Westat

Improving Lives Through Research®

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

January 31, 2025

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

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

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

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

Table of Contents

Table of Contents (continued)

Tables

Figures

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 prekindergarten (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. The evaluation investigated the demographic characteristics of the children served, classroom quality, and children's learning during the 2023–24 school year in Gardendale. During its fifth year, the partnership served more than 260 children. Of those children served, 58.6 percent were boys and 41.4 percent were girls, and the majority (95.5 percent) were Hispanic.

Westat analyzed data collected by Pre-K 4 SA, Gardendale, and a team of observers who conducted classroom quality observations, teacher report assessments, and direct child achievement assessments. Results indicated

- Partnership classrooms provided high levels of Emotional Support, midrange levels of 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 for a small subgroup of children. We observed accelerated learning, which reduced the gap for early numeracy but did not result in a significant reduction; and
- Gardendale children showed significant improvement in indicators of vocabulary, mathematics, science, and social-emotional competency over time.

Taken together, the results from the Year 5 evaluation suggest children served by the partnership experienced positive classroom environments and are benefiting from participation in Gardendale.

Introduction

Providing access to high-quality early childhood education has received, and will continue to receive, considerable attention throughout the United States (Barnett, 2011; Campbell et al., 2002; 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. Previous research indicates children from racially marginalized communities, children from economically disadvantaged backgrounds, and children whose primary language is not English are more often exposed to lower quality instruction and learning environments across the United States (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; Perlman et al., 2016).

Because of limited public funding from federal and state governments, municipal governments are increasingly using funding sources in creative ways to provide more equitable access to high-quality early childhood education and care. In Texas, some districts will increase their funding by engaging with selected partners to creatively 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 supports to children attending Gardendale. Pre-K 4 SA has used their learning model, which has shown 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., 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 2023–24 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 in 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). Moreover, these findings coincide with previous external evaluation reports of Gardendale which indicated that most children were not performing at their age level in early literacy and early numeracy (Diaz & Decker-Woodrow, 2021; Diaz et al., 2023). 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. This report highlights the 2023–24 school year.

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 5th year (2023–24 school year) of the partnership and complements the previous reports provided. The purpose of the current report is to present evaluation findings, including (1) demographics of children served; (2) classroom quality; and (3) assessment of children's understanding of and improvement in early literacy and early numeracy,

receptive vocabulary, physical development, and social-emotional learning. Following our presentation of the results in this report, we summarize, synthesize, and compare all the findings across assessments. The last section of this report outlines limitations and provides recommendations. This report is the second in a series that documents the results of the Pre-K 4 SA initiative during the 2023–24 school year.

Research Questions

The Year 5 (2023–24) evaluation of Gardendale addressed the following main research questions and subquestions:

1. What was the observed teacher–child interaction quality of Gardendale classrooms in Year 5?
2. A. How did pre-K and kindergarten Gardendale children compare to the normative sample on the Growth, Observation, and Learning (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 a random sample of Gardendale children performed at or above their age level in early literacy and early numeracy, and to what extent did the percentage change?
    B. Did a random sample of Gardendale children demonstrate significant improvement in early literacy and early numeracy?
    C. Did a random sample of Gardendale children experience accelerated learning to help narrow achievement gaps in early literacy and early numeracy?
4. A. What were the receptive vocabulary performance levels of a random sample of Gardendale children?
    B. Did a random sample of Gardendale children demonstrate significant improvement in receptive vocabulary?
    C. What types of receptive vocabulary improvement did a random sample of Gardendale children demonstrate?
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 2023–24 school year, as well as a brief discussion of the methods used.

Sample

Data were provided for 268 children attending Gardendale. Table 1 includes the demographics for the sample. Children were in pre-K (11.9 percent of total sample) through second grade (23.1 percent of total sample), with most children in kindergarten (34.3 percent of total sample). There were more boys (58.6 percent) than girls (41.4 percent). Most children were Hispanic (95.5 percent), did not receive special-education services (86.6 percent), and were not receiving English Learner services (67.5 percent).

