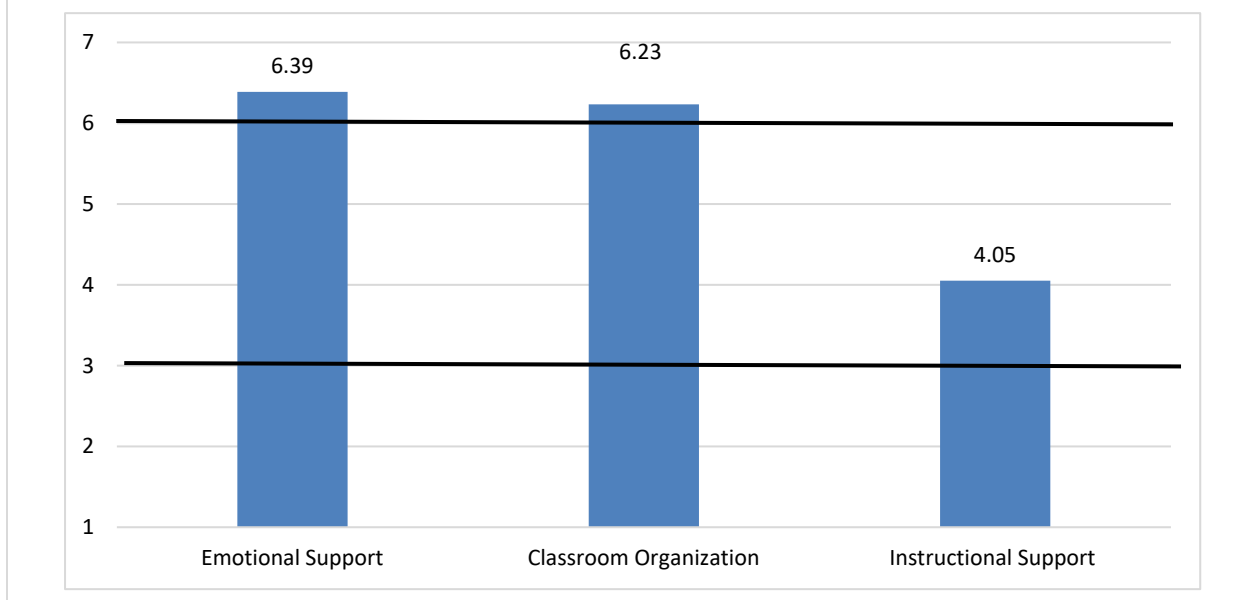
We present results for the first research question (What was the overall observed teacher–child interaction quality in Gardendale classrooms across Year 6?) in Figure 1.<sup>6</sup> Across the four grade levels, observers evaluated Gardendale classrooms ( $n = 16$ ) using CLASS, second edition, (Teachstone, 2023) during Year 6. Teacher–child interaction quality consists of three domains in the CLASS; specific findings are as follows and also appear in Figure 1.

- Scores for the **Emotional Support** domain ranged from 6.1 to 7.0 on a 1–7 scale, with an average score of 6.4; most scores were in the near-high or high range, suggesting observed teacher–child interactions in this domain were most often rated as near-high quality.
- The **Classroom Organization** domain scores ranged from 5.3 to 6.9, with an average score of 6.2, which suggests classrooms showed effective interactions regarding Classroom Organization.
- Finally, **Instructional Support** domain scores ranged from 2.7 to 5.0, with an average score in the midrange (4.1), which suggests that in some observed interactions, teachers provided support that extended children’s thinking or asked questions that encouraged children to analyze and reason. The types of interactions captured within the Instructional Support domain include interactions that facilitate higher-order thinking and cognitive development, as well as providing optimal environments for children to hear and use language. It is important to note lower ranges of Instructional Support quality are common across the United States and internationally as these types of interactions are found to be especially challenging for teachers of young children (Bassok et al., 2021; La Paro et al., 2004; Lekhal et al., 2025; Locasale-Crouch et al., 2007; Maier et al., 2022; Mashburn et al., 2008; Purtell & Ansari, 2018).

<sup>5</sup> We conducted demographic tests of differences to determine if the sample of children included in and excluded from analyses were similar. (See Appendix A, Analytic Approach, for more detailed information.)

<sup>6</sup> We provide average ratings across all CLASS dimensions in Appendix B.

Figure 1. Year 6 average classroom quality scores for Gardendale by CLASS domain



**Note:** The black horizontal lines mark the boundaries of the three score ranges: low (below 3.00), mid- (between 3.00 and 5.99), and high (6.00 and above).

## Growth, Observation, and Learning (GOLD)

Pre-K children (77.6 percent of children;  $n = 52$ ) and kindergarten children (78.9 percent of children;  $n = 56$ ) were included in analyses<sup>7</sup> if they had GOLD data for all three assessment times in at least one of the following six outcomes: Cognitive, Literacy, Mathematics, Oral Language, Physical, and Social-Emotional.

This section presents results for research question 2A (How did pre-K and kindergarten Gardendale children compare to the normative sample on GOLD outcomes?). First, we present results separately for pre-K (kindergarten readiness) and then present the results for kindergarten children. We separate the analyses because the norms vary depending on grade level.

### Pre-K

In the fall, pre-K children were significantly below the normative sample on three of the six GOLD outcomes (Cognitive, Physical, and Social-Emotional); on par<sup>8</sup> with the normative sample on two of the six GOLD outcomes (Mathematics and Oral Language); and significantly above the normative sample on the remaining outcome (Literacy). For more detailed information, see Appendix C, Tables C.1a and C.1b.

**Spring results for the Cognitive, Physical, and Social-Emotional outcomes indicated the initial gaps between Gardendale children and the normative sample were eliminated by the end of**

<sup>7</sup> We conducted demographic tests of differences to determine if the sample of children included in and excluded from analyses were similar. (See Appendix A, Analytic Approach, for more detailed information.)

<sup>8</sup> Throughout the report, “on par” indicates that though Gardendale children’s scores were different from the normative sample, these differences were not statistically significant.

**the school year.** By spring, there were no significant differences between Gardendale children and the normative sample for these outcomes. When comparing the Gardendale pre-K children to the normative sample in fall and spring, the gap for each outcome is as follows:

- Cognitive decreased 39.1 points (from scoring approximately 24.3 scale score points below the normative sample to approximately 14.8 scale score points above).
- Physical decreased 27.1 scale score points (from scoring approximately 35.2 scale score points below the normative sample in the fall to 8.1 scale score points above).
- Social-Emotional decreased 16.1 scale score points (from scoring approximately 20.8 scale score points below the normative sample to scoring approximately 4.7 scale score points below).

Across all three assessment times, Gardendale children were similar to the normative sample in Oral Language. More information is needed to understand what mechanisms might be behind Gardendale children scoring similarly to the normative sample for this outcome across all three assessment times.

Spring results for the remaining two outcomes (Literacy and Mathematics) indicated Gardendale children were significantly above the normative sample.

- For Literacy, Gardendale children were significantly above the normative sample across all three time points. More information is needed to understand what mechanisms might be behind Gardendale children scoring significantly above the normative sample for this outcome across all three assessment times.
- For Mathematics, Gardendale children surpassed and decreased approximately 18.1 scale score points (from scoring approximately 1.3 scale score points below the normative sample to scoring approximately 16.9 scale score points above).

(See Appendix C, Tables C.1a and C.1b, for more information.)

## Kindergarten

In the fall, kindergarten children were below the normative sample on two GOLD outcomes (Literacy and Oral Language) and on par with the normative sample on the remaining four GOLD outcomes (Cognitive, Mathematics, Physical, and Social-Emotional).

Spring results indicated Gardendale children scored similarly to the normative sample across all three assessment times for three outcomes (Cognitive, Mathematics, and Physical). For the other three outcomes (Literacy, Oral Language, and Social-Emotional), kindergarten children were below the normative sample. When comparing the Gardendale kindergarten children to the normative sample in fall and spring for Literacy and Oral Language, the significant gap remained for each outcome and is as follows:

- Literacy decreased 14.5 points (from scoring approximately 16.9 scale score points below the normative sample to approximately 2.3 scale score points below).
- Oral Language increased 32.9 points (from scoring approximately 50.2 scale score points below the normative sample to approximately 83.1 scale score points below).

For the Social-Emotional, a significant gap developed between the fall and spring. Children finished the year significantly below the normative sample despite starting the year on par with the normative sample. The gap in Social-Emotional increased by approximately 2.2 scale score points

(from scoring approximately 8.8 scale score points below the normative sample to scoring approximately 11.1 scale score points below). More information is needed to understand what mechanisms might be behind why Gardendale children scored similarly to the normative sample in the fall but then scored significantly below the normative sample in the spring for Social-Emotional. (See Appendix C, Tables C.2a and C.2b, for more information.)

## **Pre-K and Kindergarten Growth**

Results for research question 2B (Did pre-K and kindergarten Gardendale children demonstrate significant improvement on GOLD outcomes?) indicated there was significant improvement from fall to spring for pre-K and kindergarten children across all six outcomes. For pre-K children, the growth ranged from 42.3 scale score points for the Literacy domain to 117.6 scale score points for the Cognitive domain. For kindergarten children, the growth ranged from 60.8 scale score points for the Oral Language domain to 91.7 scale score points for the Mathematics domain. (See Appendix C, Table C.3, for more information.)

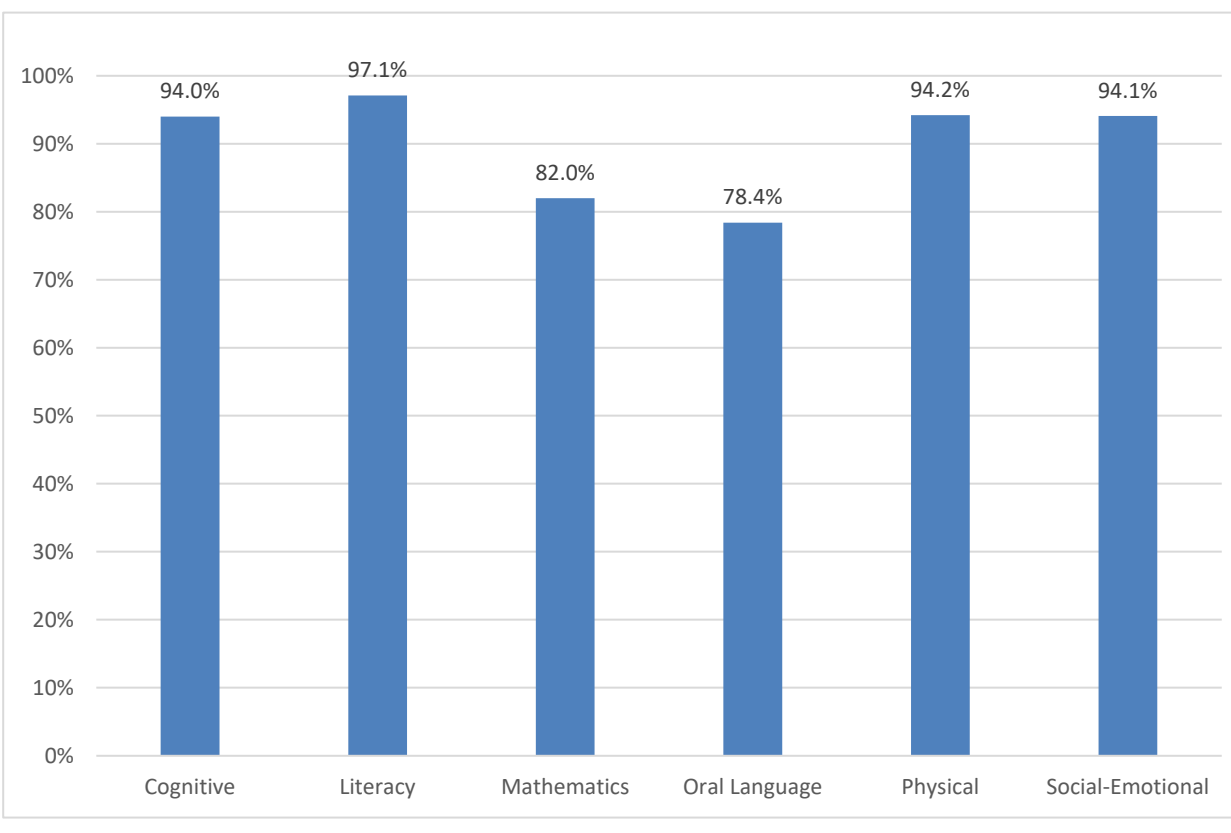
## **Kindergarten Readiness**

To find results for research question 2C (What percentage of Gardendale children demonstrated kindergarten readiness as measured by GOLD outcomes?), Westat conducted separate analyses for pre-K and kindergarten children to determine (1) whether pre-K children finished the year ready for kindergarten, and (2) whether kindergarten children started the year ready for kindergarten.

### **Pre-K Children**

Results indicated the majority of pre-K children demonstrated kindergarten readiness at the end of the year across all six outcomes. The readiness percentages ranged from 78.4 percent for the Oral Language domain to 97.1 percent for the Literacy domain, as shown in Figure 2. (See Appendix C, Table C.4, for more information.)

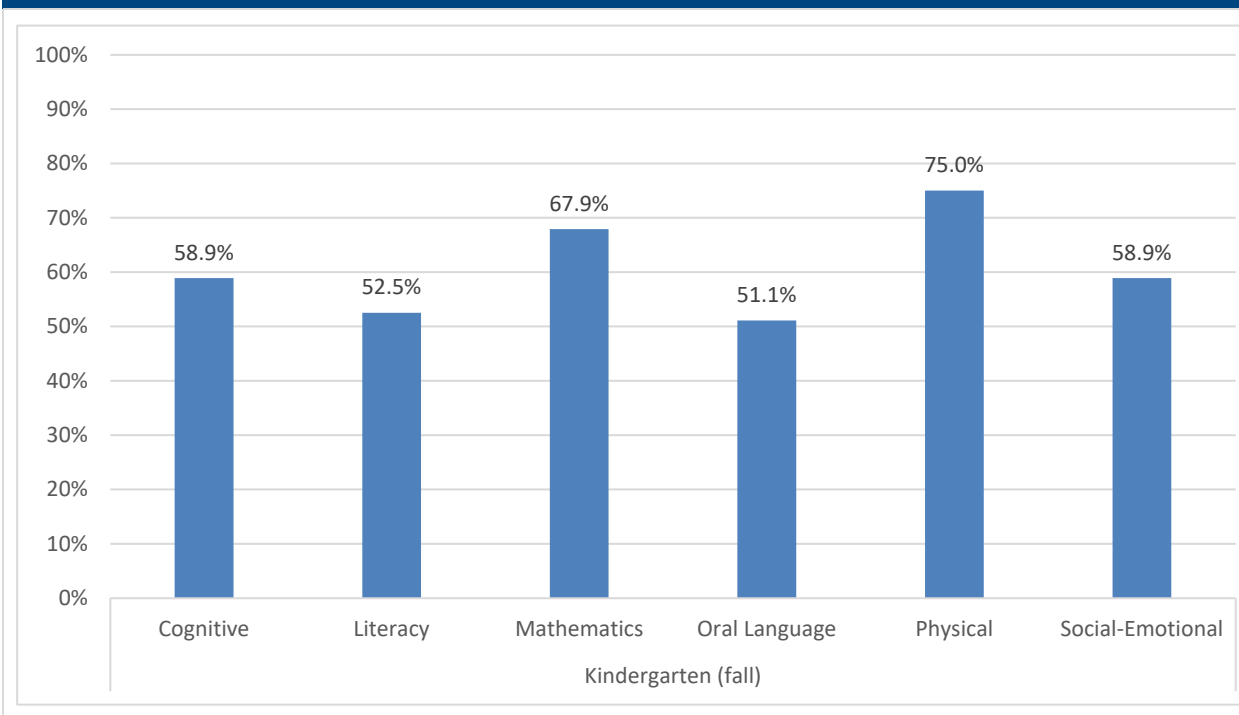
**Figure 2. Percentage of pre-K Gardendale children demonstrating kindergarten readiness in spring 2025 by GOLD outcome**



### Kindergarten Children

The majority of kindergarten children demonstrated readiness at the start of the year across all six outcomes. The readiness ranged from 51.1 percent for the Oral Language domain to 75.0 percent for the Physical domain, as shown in Figure 3. It is important to note that, of the children included in this analysis, 39 (69.6 percent) attended Gardendale for the first time and 17 (30.4 percent) attended for their second year. Therefore, it is possible some children did not start the year ready for kindergarten as the majority did not attend Gardendale last year. (See Appendix C, Table C.4, for more information.)

**Figure 3. Percentage of kindergarten Gardendale children demonstrating kindergarten readiness in fall 2024 by GOLD outcome**



## Direct Child Assessments

### Woodcock-Johnson (WJ) and Bateria

Westat analyzed data collected by Pre-K 4 SA from a random sample ( $n < 75$ ) on two subtests of a direct child achievement assessment: Letter-Word and Applied Problems from the WJ and the Bateria. Based on the results for Letter-Word, there were two patterns of early literacy achievement. The first pattern represented the majority of children ( $n = 59$ ), which we will present in this section. The second group represents a smaller subset ( $n < 15$ ), which demonstrated high achievement;<sup>9</sup> we will present these results in a later section. Based on the results for Applied Problems, there were also two patterns of early numeracy achievement. The first pattern represented the majority of children ( $n = 63$ ), which we will present in this section, and the second group represents a smaller subset ( $n < 10$ ) demonstrating high achievement, which we will present in a later section.

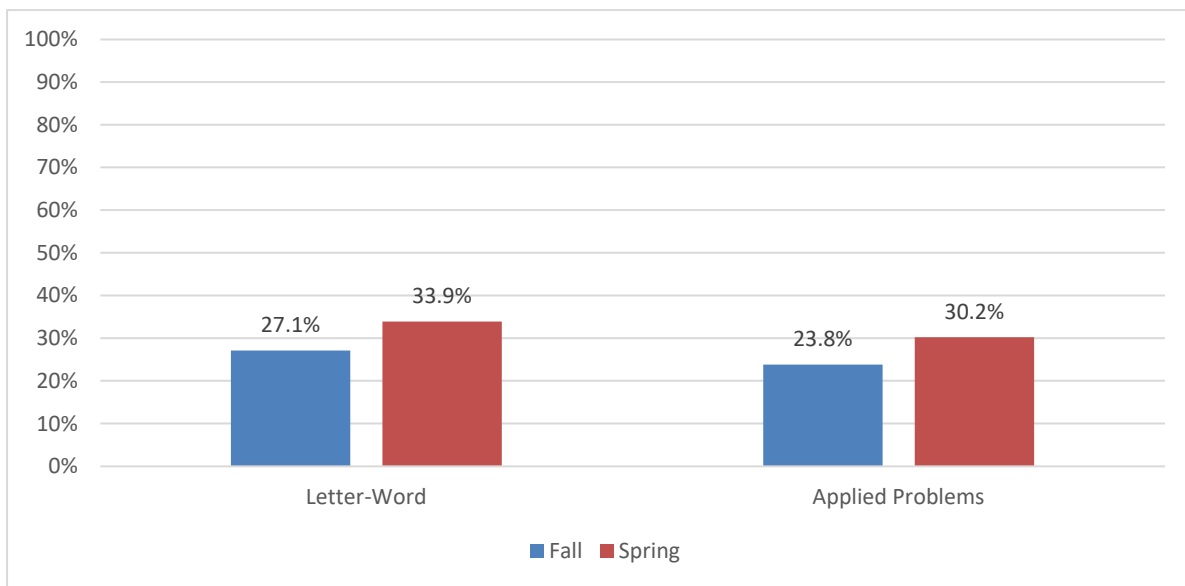
Results for the first part of research question 3A (What percentage of Gardendale children performed at or above their age level in early literacy and early numeracy?) showed that in the fall, 27.1 percent of children in Gardendale performed at or above their age level, and in the spring, 33.9 percent of children performed at or above their age level for early literacy, as shown in Figure 4. For early numeracy, in the fall, 23.8 percent of children in Gardendale performed at or above their age level, and in the spring, 30.2 percent of children performed at or above their age level. These findings imply that most children are performing below their age level at both assessment times

<sup>9</sup> We defined the high achievement subgroup to be children with more than 18 months of growth from fall to spring.

and are behind what would be considered “ready” for the next grade from a nationally representative lens.

However, for early numeracy, these levels in fall and spring are both more favorable than we observed in the Year 5 evaluation report for a slightly different group of students and one year ago.