Table 1. Demographics of the overall sample

Demographic characteristic		*N* (percentage)
Gender	Male	157 (58.6%)
	Female	111 (41.4%)
Grade level	Pre-K	32 (11.9%)
	Kindergarten	92 (34.3%)
	First	82 (30.6%)
	Second	62 (23.1%)
Race/ethnicity	Hispanic/Latino	256 (95.5%)
	Not Hispanic/Latino	12 (4.5%)
Economically disadvantaged	Yes	227 (84.7%)
	No	41 (15.3%)
Receiving special-education services	Yes	36 (13.4%)
	No	232 (86.6%)
Receiving English Learner services	Yes	87 (32.5%)
	No	181 (67.5%)

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

To answer the first research question pertaining to teacher–child interaction quality, a total of 13 classrooms with observation data from the spring were included. One of the classrooms was pre-K (7.7 percent of the total classrooms assessed), five were kindergarten (38.5 percent of the total classrooms assessed),[1] four were first grade (30.8 percent of the total classrooms assessed), and three were second grade (23.1 percent of the total classrooms assessed).

[1] One of these five classrooms was a combined prekindergarten and kindergarten-aged classroom. It was categorized as kindergarten since most children were in kindergarten.

Methods

All research questions were addressed through analysis of existing Pre-K 4 SA and Gardendale databases and classroom observations. To answer the first question, data were collected and 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. Scores were assigned during various classroom activities over the course of a morning and then averaged across all observation periods (or cycles) for overall quality scores in three domains. Interactions were measured through 10 different dimensions that are sorted into the 3 larger domains. The Emotional Support domain is measured using four dimensions (positive climate, negative climate, educator sensitivity, and regard for child perspectives); the Classroom Organization domain is measured using three dimensions (productivity, behavior management, and instructional learning formats); and the Instructional Support domain is measured using three dimensions (concept development, quality of feedback, and language modeling).

CLASS uses a 7-point Likert-type scale, 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. Each dimension and domain are assigned a score during each of five 20-minute cycles. The number of children and adults in the classroom was also recorded during each of the five 20-minute cycles. (See Appendix A for more detailed information.)

To address the second set of research questions, descriptive and inferential analyses were conducted 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 (Lambert, 2020; see Appendix A for more detailed information).

To address the third set of questions, data collected by Pre-K 4 SA were submitted to Westat and analyzed descriptively and inferentially. Two direct assessments, early literacy (Letter-Word) and early numeracy (Applied Problems), were administered to a random sample of Gardendale children in the fall and spring. These two assessments are subtests from the Woodcock-Johnson IV Tests of Achievement (WJ; Schrank et al., 2014) and matching subtests from the Batería III Spanish assessment (Muñoz-Sandoval et al., 2005; see Appendix A for more detailed information). They were chosen because they are widely used in early childhood and 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, data collected by Pre-K 4 SA were submitted to Westat and analyzed descriptively and inferentially. A direct assessment of vocabulary was administered to a random sample of children in fall and spring using the Peabody Picture Vocabulary Test-5 (PPVT; Dunn & Dunn, 2019; see Appendix A for more detailed information). Like WJ IV, this assessment was chosen because it is widely used in early childhood and complements the GOLD findings by providing additional insights from a trained assessor as compared to a teacher report. (Puma et al., 2010). While the GOLD findings provide an overall perspective and measure multiple aspects of

early literacy and numeracy, the PPVT findings are more nuanced and measure receptive vocabulary knowledge and understanding.

To address the fifth set of questions, data collected by Gardendale were submitted to Westat and analyzed descriptively and inferentially. An electronic assessment of mathematics, reading, and science, the Measures of Academic Progress (MAP; NWEA, 2023), was administered to children in kindergarten through second grade three times throughout the school year. Participating children entered their responses on an iPad. (See Appendix A for more detailed information.)

To address the sixth set of questions, data collected by Gardendale were submitted to Westat and analyzed descriptively and inferentially. An assessment of early literacy, the mCLASS, was administered 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 phonemic awareness, phonics, fluency, vocabulary, and comprehension. Children in kindergarten were given an online assessment; children in first and second grade were administered a reading booklet, and their teacher entered their responses into the data system (Biancarosa et al., 2021; see Appendix A for more detailed information).

To address the seventh set of questions, data collected by Gardendale were submitted to Westat and analyzed descriptively and inferentially. A teacher report assessment of social-emotional competence, the Devereux Early Childhood Assessment (DECA; LeBuffe & Naglieri, 2012) was administered to children in pre-K in the fall and spring, and the Devereux Student Strengths Assessment (DESSA; LeBuffe et al., 2014) was administered to children in kindergarten through second grade in the fall and spring. (See Appendix A for more detailed information).