**Figure 4. Percentage of Gardendale children meeting age equivalency by subtest and assessment time point**



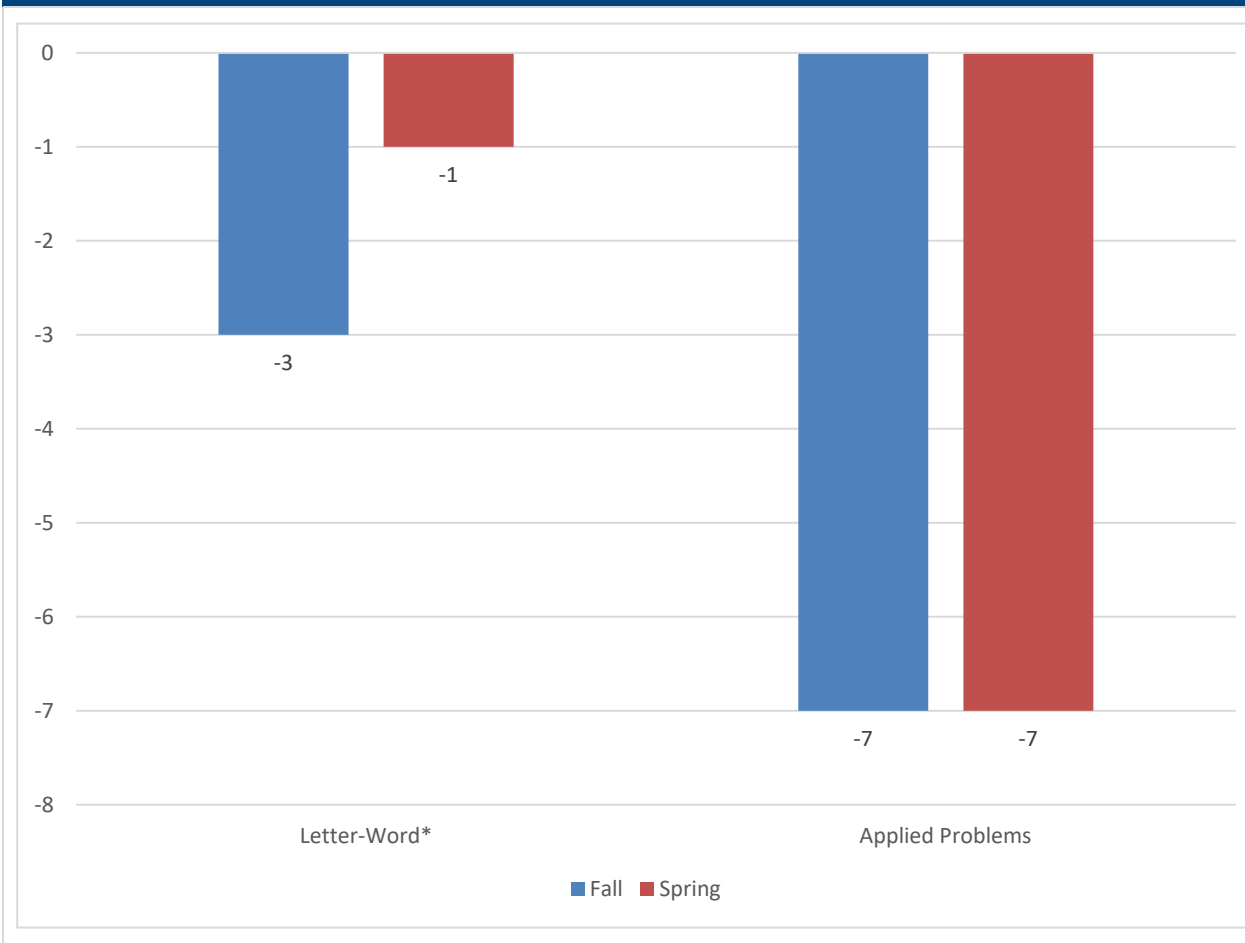
**Note:** Letter-Word measures early literacy skills, and Applied Problems measures early numeracy skills.

Results for the second part of research question 3A (To what extent did the percentage change?) showed no significant difference in children performing at or above their age level in early literacy and early numeracy in the spring compared to the fall. (For more detailed information and results by grade level, see Appendix D, Table D.1.)

Results for research question 3B (Did Gardendale children demonstrate significant improvement in early literacy and early numeracy?) revealed there was significant improvement for both outcomes. For early literacy, there was approximately 8 months of growth in learning in 6 calendar months. In early numeracy, there was approximately 7 months of growth in learning in 6 calendar months. Therefore, these findings suggest children gained significant knowledge of early literacy and numeracy from fall to spring. (For more detailed information and results by grade level, see Appendix D, Table D.2.)

Results for research question 3C (Did Gardendale children experience accelerated learning to help narrow achievement gaps in early literacy and early numeracy?) indicated that accelerated learning occurred for early literacy, but not early numeracy. For early literacy, children were on average 3 months below the norms in the fall, and in the spring, children were only 1 month below the norms (see Figure 5). In other words, the gap between Gardendale children and the national norm significantly decreased by 2 months for early literacy. For early numeracy, children were on average 7 months below the norms in both the fall and spring. It is worth noting that although this is not a positive finding, children’s performance did not decline further from fall to spring. (For more detailed information and results by grade level, see Appendix D, Table D.3.)

**Figure 5. Size of achievement gap (in months) between Gardendale and normative sample by assessment and time**



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

\* = Statistically significant increase ( $p < 0.05$ ).

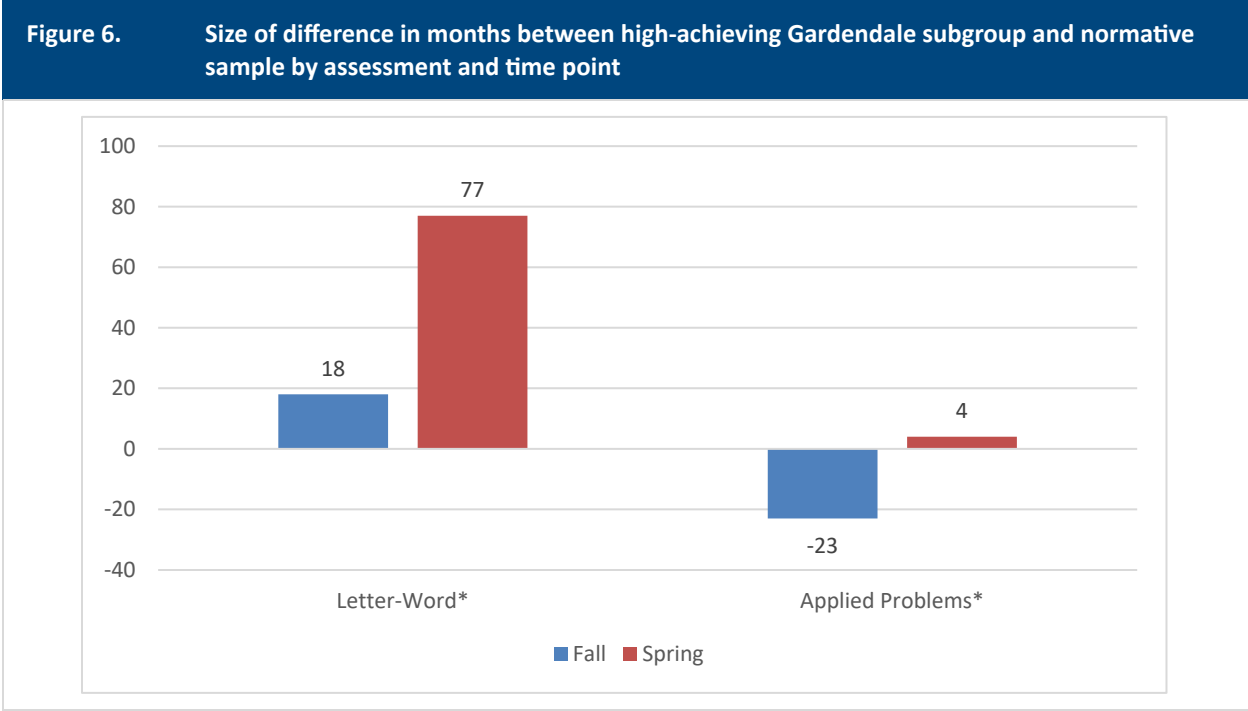
### High-Achieving Subgroup

In our evaluation, we found that there was a subgroup of children demonstrating high achievement in early literacy and early numeracy. We adapted research question 3A (What percentage of Gardendale children performed at or above their age level in early literacy?) to focus on two high-achieving subgroups ( $n < 15$  for early literacy and  $n < 10$  for early numeracy) of Gardendale children. Results showed the majority of the high-achieving early literacy subgroup tested at or above their age level for early literacy at both assessment times (80.0 percent in fall and 90.0 percent in spring) throughout the school year. Because of low sample sizes, we cannot report the results of the high-achieving early numeracy subgroup.

When applying the second part of research question 3A (To what extent did the percentage change?) to this high-achieving subgroup, results showed no significant difference in children performing at or above their age level in early literacy in the spring compared to fall. For early numeracy, it was not possible to conduct analyses because of low sample sizes. (For more detailed information and results by grade level, see Appendix D, Table D.4.)

Reframing research question 3B (Did Gardendale children demonstrate significant improvement in early literacy and early numeracy?) to focus on this high-achieving subgroup revealed that there was significant improvement. For early literacy, there was approximately 5 years and 5 months of growth in learning during 6 calendar months of time. For early numeracy, there was approximately 2 years and 9 months of growth in learning during 6 calendar months of time. Therefore, this finding suggests this small subgroup of children gained significant knowledge of early literacy and early numeracy from fall to spring: approximately eleven and five times the actual amount of time between assessments for early literacy and early numeracy, respectively. (For more detailed information and results by grade level, see Appendix D, Table D.5.)

Similarly, reframing research question 3C (Did Gardendale children experience accelerated learning to help narrow achievement gaps in early literacy and early numeracy?) to focus instead on this high-achieving subgroup indicated that significant accelerated learning did occur in early literacy and early numeracy. Given the small number of children, these findings do not represent all Gardendale children. In the fall, children were 18 total months ahead of the norms for early literacy on average, and in spring, children were 77 total months ahead of the norms (see Figure 6). In the fall, children were 23 total months behind the norms for early numeracy on average, and in spring, children were 4 total months ahead of the norms (see Figure 6). This implies this small subgroup of Gardendale children was outperforming the national norms at both assessment points and further surpassed the norms in the spring for early literacy. For early numeracy, this small subgroup of Gardendale children underperformed the national norms in the fall but surpassed the norms in the spring. Therefore, this small subgroup of Gardendale children surpassed the difference between their scores and the national norm by 4 years and 11 months for early literacy, and by 2 years and 3 months for early numeracy. (For more detailed information and results by grade level, see Appendix D, Table D.6.)



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

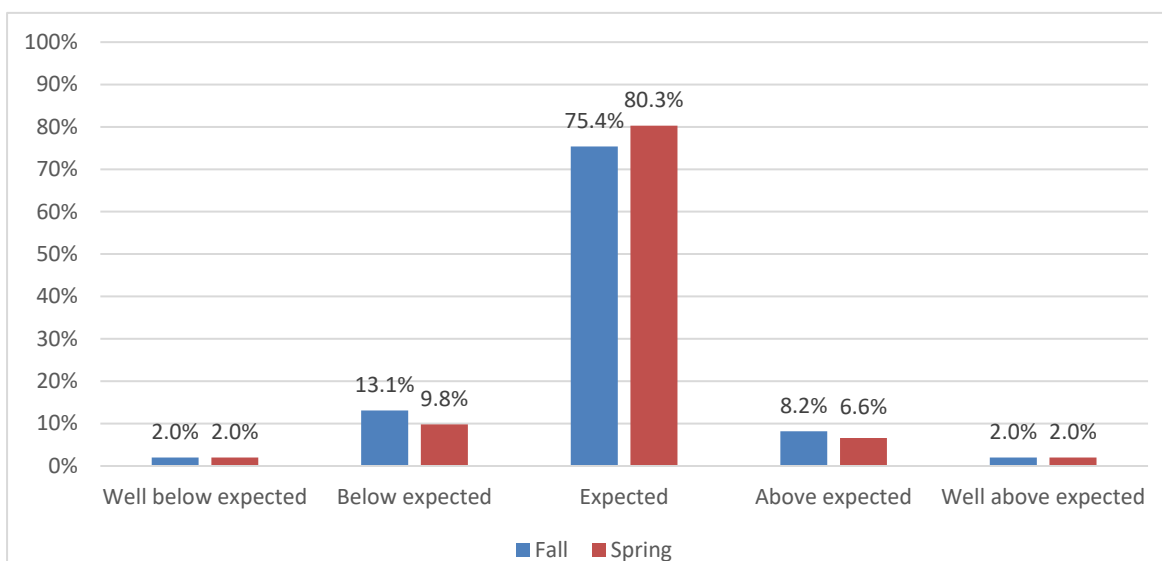
\* = Statistically significant increase ( $p < 0.05$ ).

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

To measure children’s receptive vocabulary, Westat analyzed data collected by Pre-K 4 SA from a random sample ( $n = 61$  for each assessment) on the ROWPVT and EOWPVT.<sup>10</sup> To evaluate children’s understanding, Westat converted their scores into five performance levels: (1) well below expected, (2) below expected, (3) expected, (4) above expected, and (5) well above expected. These levels are based on a normative sample and represent the developmental trajectory of children based on their age. To better understand how children were progressing throughout the year, we conducted analyses of vocabulary growth to assess changes over time and gain data points to compare to a normative sample. Together, these findings provided a holistic perspective of children’s vocabulary learning across the year.

Results for research question 4A (What were the receptive and expressive vocabulary performance levels of Gardendale children, and to what extent did the performance levels change over the year?) demonstrated children performed similarly in the fall and spring for both the receptive and expressive vocabulary tests. The majority of children were performing in the expected range in the fall and the spring for both the receptive (75.4 percent in fall and 80.3 percent in spring, see Figure 7) and expressive vocabulary tests (65.6 percent in fall and 63.9 in the spring, see Figure 8; for more detailed information and results by grade level, see Appendix E, Table E.1).

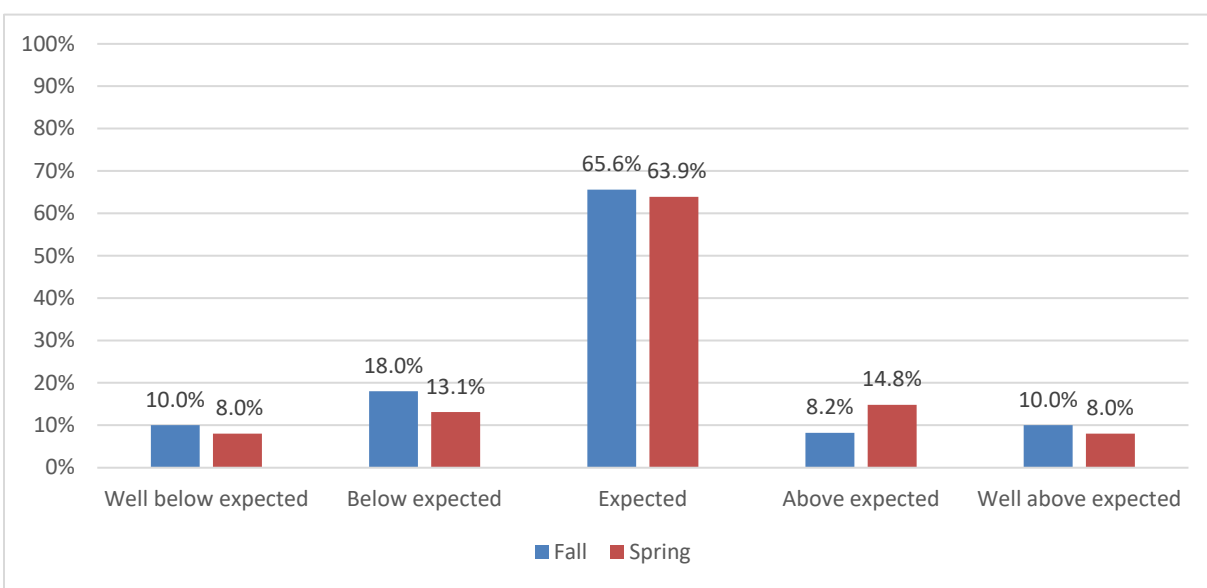
**Figure 7.** Percentage of children within each performance level for receptive vocabulary by assessment time



**Note:** Some percentages are rounded to protect confidential data, and totals may be more than 100%.

<sup>10</sup> This marks a transition in the vocabulary assessments data collection as this is first year collecting data for these assessments in English and Spanish. We previously used the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2019), which was only available in English. The Spanish version of the PPVT, Test de Vocabulario en Imágenes Peabody (Dunn et al., 1986), was discontinued by the publisher.

**Figure 8. Percentage of children within each performance level for expressive vocabulary by assessment time**



**Note:** Some percentages are rounded to protect confidential data, and totals may be more than 100%.

Results for research question 4B (Did Gardendale children demonstrate significant improvement in receptive and expressive vocabulary?) indicated children overall experienced significant improvement and gained additional vocabulary for both tests. There were approximately 3 (3.3) standard score points of growth for the receptive test and three (3.6) standard score points of growth for the expressive. (For more detailed information and results by grade level, see Appendix E, Table E.2.)

## Measures of Academic Progress (MAP)

Gardendale uses MAP for children beginning in kindergarten, and therefore these sections focus on kindergarten through second grade students, and only the subset of those students with data.

### Mathematics

We included a little over three-quarters of kindergarten children (77.5 percent;  $n = 55$ ), first-grade children (81.3 percent;  $n = 65$ ), and second-grade children (80.0 percent;  $n = 52$ ) in analyses. Results for the MAP mathematics assessment varied by grade level (addressing part of research question 5A: How did Gardendale children in kindergarten through second grade compare to the normative sample on MAP mathematics and reading?).

- **Kindergarten** children were on par with the normative sample in the fall and spring. Their initial gap decreased by 1.0 points (from scoring approximately 1.8 scale score points below the normative sample to approximately 0.8 scale score points above).
- **First-grade** children were significantly below the normative sample in the fall and on par with the normative sample in the spring. Their initial gap decreased by 2.7 points (from scoring approximately 4.8 scale score points below the normative sample to approximately 2.0 scale score points below).

- **Second-grade** children were significantly below the normative sample in the fall and spring. Their initial gap decreased by 4.9 points (from scoring approximately 10.5 scale score points below the normative sample to approximately 5.6 scale score points below).

(For more detailed information, see Appendix F, Tables F.1, F1.2, and F.3.)

Results for the first MAP test discussed in research question 5B (Did Gardendale children in kindergarten through second grade demonstrate significant improvement on MAP mathematics and reading?) indicated there was significant growth in mathematics knowledge for all three grade levels between the fall and the spring. Kindergarten children’s mathematics knowledge grew by 19.2 points; first-grade children’s mathematics knowledge grew by 18.4 points; and second-grade children’s mathematics knowledge grew by 19.5 points. (For more detailed information, see Appendix F, Table F.4.)

## Reading

We included about half of kindergarten children (54.9 percent;  $n = 39$ ), almost half of first-grade children (48.8 percent;  $n = 39$ ), and a little over half of second-grade children (56.9 percent;  $n = 37$ ) in analyses. Results for the MAP reading assessment varied by grade level (addressing part of research question 5A: How did Gardendale children in kindergarten through second grade compare to the normative sample on MAP mathematics and reading?).

- **Kindergarten** children were on par with the normative sample in the fall and spring. Their initial gap over the normed sample increased by 1.9 points (from scoring approximately 1.5 scale score points above the normative sample to approximately 3.5 scale score points above).
- **First-grade** children scored significantly below the normative sample in the fall and spring. Their initial gap below the normed sample decreased by 4.4 points (from scoring approximately 9.0 points below the normative sample to scoring 4.6 points below the normative sample).
- **Second-grade** children scored significantly below the normative sample in the fall and on par with the normative sample in the spring. Their initial gap below the normed sample decreased by 4.8 points (from scoring approximately 8.4 points below the normative sample to scoring 3.6 points below the normative sample).

For more detailed information, see Appendix F, Tables F-5, F-6, and F-7.

Results for the second MAP test mentioned in research question 5B (Did Gardendale children in kindergarten through second grade demonstrate significant improvement on MAP mathematics and reading?) indicated there was significant growth in reading knowledge from fall to spring for all three grade levels. Kindergarten children’s reading knowledge grew by 15.8 points; first-grade children’s reading knowledge grew by 16.9 points; and second-grade children’s reading knowledge grew by 16.4 points. (For more detailed information, see Appendix F, Table F-8.)

## Science

We included about three-quarters of second-grade children (75.4 percent;  $n = 49$ ) in the analysis. Results for research question 5C (How did second-grade Gardendale children compare to the normative sample on MAP science?) demonstrate that children were significantly below the normative sample in the fall and spring. Their initial gap below the normed sample decreased by 4.0 points (from scoring approximately 8.3 points below the normative sample to scoring 4.3 points below the normative sample). (For more detailed information, see Appendix F, Table F-9.)

Results for research question 5D (Did second-grade Gardendale children demonstrate significant improvement on MAP science?) indicated significant growth in science knowledge from fall to spring. Specifically, second-grade children’s science knowledge grew by 13.1 points. (For more detailed information, see Appendix F, Table F-10.)