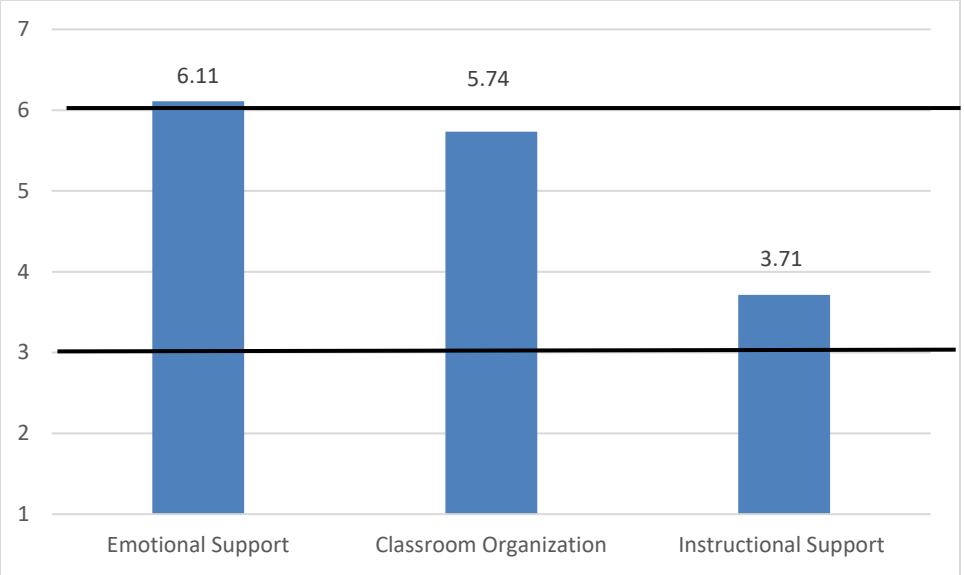
Evaluation Results

Teacher–Child Interaction Quality

Results for the first research question (What was the overall observed teacher–child interaction quality in Gardendale classrooms across Year 5?) are presented in Figure 1.[2] Across the four grade levels, Gardendale classrooms (n = 13) were observed using CLASS, second edition, (Teachstone, 2023) during Year 5. Scores for the Emotional Support domain ranged from 5.2 to 6.9 on a 1–7 scale, with an average score of 6.1; 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 4.6 to 6.8, with an average score of 5.7, which suggests classrooms sometimes showed effective interactions regarding Classroom Organization. Finally, Instructional Support domain scores ranged from 2.7 to 5.4, with an average score in the midrange (3.7), which suggests 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 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; Locasale-Crouch et al., 2007; Maier et al., 2022; Mashburn et al., 2008; Purtell & Ansari, 2018). Each of the Year 5 CLASS domain scores is represented visually in Figure 1.

[2] Average ratings across all CLASS dimensions are provided in Appendix B.

Figure 1. Year 5 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), mid- (between 3 and 6), and high (6 and above).

Kindergarten and First-Grade Readiness: Growth, Observation, and Learning (GOLD) Results

Pre-K children (87.5 percent of children; *n* = 28) and kindergarten children (83.7 percent of children; *n* = 77) were included in analyses[3] if they had GOLD data for all three assessment times[4] in at least one of the following six outcomes: cognitive, literacy, mathematics, oral language, physical, and social-emotional.

As data were not collected on a comparison or control group, comparisons were conducted using the nationally representative normed data based on age bands for the GOLD assessment (Lambert, 2020). The results for research question 2A (How did pre-K and kindergarten Gardendale children compare to the normative sample on GOLD outcomes?) are presented separately for pre-K and kindergarten children as the norms vary depending on grade level.

Pre-K (Kindergarten Readiness) Results

In the fall, pre-K children were significantly below the normative sample on two of the six GOLD outcomes (cognitive and literacy) and on par[5] with the normative sample on the remaining four of

[3] As children were not randomly sampled, demographic tests of differences were conducted to determine if the sample of children included in and excluded from analyses were similar. See Appendix A, Analytic Approach for more detailed information.

[4] There was a small subset of pre-K and kindergarten children (*n* < 10) not able to be included in GOLD analyses due to invalid data for at least one assessment time. These children were excluded from all GOLD outcome analyses.

[5] While Gardendale children's scores were different from the normative sample, none of these differences were statistically significant.

the six GOLD outcomes (mathematics, oral language, physical, and social-emotional; for more detailed information, see Appendix C, Table C-1).