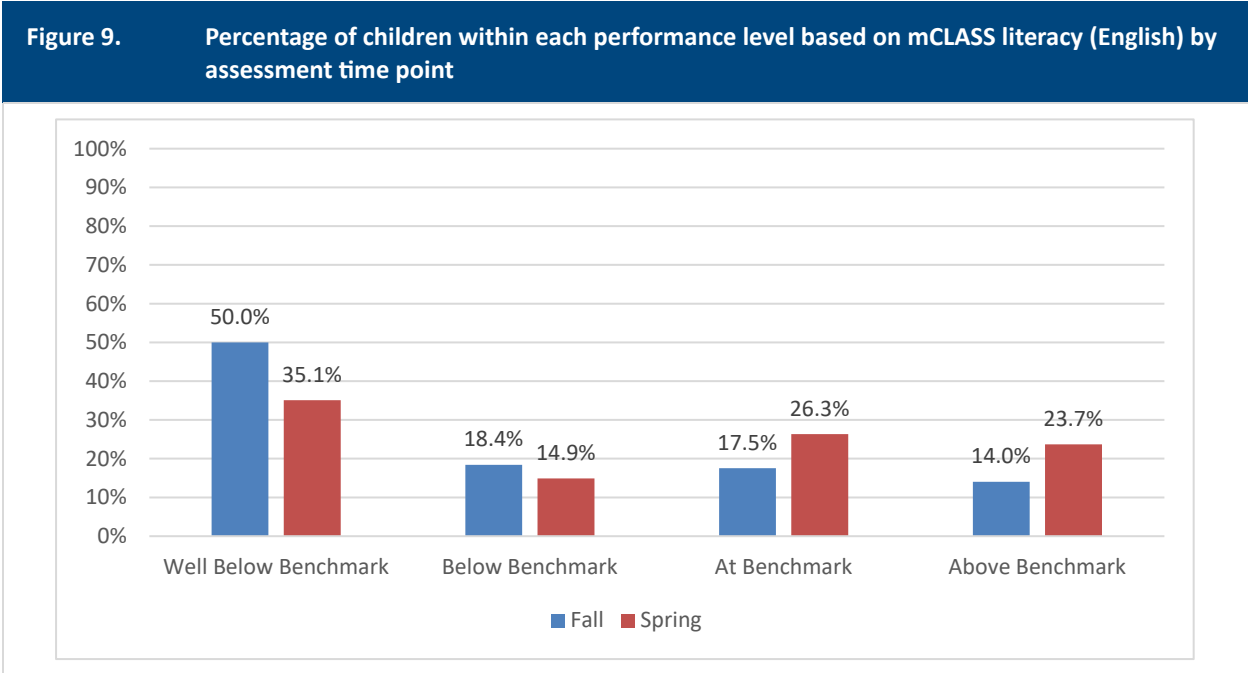
**mCLASS**

To measure children’s early literacy, Westat analyzed data collected by Gardendale for children in kindergarten through second grade (78.7 percent of total sample; *n* = 170) on the mCLASS, administered in the fall, winter, and spring of the school year. To evaluate children’s understanding, Westat analyzed the four performance levels: (1) well below benchmark, (2) below benchmark, (3) at benchmark, and (4) above benchmark. These levels are based on a comparison to the normative sample and represent the developmental trajectory of children based on their grade level and language of assessment (English and Spanish).

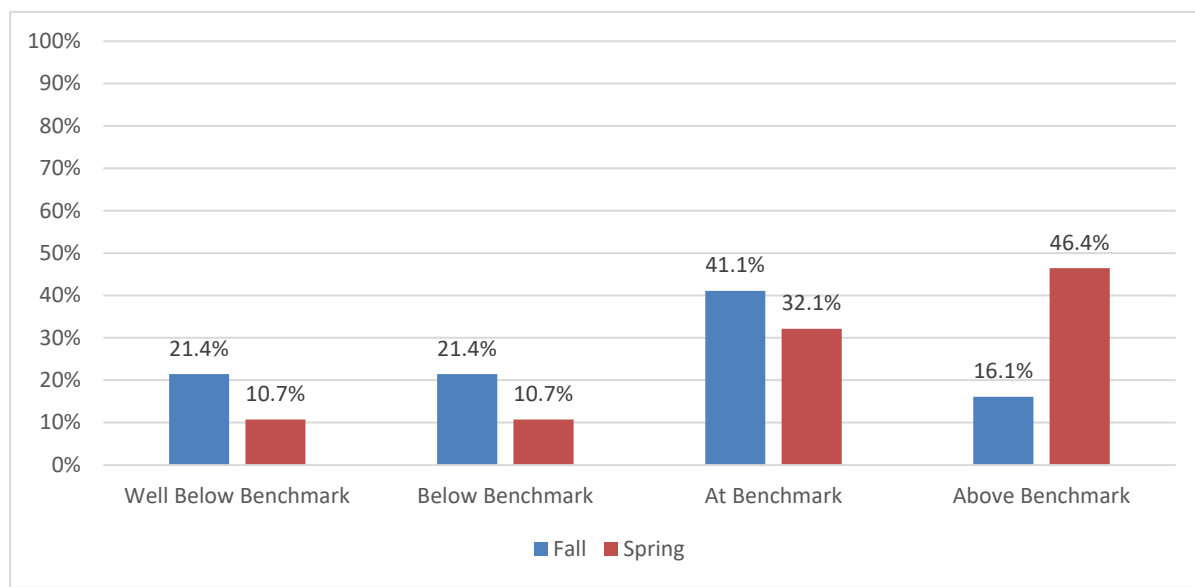
Results for research question 6A (What were the performance levels of Gardendale children in kindergarten through second grade in mCLASS literacy?) demonstrated significant positive movement for children assessed in English and Spanish. Findings by language of testing revealed different trends:

- For children assessed in English, a majority (50.0 percent) tested at the lowest level (well below benchmark) in the fall. However, in the spring, less than half of the assessed children (35.1 percent) were testing at the lowest level. Similarly, more children tested in the highest two levels (at benchmark and above benchmark) in the spring compared to the fall.
- For children assessed in Spanish, 16.1 percent of the assessed children tested at the highest level (above benchmark) in the fall, and 46.4 percent of the assessed children tested at the highest level in the spring. There was evidence of positive movement from fall to spring as more children tested in the highest two levels (at benchmark and above benchmark) in the spring.

See Figures 9 and 10 for illustration, and Appendix F, Tables F-11 and F-12, for more detail.



**Figure 10. Percentage of children within each performance level based on mCLASS literacy (Spanish) by assessment time point**



Results for research question 6B (Did Gardendale children in kindergarten through second grade demonstrate significant improvement in mCLASS literacy?) indicated there was significant improvement from fall to spring across all three grade levels. Children assessed in English improved by 123.6 points, and children assessed in Spanish improved by 122.5 points. For more detailed information and results by grade level, see Appendix F, Table F.13.

### Social-Emotional Assessment

To measure children’s social-emotional competencies, Westat analyzed data collected from teacher ratings of children using the Devereux Early Childhood Assessment (DECA) and Devereux Student Strengths Assessment (DESSA) in the fall and spring.<sup>11</sup> The DECA outcomes are Initiative, Self-Control, Attachment, Total Protective Factors, and Behavioral Concerns. Initiative, Self-Control, and Attachment are protective factors taken together to form an overall level of social-emotional competencies (or the Total Protective Factors). The full DESSA outcomes are Optimistic Thinking, Relationship Skills, Responsible Decision- Making, Self-Awareness, Self-Management, and Social Awareness. When taken together, these outcomes form the Overall Total.

### Devereux Early Childhood Assessment (DECA)

We included about three-quarters of pre-K children (77.6 percent;  $n = 52$ ) in DECA analyses;<sup>12</sup> we included children in our analyses if they had outcome data for both assessment times. Results for

<sup>11</sup> Pre-K 4 SA selected two different assessments because they are tailored to be developmentally appropriate based on grade level: Teachers administered the DECA to pre-K children, and the DESSA to children in kindergarten through second grade.

<sup>12</sup> We conducted demographic tests of differences to determine if the sample of children included in and excluded from analyses were similar. (See Appendix A, Analytic Approach, for more detailed information.)

research question 7A (What were the levels of Gardendale children’s social-emotional competency, and to what extent did the levels change?) revealed significant positive movement between levels for all outcomes except Behavioral Concerns. The results showed an increasing percentage of children scoring at the highest level (Strengths) between the fall and spring for

- Initiative (a 36.5 percent difference)
- Self-Control (a 26.9 percent difference)
- Attachment (a 46.2 percent difference)
- The combined score for Total Protective Factors (a maximum of a 61.5 percent difference)

The results also showed a declining percentage of children scoring at the lowest level (Needs Instruction) between the fall and spring for

- Initiative (a –13.5 percent difference)
- Self-Control (a –1.9 percent difference)
- Attachment (a –13.5 percent difference)
- Total Protective Factors (a –9.6 percent difference)

Behavioral Concerns showed no change in children demonstrating either need or typical.<sup>13</sup> (For more detailed information, see Appendix G, Table G.1.)

Results for research question 7B (Did Gardendale children demonstrate significant growth in social-emotional competence?) revealed that there was significant growth across all outcomes except Behavioral Concerns. The mean between fall and spring grew 10.1 points for Initiative, 7.8 for Self-Control, 12.8 points for Attachment, and 11.8 points for Total Protective Factors. For Behavioral Concerns, there was no significant change between the fall and spring, but the mean decreased by 0.7 points, which indicates an overall reduction in problematic behavior. (For more detailed information, see Appendix G, Table G.2.)

## Devereux Student Strengths Assessment (DESSA)

Teachers administered two versions of the DESSA. Pre-K 4 SA instructed teachers to administer a mini-DESSA to all children in the fall and spring, and a full DESSA to any child with a mini-DESSA score in the lowest category (Needs Instruction) in the fall and spring. The mini-DESSA is a shortened form of the full DESSA and measures social-emotional competency using an overall score. The full DESSA provides deeper understanding and measures six aspects of social-emotional competency.<sup>14</sup>

---

<sup>13</sup> The Behavioral Concerns outcome has only two levels by design. It is intended to measure problematic behaviors and is reversed from the other outcomes measuring positive behaviors. Therefore, the category Area of Need comes from high scores and all other scores fall in the Typical range.

<sup>14</sup> The six aspects are the scales, which include: Optimistic Thinking, Relationship Skills, Responsible Decision-Making, Self-Awareness, Self-Management, and Social Awareness.

## Mini DESSA

We included children in kindergarten through second grade (71.3 percent of total sample;  $n = 154$ ) in the mini-DESSA analyses<sup>15</sup> if they had outcome data for both assessment time points. Results for research question 7A (What were the levels of Gardendale children’s social-emotional competency, and to what extent did the levels change?) revealed most children tested at the Typical level, and there was no significant movement between levels from the fall to the spring. (For more detailed information, see Appendix G, Table G.3.)

Results for research question 7B (Did Gardendale children demonstrate significant improvement in social-emotional competence?) showed children had no significant growth from fall to spring. (For more detailed information, see Appendix G, Table G.4.)

## Full DESSA

We included children in kindergarten through second grade (7.9 percent<sup>16</sup> of total sample;  $n = 17$ ) in the full DESSA analyses if they had outcome data for both time points. Results for research question 7A (What were the levels of Gardendale children’s social-emotional competency, and to what extent did the levels change?) revealed no significant positive movement between levels across all outcomes. (For more detailed information, see Appendix G, Table G.5.)

Results for research question 7B (Did Gardendale children demonstrate significant improvement in social-emotional competence?) revealed that there was no significant growth across all outcomes. (For more detailed information, see Appendix G, Table G.6.)

# Conclusions and Looking Ahead

## Overview of Findings

The evaluation results of the Gardendale partnership in 2024–25 reflect the unique post-pandemic environment observed. The evaluation team highlights these five findings:

1. **Quality learning.** Gardendale continues to educate children through safe and supportive classroom environments that are organized and well-managed, and that provide opportunities for higher-order thinking. Considering the importance of emotionally supportive environments when safety, security, and well-being are crucial for young children, this finding is important. Moreover, average results for all three CLASS domains (Emotional Support, Classroom Organization, and Instructional Support) were higher this year compared to last year.<sup>17</sup>
2. **Significant improvement in academic and social-emotional skills.** Children demonstrated significant growth across multiple outcomes and measures over the 2024–25 school year. These outcomes and measures include kindergarten and first-grade readiness (GOLD); early literacy and numeracy (WJ and Bateria); receptive vocabulary (ROWPVT); expressive vocabulary (EOWPVT); mathematics, reading, and science (MAP); literacy (mCLASS); and social-emotional

<sup>15</sup> We conducted demographic tests of differences to determine if the sample of children included in and excluded from analyses were similar. (See Appendix A, Analytic Approach, for more detailed information.)

<sup>16</sup> It is not surprising that the sample includes a low percentage of children. Teachers only administered the full DESSA to children with a mini-DESSA score in the lowest category (Needs Instruction).

<sup>17</sup> The Emotional Support average went from 6.11 to 6.39; the Classroom Organization went from 5.74 to 6.23; and the Instructional Support average went from 3.71 to 4.05.

competency (DECA). The results suggest children benefit from their educational experience at Gardendale and provide empirical evidence that the partnership is supporting children’s achievement and learning. Given concerns among the broader education community about necessary learning support in response to the pandemic, these results provide one empirically evaluated example of a partnership that supports and achieves children’s learning.

3. **Different findings across assessments.** Results across multiple measures and outcomes, such as early literacy and numeracy (WJ and Batería), literacy and oral language (GOLD), mathematics (GOLD), and reading and mathematics (MAP), point to different achievement patterns based on grade level. We will expand on this finding in the next section.
4. **Comparing findings across partnership years.** There were comparable results across all three years for mathematics (MAP), literacy (mCLASS), and social-emotional (DECA), suggesting steady performance. However, other social-emotional (DESSA) analyses suggested mixed findings.
  - For MAP mathematics, results were similar across all three years for second-grade children. Results for kindergarten and first-grade children demonstrated varied findings by year. Across all three years, children consistently demonstrated significant growth.
  - For mCLASS literacy, previous results demonstrated comparable findings across all three years, suggesting that children exhibited the same performance. Overall, children consistently demonstrated significant positive movement into higher levels and demonstrated significant growth.
  - For DECA, previous results demonstrated comparable results across all three years. Children consistently demonstrated significant positive movement into higher levels and demonstrated significant growth for Total Protective Factors. Children performed similarly over time for Behavioral Concerns.
  - For mini-DESSA, the results demonstrated varied findings by year, demonstrating no consistent patterns over time. For this year and the 2022–23 school year, children did not demonstrate significant growth, but they did in the 2023–24 school year.
5. **Unfinished learning and learning loss.** Based on the prior evaluation reports and national trends, it is highly likely many Gardendale children were experiencing unfinished learning and learning loss from prior school years because of the pandemic (Center for Education Policy Research, 2023; Dewey et al., 2025; Diaz & Decker-Woodrow, 2021; Jung & Barnett, 2021; Socol, 2022; Weiland et al., 2021). Therefore, Gardendale children likely entered the 2024–25 school year with preexisting learning gaps and achievement challenges.

Taken together, these findings support the partnership between Pre-K 4 SA and Gardendale by not only providing findings for this school year, but also expounding on the existing results that demonstrate longitudinal patterns of learning and achievement. They also add to the larger conversation on the need for learning recovery and accelerating learning for all children.

## Comparing Assessment Results Across Measures

---

For most measures and outcomes, Gardendale children demonstrated significant improvement. However, sometimes different conclusions arose when comparing early literacy, early numeracy, and social-emotional competency findings across multiple measures to the normative samples. We provide a summary of reasons that may lead to different conclusions across measures to set context for readers. Then we give the cross-measure results for each of three categories of findings: literacy, numeracy, and social-emotional.

## Context and Caveats for Comparing Across Measures

Inconsistent findings across assessment tools are not specific to this Gardendale evaluation. Indeed, previous research comparing GOLD and direct child assessments found, in summary, prior research found children’s skills to be much more similar within a classroom when assessed using GOLD as compared to the direct assessments (Miller-Bains et al., 2017; Qiu et al., 2021; Russo et al., 2019).

We identified three pertinent reasons why these assessments could provide different conclusions.

- 1. Different scoring methods for each assessment.** GOLD results are based on comparing children to a single normative average; WJ, ROWPVT, and EOWPVT have age-specific (measured in months) normative averages; and DECA measures across three levels based on score transformations using a normative sample. Therefore, it may be possible for children to show more nuanced understanding on WJ, ROWPVT, and EOWPVT than on GOLD, leading to different conclusions.
- 2. Different assessors.** The GOLD and DECA are teacher-reported assessments, and the WJ IV, ROWPVT, and EOWPVT direct assessments collected by an independent assessor. Therefore, differences could be attributable to the data collector or collection method (e.g., teacher or independent assessor bias and teacher or independent assessor training) and not the content intended to be captured by the assessment.
- 3. Unclear or absent cross-measure reliability.** Notably, there is limited validity research that has compared these measures to determine how much content is similar or different across them (e.g., Barghaus et al., 2022; Miller-Bains et al., 2017; Qiu et al., 2021; Russo et al., 2019). Given this, there could be different levels or content measured (e.g., literacy, vocabulary, letter recognition, and phonemic awareness). Therefore, these measures may be measuring different and distinct aspects of early literacy, early numeracy, receptive vocabulary, and social-emotional competency.

## Results From Comparing Across Measures

**Early literacy.** All literacy findings indicated children demonstrated significant improvement as shown in Table 2. When comparing assessment findings for the remaining analyses, the results are mixed. For the GOLD assessment, pre-K children performed significantly above the normative sample on Literacy and on par with the normative sample in Oral Language. However, kindergarten children performed significantly below the normative sample on both Literacy and Oral Language. For Letter-Word, the majority of children performed below the normative samples.<sup>18</sup> For MAP reading, children performed significantly below or on par with the normative sample. When comparing findings according to the mCLASS, the majority of children performed below benchmark levels in fall and half performed below benchmark levels in spring when assessed in English. For children assessed in Spanish, the majority performed above benchmark levels. For receptive vocabulary (ROWPVT) and expressive vocabulary (EOWPVT), the majority of children performed in the expected range of understanding.

---

<sup>18</sup> This finding only holds for the majority of children taken from a random sample; it does not apply to the small subgroup of high-achieving children, which we analyzed separately.

**Table 2. Summary of early literacy findings across assessments**

| Outcome          | Assessment          | Grade  | Domain                       | Growth         | Percentage change | Cap closure | Benchmark and performance levels <sup>a</sup> |        | Norm comparisons <sup>b</sup> |        |   |
|------------------|---------------------|--------|------------------------------|----------------|-------------------|-------------|---|--------|-------------------------------|--------|---|
|                  |                     |        |                              | Fall to Spring |                   |             | Fall  | Spring | Fall                          | Spring |   |
| Early Literacy   | GOLD                | PK     | Literacy                     | ↑              |                   |             |   |        | ↑                             | ↑      |   |
|                  |                     |        | Oral Language                | ↑              |                   |             |   |        | —                             | —      |   |
|                  |                     | K      | Literacy                     | ↑              |                   |             |   |        | ↓                             | ↓      |   |
|                  |                     |        | Oral Language                | ↑              |                   |             |   |        | ↓                             | ↓      |   |
|                  | WJ <sup>c</sup>     | PK-2nd | Letter-Word                  | ↑              | —                 | ↑           |   |        | ↓                             | ↓      |   |
|                  |                     | PK-2nd | Letter-Word (high achieving) | ↑              | —                 | ↑           |   |        | ↑                             | ↑      |   |
|                  | ROWPVT <sup>c</sup> | PK-2nd | Receptive vocabulary         | ↑              | —                 |             | ↑   | ↑      |                               |        |   |
|                  | EOWPVT <sup>c</sup> | PK-2nd | Expressive vocabulary        | ↑              | —                 |             | ↑   | ↑      |                               |        |   |
|                  | MAP                 | K      | Reading                      |                | ↑                 |             |   |        |                               | —      | — |
|                  |                     |        |                              | 1st            | ↑                 |             |   |        |                               | ↓      | ↓ |
|                  |                     |        |                              | 2nd            | ↑                 |             |   |        |                               | ↓      | — |
|                  | mCLASS              | K-2nd  | English literacy             | ↑              | ↑                 |             | ↓   | —      |                               |        |   |
| Spanish literacy |                     |        | ↑                            | ↑              |                   | ↑           | ↑   |        |                               |        |   |

**Note:** A green arrow or triangle that points up indicates a positive significant result; a dash or a yellow bar indicates a nonsignificant result; a red arrow or triangle that points down indicates a negative significant result. We conducted analyses based on the assessment scoring methods indicated in the technical manuals. Columns and rows without icons indicate we did not conduct those analyses.

GOLD = Growth, Observation, and Learning; WJ = Woodcock-Johnson and Bateria; ROWPVT = Receptive One-Word Picture Vocabulary Test; EOWPVT = Expressive One-Word Picture Vocabulary Test; MAP = Measures of Academic Progress; PK = Pre-K; K = Kindergarten.

<sup>a</sup> ROWPVT and EOWPVT findings for performance levels are based on descriptive statistics; we did not conduct inferential tests.

<sup>b</sup> WJ findings for norm comparisons are based on descriptive statistics; we did not conduct inferential tests. Across all assessments except MAP, researchers created the norm comparisons prior to the COVID-19 pandemic and do not reflect pandemic-related disruptions to learning and well-being. Therefore, they represent a normative sample taken from environments which are most likely quite different from the environments experienced by Gardendale children. MAP revised its norm comparisons based on data from the 2022–23 school year.

<sup>c</sup> We conducted assessments for a random sample of children.

**Early numeracy.** All numeracy findings indicated children demonstrated significant improvement as shown in Table 3. When comparing assessment findings for the remaining analyses, the results are mixed. Second- grade assessment results for MAP demonstrated comparable results as the Applied Problems findings.<sup>19</sup> However, in the GOLD Mathematics domain, pre-K and kindergarten children were either significantly above or on par with the normative sample. This disagreed with Applied Problems findings, which demonstrated children were significantly below the normative sample. According to MAP findings, children demonstrated different patterns from fall to spring compared to the normative sample and based on grade level.

<sup>19</sup> This finding only holds for the majority of children taken from a random sample; it does not apply to the small subgroup of high-achieving children, which we analyzed separately.

**Table 3. Summary of early numeracy findings across assessments**

| Outcome        | Assessment      | Grade  | Domain                            | Growth | Percentage change | Gap closure | Benchmark and performance levels <sup>b</sup> |        | Norm comparisons <sup>b</sup> |        |
|----------------|-----------------|--------|-----------------------------------|--------|-------------------|-------------|---|--------|-------------------------------|--------|
|                |                 |        |                                   |        | Fall to Spring    |             | Fall  | Spring | Fall                          | Spring |
| Early Numeracy | GOLD            | PK     | Mathematics                       | ↑      |                   |             |   |        | —                             | ↑      |
|                |                 | K      |                                   | ↑      |                   |             |   |        | —                             | —      |
|                | WJ <sup>c</sup> | PK-2nd | Applied Problems                  | ↑      | —                 | —           |   |        | ↓                             | ↓      |
|                |                 |        | Applied Problems (high achieving) | ↑      | —                 | ↑           |   |        | ↓                             | ↑      |
|                | MAP             | K      | Mathematics                       | ↑      |                   |             |   |        | —                             | —      |
|                |                 | 1st    |                                   | ↑      |                   |             |   |        | ↓                             | —      |
|                |                 | 2nd    |                                   | ↑      |                   |             |   |        | ↓                             | ↓      |

**Note:** A green arrow or triangle that points up indicates a positive significant result; a dash or yellow bar indicates a nonsignificant result; a red arrow or triangle that points down indicates a negative significant result. Westat conducted analyses based on the assessment scoring methods indicated in the technical manuals. Columns and rows without icons indicate we did not conduct those analyses.