Spring results for the cognitive outcome indicated the initial gap between Gardendale children and the normative sample was eliminated by the end of the school year. By spring, no significant difference was found between Gardendale children and the normative sample for this outcome. The gap in cognitive was reduced by 63.5 percent (a decrease from an initial gap of approximately 34 scale score points to a gap of approximately 12 scale score points). A significant gap remained for the literacy outcome, which increased from an initial gap of approximately 64 scale score points to a gap of approximately 71 scale score points: an increase of 7 scale score points.

Across all three assessment times, Gardendale children were similar to the normative sample in mathematics and oral language. For the mathematics outcome, the same pattern was observed last year (2022–23 school year). More information is needed to understand what mechanisms might be behind Gardendale children scoring similarly to the normative sample for these outcomes across all three assessment times.

Spring results for the remaining two outcomes (physical and social-emotional) indicated Gardendale children were significantly below the normative sample. The gap in physical increased from approximately 11 scale score points in the fall to approximately 52 scale score points in the spring, and the gap in social-emotional increased from approximately 8 scale score points in the fall to 18 scale score points in the spring. More information is needed to understand what mechanisms might be behind Gardendale children scoring similarly to the normative sample in the fall but then scoring significantly below the normative sample in the spring. (See Appendix C, Table C-1 for more information.)

Kindergarten (First-Grade Readiness) Results

In the fall, kindergarten children were below the normative sample on two GOLD outcomes (literacy and oral language), on par[6] with the normative sample on three GOLD outcomes (cognitive, physical, and social-emotional), and significantly above the normative sample on the remaining GOLD outcome (mathematics). The effect size (Hedges' g) for the significant result is large: 1.3.

Spring results indicated a significant gap remained for the literacy and oral language outcomes. For literacy, the initial gap increased approximately 17 scale score points (from an initial gap of approximately 76 scale score points to a gap of approximately 93 scale score points). For the physical outcome, results indicated a significant gap developed between the fall and spring assessments. Children finished the year significantly below the normative sample despite starting the year on par with the normative sample: The gap in physical increased from approximately 8 scale score points to approximately 17 scale score points. More information is needed to understand what mechanisms might be behind the Gardendale children scoring similarly to the normative sample in the fall for physical but then scoring significantly below the normative sample in the spring. (See Appendix C, Table C-2 for more information.)

Across all three assessment times, Gardendale children were similar to the normative sample in cognitive and social-emotional. For the remaining outcome (mathematics), children finished the year on par[7] with the normative sample despite starting the year significantly ahead of the

[6] While Gardendale children's scores were different from the normative sample, none of these differences were statistically significant.

[7] While Gardendale children's scores were different from the normative sample, none of these differences were statistically significant.

normative sample. More information is needed to understand what mechanisms might be behind the Gardendale children scoring similarly to the normative sample for these outcomes. (For more detailed information, see Appendix C, Table C-2.)