GOLD = Growth, Observation, and Learning; WJ = Woodcock-Johnson and Batería; MAP = Measures of Academic Progress; PK = Pre-K; K = Kindergarten.

<sup>a</sup> WJ findings for norm comparisons are based on descriptive statistics; we did not conduct inferential tests. Across all assessments except MAP, researchers created the norm comparisons prior to the COVID-19 pandemic and do not reflect pandemic-related disruptions to learning and well-being. Therefore, they represent a normative sample taken from environments which are most likely quite different from the environments experienced by Gardendale children. MAP revised its norm comparisons based on data from the 2022–23 school year.

<sup>b</sup> We conducted assessments for a random sample of children.

**Social-emotional competency.** Pre-K social-emotional findings indicated children demonstrated significant improvement as shown in Table 4. Pre-K and kindergarten children were either significantly below or on par with the normative sample in the GOLD Social-Emotional domain. For DECA and the mini-DESSA<sup>20</sup>, findings demonstrate that the majority of children were scoring at the Typical or Strengths levels.

<sup>20</sup> The full DESSA findings are based on a small subgroup of children, so it would not be appropriate to include them with the overall findings.

**Table 4. Summary of social-emotional findings across assessments**

| Outcome          | Assessment   | Grade | Domain                           | Growth | Percentage change | Gap closure | Benchmark and performance levels <sup>c</sup> |        | Norm comparisons <sup>b</sup> |        |
|------------------|--------------|-------|----------------------------------|--------|-------------------|-------------|---|--------|-------------------------------|--------|
|                  |              |       |                                  |        | Fall to Spring    |             | Fall  | Spring | Fall                          | Spring |
| Social-Emotional | GOLD         | PK    | Social-Emotional                 | ↑      |                   |             |   |        | ↓                             | —      |
|                  |              | K     |                                  | ↑      |                   |             |   |        | —                             | ↓      |
|                  | DECA         | PK    | Initiative                       | ↑      | ↑                 |             | ↑   | ↑      |                               |        |
|                  |              |       | Self-Control                     | ↑      | ↑                 |             | ↑   | ↑      |                               |        |
|                  |              |       | Attachment                       | ↑      | ↑                 |             | ↑   | ↑      |                               |        |
|                  |              |       | Total Protective Factors         | ↑      | ↑                 |             | ↑   | ↑      |                               |        |
|                  |              |       | Behavioral Concerns <sup>c</sup> | —      | —                 |             | ↑   | ↑      |                               |        |
|                  | DESSA (mini) | K-2nd | Overall                          | —      | —                 |             | ↑   | ↑      |                               |        |
|                  | DECA (full)  | K-2nd | Self-Awareness                   | —      | —                 |             | ↓   | ↓      |                               |        |
|                  |              |       | Self-Management                  | —      | —                 |             | ↓   | ↓      |                               |        |
|                  |              |       | Social Awareness                 | —      | —                 |             | ↑   | ↑      |                               |        |
|                  |              |       | Relationship Skills              | —      | —                 |             | ↑   | ↑      |                               |        |
|                  |              |       | Decision-Making                  | —      | —                 |             | ↓   | ↓      |                               |        |
|                  |              |       | Optimistic Thinking              | —      | —                 |             | ↑   | ↓      |                               |        |
|                  |              |       | Overall                          | —      | —                 |             | ↓   | ↓      |                               |        |

**Note:** A green arrow or triangle that points up indicates a positive significant result; a dash or a yellow bar indicates a nonsignificant result; a red arrow or triangle that points down indicates a negative significant result. Westat conducted analyses based on the assessment scoring methods indicated in the technical manuals. Columns and rows without icons indicate we did not conduct those analyses.

GOLD = Growth, Observation, and Learning; DECA = Devereux Early Childhood Assessment; DESSA = Devereux Student Strengths Assessment; PK = Pre-K; K = Kindergarten.

<sup>a</sup> These findings are based on descriptive statistics; we did not conduct any inferential tests.

<sup>b</sup> Across all assessments, researchers created the norm comparisons prior to the COVID-19 pandemic and do not reflect pandemic-related disruptions to learning and well-being. Therefore, they represent a normative sample taken from environments which are most likely quite different from the environments experienced by Gardendale children.

<sup>c</sup> The Behavioral Concerns outcome is intended to measure problematic behaviors, which is reversed from the other outcomes measuring positive behaviors. A positive significant finding in Behavioral Concerns indicates a significant reduction in challenging behaviors.

## Limitations

The evaluation findings have four key limitations.

- 1. Resource limitations constrained data collection.** Westat was not able to collect data from a comparison school with which to compare the Gardendale children because of resource constraints. Therefore, we used normative samples for comparisons. There can be more confidence in interpreting resulting differences on outcomes when a comparison or control group is formed with children who are similar to the Gardendale children and experienced

learning during the pandemic. Furthermore, there can be more confidence that differences can be attributed to Gardendale and are not a result of other factors.<sup>21</sup>

- 2. Pandemic context influenced these children’s trajectories.** Researchers create normative samples to be reflective of the demographic proportions similar to those found in the U.S. Census data. For all assessments except MAP,<sup>22</sup> researchers collected normative data prior to the COVID-19 pandemic and therefore do not consider the pandemic-related disruption to learning and well-being that has occurred for children and families. The children in this normative data sample are most likely quite different from the Gardendale children and did not experience learning in the same context. Hence, using normative comparisons (except for MAP) is not ideal, but these normative samples are the best research evidence currently available for comparison. Therefore, readers should interpret these results with caution. Moving forward, it will be necessary to obtain new normative sample results from test publishers in order to perform more comparable analyses. Given the amount of effort it takes to create normative samples, at a minimum it will be several years before it is possible to conduct such analyses.
- 3. High-achieving subgroup findings do not represent all children.** The early literacy and numeracy results from Letter-Word for the high-achieving children are not representative of all Gardendale children. Rather, they represent a small subgroup of children who demonstrated significant above-grade-level knowledge and experienced significant accelerated growth. Therefore, readers should interpret these findings with caution.
- 4. GOLD, MAP, mCLASS, DECA, and DESSA findings do not represent all Gardendale children.** Findings from GOLD, MAP, mCLASS, DECA, and DESSA do not represent the full population of enrolled children.<sup>23</sup> We included children with complete data across all assessment time points in analyses and excluded children who enrolled after the initial assessment window or withdrew prior to subsequent data collection. Accordingly, results reflect only those children with complete participation across assessment periods.

## Directions for Future Research

Taken together, these findings demonstrated children benefited from attending Gardendale. We would like to draw attention to four directions for future research.

- 1. Improve instructional quality.** Given the midrange CLASS score for Instructional Support, children would benefit from increased opportunities to think critically, be asked questions to prompt their thought processes and scaffold learning, and hear and use language. Therefore, Pre-K 4 SA should consider continuing to focus on and invest resources in improving instructional quality (e.g., teacher professional learning, mentoring, and coaching).
- 2. Encourage accelerated learning.** The significant growth in early literacy and numeracy, and receptive and expressive vocabulary provide empirical evidence of a step in the right direction. However, many results demonstrated that children were performing below what would be expected for their grade based on national norms. Pre-K 4 SA should consider continuing to focus on and invest targeted resources (e.g., evidence-based enrichment activities or engaging

---

<sup>21</sup> One way to form a comparison group of children similar in nature to Gardendale children in the future would be to work with Teaching Strategies to create a matched comparison group from the normative sample of children.

<sup>22</sup> MAP revised its norm comparisons based on data from the 2022–23 school year.

<sup>23</sup> We conducted demographic tests of differences to determine if the sample of children included in and excluded from analyses were similar. (See Appendix A, Analytic Approach, for more detailed information.)

families and caregivers to support their child’s education) for supporting and bringing children’s academic understanding up to grade level.

3. **Examine the high- achieving subgroup.** In contrast, there is a small subgroup of children who experienced accelerated learning and significantly reduced existing learning gaps in early literacy and numeracy. Notably, the previous evaluation demonstrated the early literacy finding last year, which suggests a potential positive trend. Pre-K 4 SA should observe these classrooms and talk with these teachers to fully understand the mechanisms behind such strong improvement. This could also shape professional development for teachers to improve program effectiveness and replicate this success in other classrooms.
4. **Investigate Pre-K Literacy and Kindergarten GOLD findings.** Across all three assessment times in the 2024-25 school year, pre-K Gardendale children were significantly above the normative samples in Literacy. For kindergarten Gardendale children started the school year on par with the normative sample and finished the year significantly below the normative sample. Pre-K 4 SA should collaborate with Gardendale educators to consider investigating what mechanisms might be behind Gardendale children’s performance for these two outcomes over time.

## References

- Barghaus, K., Fantuzzo, J., Buek, K., & Gullo, D. (2022). Neglected validities: A diagnostic look at the state of early childhood assessment. *Early Childhood Research Quarterly, 58*(2), 287–99. <https://doi.org/10.1016/j.ecresq.2021.09.007>
- Barnett, W. S. (2011). Four reasons the United States should offer every child a preschool education. In E. Zigler, W. Gilliam, & W.S. Barnett (Eds.), *The pre-K debates: Current controversies and issues* (pp. 34–39). Brookes Publishing.
- Bassok, D., & Galdo, E. (2016). Inequality in preschool quality? Community-level disparities in access to high-quality learning environments. *Early Education and Development, 27*(1), 128–44. <https://doi.org/10.1080/10409289.2015.1057463>
- Bassok, D., Magouirk, P., & Markowitz, A. J. (2021). Systemwide quality improvement in early childhood education: Evidence from Louisiana. *AERA Open, 7*, 1–19. <https://doi.org/10.1177/23328584211011610>
- Biancarosa, G., Kennedy, P. C., Park, S., & Otterstedt, J. (2021). *mCLASS DIBELS 8th Edition: Administration and scoring guide*. Center on Teaching & Learning.
- Bloom, H. S., & Weiland, C. (2014). *To what extent do the effects of Head Start on enrolled children vary across sites?* (working paper). MDRC.
- Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25*(2), 166–76. <https://doi.org/10.1016/j.ecresq.2009.10.004>
- Campbell, F.A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6*(1), 42–57. [https://doi.org/10.1207/S1532480XADS0601\\_05](https://doi.org/10.1207/S1532480XADS0601_05)
- Center for Education Policy Research. (2023). *New research finds that pandemic learning loss impacted whole communities, regardless of student race or income*. Harvard University. Retrieved from <https://cepr.harvard.edu/news/new-research-finds-pandemic-learning-loss-impacted-whole-communities-regardless-student>.
- Center on Reinventing Public Education (CPRE). (2024). *The state of the American student: Fall 2024*. [https://crpe.org/wp-content/uploads/CRPE\\_SOS2024\\_FINAL.pdf](https://crpe.org/wp-content/uploads/CRPE_SOS2024_FINAL.pdf)
- Coogle, C. G., Storie, S., & Rahn, N. L. (2021). A framework for promoting access, increasing participation, and providing support in early childhood classrooms. *Early Childhood Education Journal, 50*(5), 867–77. <https://doi.org/10.1007/s10643-021-01200-6>
- Decker-Woodrow, L., Diaz, E., Adachi, E., Barfield, D., & Lamey, G. (2018). *Pre-K 4 SA evaluation report: Year 4*. Westat.
- Decker-Woodrow, L., Diaz, E., Barfield, D., & Lamey, G. (2017). *Pre-K 4 SA evaluation report: Year 5*. Westat.
- Decker-Woodrow, L., Diaz, E., Lamey, G., Hartman, N., Adachi, E., & Barfield, D. (2019). *Pre-K 4 SA evaluation report: Year 6*. Westat.
- Decker-Woodrow, L., & Price, E. (2016). *Pre-K 4 SA evaluation report: Year 3*. Edvance Research.

- Dewey, D. C., Fahle, E., Kane, T. J., Reardon, S. F., & Staiger, D. O. (2025). *Education Recovery Scorecard: Pivoting from pandemic recovery to long-term reform: A district-level analysis*. Center for Education Policy Research,, Harvard University. <https://educationrecoveryscorecard.org/wp-content/uploads/2025/02/Pivoting-from-Pandemic-Recovery-to-Long-Term-Reform-A-District-Level-Analysis.pdf>
- Diaz, E., & Decker-Woodrow, L. (2021). *Pre-K 4 SA and Edgewood ISD 1882 partnership evaluation report*. Westat.
- Diaz, E., Decker-Woodrow, L., & Adachi, E. (2023). *Pre-K 4 SA and Edgewood Independent School District Year 3 Gardendale evaluation report*. Westat.
- Diaz, E., Decker-Woodrow, L., & Adachi, E. (2022). *An exploration of Pre-K 4 SA education centers post-pandemic*. Westat.
- Diaz, E., Sergueev, V., Glidden, S., Harper, G., Liu, A., & Decker-Woodrow, L., (2024). *Pre-K 4 SA and Gardendale Early Learning Program: Year 5 technical report*. Westat.
- Dunn, L., & Dunn, L. (2019). *Peabody Picture Vocabulary Test, fifth edition (PPVT -V)*. NCS Pearson.
- Dunn, L. M., Lugo, D. E., Padilla, E. R., & Dunn, L. M. (1986). *Test de Vocabulario en Imagenes Peabody - TVIP*. American Guidance Services.
- Edvance Research. (2015). *Pre-K 4 SA evaluation report: Year 2*. Edvance Research.
- Edvance Research. (2014). *Pre-K 4 SA evaluation report: Year 1*. Edvance Research.
- Fahle, E. M., Kane, T. J., Patterson, T., Reardon, S. F., Staiger, D. O., & Stuart, E. A. (2023). *School district and community factors associated with learning loss during the COVID-19 pandemic*. Center for Education Policy Research, Harvard University. [https://cepr.harvard.edu/sites/hwpi.harvard.edu/files/cepr/files/explaining\\_covid\\_losses\\_5.23.pdf](https://cepr.harvard.edu/sites/hwpi.harvard.edu/files/cepr/files/explaining_covid_losses_5.23.pdf).
- Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the High/Scope Perry Preschool Program. *Journal of Public Economics*, 94(1–2), 114–28. <https://doi.org/10.1016/j.jpubeco.2009.11.001>
- Hill, C. J., Gormley, W. T., Jr., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Childhood Research Quarterly*, 32, 60–79. <https://doi.org/10.1016/j.ecresq.2014.12.005>
- Ingvanson, E. M., Perry, L. K., VanDam, M., & Grieco-Calub, T. M. (2023). Comparing scores on the Peabody Picture Vocabulary Test and Receptive One-Word Picture Vocabulary Test in preschoolers with and without hearing loss. *American Journal of Speech-Language Pathology*, 32(3), 1610–1619. [https://doi.org/10.1044/2023\\_AJSLP-22-00352](https://doi.org/10.1044/2023_AJSLP-22-00352)
- Jung, K., & Barnett, W. S. (2021). *Impacts of the pandemic on young children and their parents: Initial findings from NIEER's May-June 2021 preschool learning activities survey*. National Institute for Early Education Research. <https://nieer.org/research-report/impacts-of-the-pandemic-on-young-children-and-their-parents-initial-findings-from-nieers-may-june-2021-preschool-learning-activities-survey>.
- La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system (CLASS): Findings from the prekindergarten year. *Elementary School Journal*, 104(5), 409–26. <https://doi.org/10.1086/499760>

- 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. (2012). *Devereux Early Childhood Assessment for Preschoolers, Second Edition (DECA-P2): User's guide and technical manual*. Kaplan Press.
- Lekhal, R., Karlsen, L., Wilhelmsen, T., Vandell, D. L., Lydersen, S., & Rydland, V. (2025). Promoting classroom quality through professional development: Results from the Oslo early education study, a randomized controlled trial. *Studies in Educational Evaluation*, 86, Article 101462. <https://doi.org/10.1016/j.stueduc.2025.101462>
- Lippard, C. N., La Paro, K. M., Rouse, H. L., & Crosby, D. A. (2018). A closer look at teacher–child relationships and classroom emotional context in preschool. *Child & Youth Care Forum*, 47, 1–21. <https://doi.org/10.1007/s10566-017-9414-1>
- Locasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly*, 22(1), 3–17. <https://doi.org/10.1016/j.ecresq.2006.05.001>
- Maier, M. F., McCormick, M. P., Xia, S., Hsueh, J., Weiland, C., Morales, A., Boni, M., Tonachel, M., Sachs, J., & Snow, C. (2022). Content-rich instruction and cognitive demand in prek: Using systematic observations to predict child gains. *Early Childhood Research Quarterly*, 60(3), 96–109. <https://doi.org/10.1016/j.ecresq.2021.12.010>
- 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.
- Martinson, B., Supe, I., Stokenberga, I., Damberga, I., Cefai, C., Camilleri, L., Bartolo, P., O'Riordan, M. R., & Grazzani, I. (2022). Social emotional competence, learning outcomes, emotional and behavioral difficulties of preschool children: Parent and teacher evaluations. *Frontiers in Psychology*, 12, Article 760782. <https://doi.org/10.3389/fpsyg.2021.760782>
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D. M., & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79(3), 732–49. <https://doi.org/10.1111/j.1467-8624.2008.01154.x>
- McCormick, M. (2022). *Building better evidence on pre-K by strengthening assessments of children's skills*. MDRC. <https://www.mdrc.org/publication/building-better-evidence-pre-k-strengthening-assessments-children-s-skills>
- Miller-Bains, K. L., Russo, J. M., Williford, A. P., DeCoster, J., & Cottone, E. A. (2017). Examining the validity of a multidimensional performance-based assessment at kindergarten entry. *AERA Open*, 3(2). <https://doi.org/10.1177/2332858417706969>

- Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *The Batería III Woodcock-Muñoz: Pruebas de aprovechamiento*. Riverside.
- 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/>
- Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the classroom assessment scoring system) in early childhood education and care settings and child outcomes. *PloS one*, *11*(12), Article e0167660.
- Puma, M., Bell, S., Cook, R., & Heid, C. (2010). *Head Start Impact Study. Final report*. U.S. Department of Health and Human Services, Administration for Children and Families.
- Purtell, K. M., & Ansari, A. (2018). Classroom age composition and preschoolers' school readiness: The implications of classroom quality and teacher qualifications. *AERA Open*, *4*(1), 1–13.
- Qiu, Y., Leite, W. L., Rodgers, M. K., & Hagler, N. (2021). Construct validation of an innovative observational child assessment system: Teaching Strategies GOLD® birth through third grade edition. *Early Childhood Research Quarterly*, *56*, 41–51. <https://doi.org/10.1016/j.ecresq.2021.02.005>.
- Reynolds, A. J., Temple, J. A., White, B. A., Ou, S. R., & Robertson, D. L. (2011). Age-26 cost-benefit analysis of the Child-Parent Center early education program. *Child Development*, *82*(1), 379–404.
- 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.
- Rolnick, A., & Grunewald, R. (2003). Early childhood development: Economic development with a high public return. *The Region*, *17*(4), 6–12.
- Russo, J. M., Williford, A. P., Markowitz, A. J., Vitiello, V. E., & Bassok, D. (2019). Examining the validity of a widely-used school readiness assessment: Implications for teachers and early childhood programs. *Early Childhood Research Quarterly*, *48*, 14–25. <https://doi.org/10.1016/j.ecresq.2019.02.003>.
- Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Achievement*. Riverside.
- Socol, A. R. (2022, September 1). *NAEP results show dismal learning loss due to pandemic. What can be done?* The Education Trust. Retrieved from <https://edtrust.org/the-equity-line/naep-results-show-dismal-learning-loss-due-to-pandemic-what-can-be-done/>.
- Teachstone (2023). *Classroom Assessment Scoring System 2nd Edition: Pre-K–3rd Observation Field Guide*. Teachstone.
- Valentino, R. (2018). Will public pre-K really close achievement gaps? Gaps in prekindergarten quality between students and across states. *American Educational Research Journal*, *55*(1), 79–116.
- Vance, B., West, R., & Kutsick, K. (1989). Prediction of Wechsler Preschool and Primary Scale of Intelligence IQ scores for preschool children using the Peabody Picture Vocabulary Test–R and the Expressive One Word Picture Vocabulary Test. *Journal of Clinical Psychology*, *45*(4), 642–44.

- Villareal, M. (2019). *Impact study of prekindergarten for San Antonio*. Urban Education Institute at the University of Texas at San Antonio.
- Weiland, C. (2016). Impacts of the Boston prekindergarten program on the school readiness of young children with special needs. *Developmental Psychology*, 52(11), 1763–76.
- Weiland, C., Greenberg, E., Bassok, D., Markowitz, A., Rosada, P. G., Luetmer, G., Abenavoli, R., Gomez, C., Johnson, A., Jones-Harden, B., Maier, M., McCormick, M., Morris, P., Nores, M., Phillips, D., & Snow, C. (2021). *Historical crisis, historic opportunity: Using evidence to mitigate the effects of the COVID-19 crisis on young children and early care and education programs*. Education Policy Initiative, University of Michigan, and Urban Institute.  
<https://edpolicy.umich.edu/sites/epi/files/uploads/EPI-UI-Covid%20Synthesis%20Brief%20June%202021.pdf>.