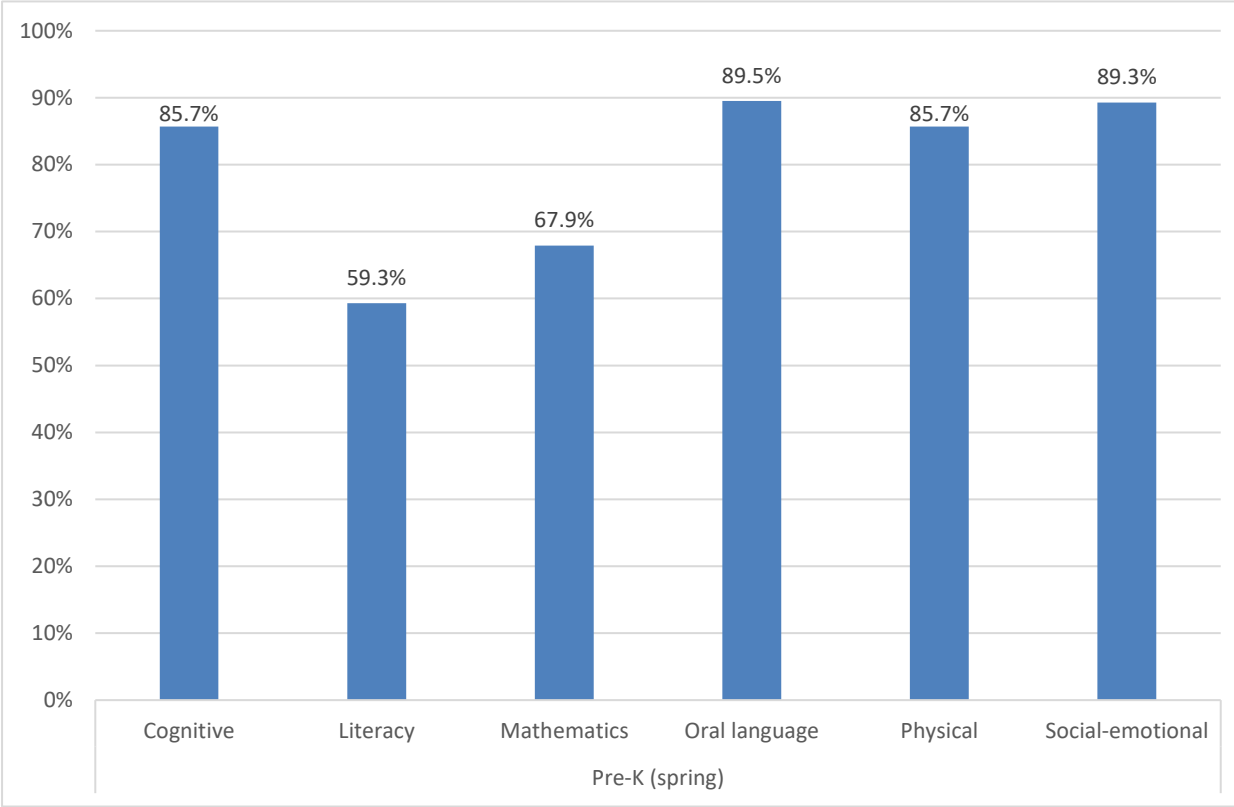
Pre-K and Kindergarten Growth Results

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 39.6 scale score points for the physical domain to 100.0 scale score points for the cognitive domain. For kindergarten children, the growth ranged from 58.8 scale score points for the literacy domain to 94.0 scale score points for the oral language domain. (See Appendix C, Table C-3 for more information.)

Kindergarten Readiness Results

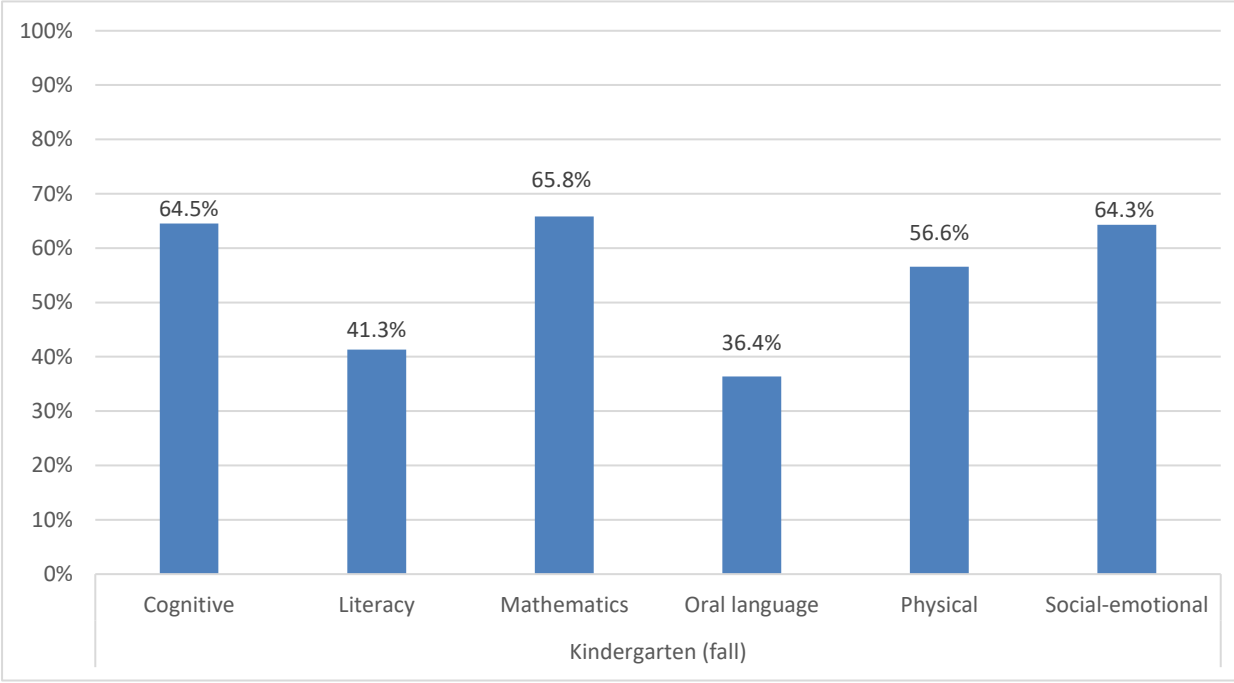
Results for research question 2C (What percentage of Gardendale children demonstrated kindergarten readiness as measured by GOLD outcomes?) were conducted separately for pre-K and kindergarten children to determine (1) whether pre-K children finished the year ready for kindergarten, and (2) whether kindergarten children started the year ready for kindergarten. 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 67.9 percent for the mathematics domain to 89.5 percent for the oral language 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 by GOLD outcome



For the kindergarten children, the majority demonstrated readiness at the start the year in four out of six outcomes (cognitive, mathematics, physical, and social-emotional). For the remaining two outcomes (oral language and literacy), the majority demonstrated they were not ready for kindergarten. The readiness ranged from 36.4 percent for the oral language outcome to 65.8 percent for the mathematics outcome, as shown in Figure 3. (See Appendix C, Table C-4 for more information.)

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

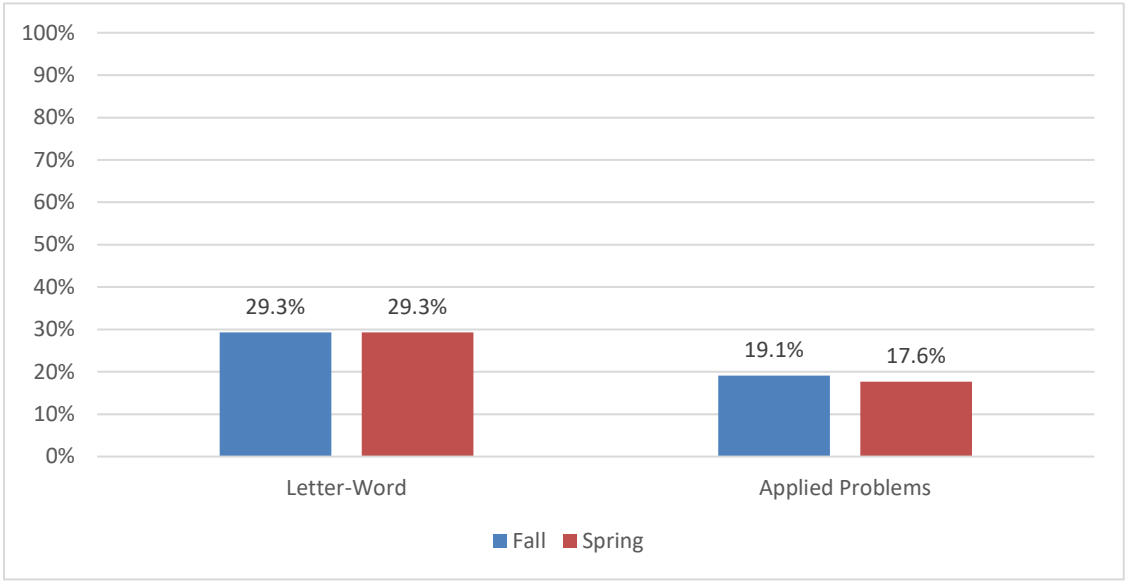


Direct Child Assessments

Woodcock-Johnson (WJ) and Batería

Westat analyzed data collected by Pre-K 4 SA from a random sample (n = 66) on two subtests of a direct child achievement assessment: Letter-Word and Applied Problems from the WJ and the Batería. Based on the results for Letter-Word, there were two patterns of early literacy achievement. The first pattern represented the majority of children (n = 58), which we will present in this section, and the second group represents a smaller subset (n < 10), which we will present in a later section. Results for the first part of research question 3A (What percentage of a random sample of Gardendale children performed at or above their age level in early literacy and early numeracy?) showed that 29.3 percent of children in Gardendale were at or above their age level in fall and spring, as shown in Figure 4. For early numeracy, in the fall 19.1 percent of children in Gardendale were at or above their age level, and in the spring 17.6 percent of children were at or above their age level. These findings imply most children are performing below their age level at both assessment times and are behind what would be considered ready for the next grade from a nationally representative lens. Moreover, these findings mirror the results we observed last school year (2022–23).

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



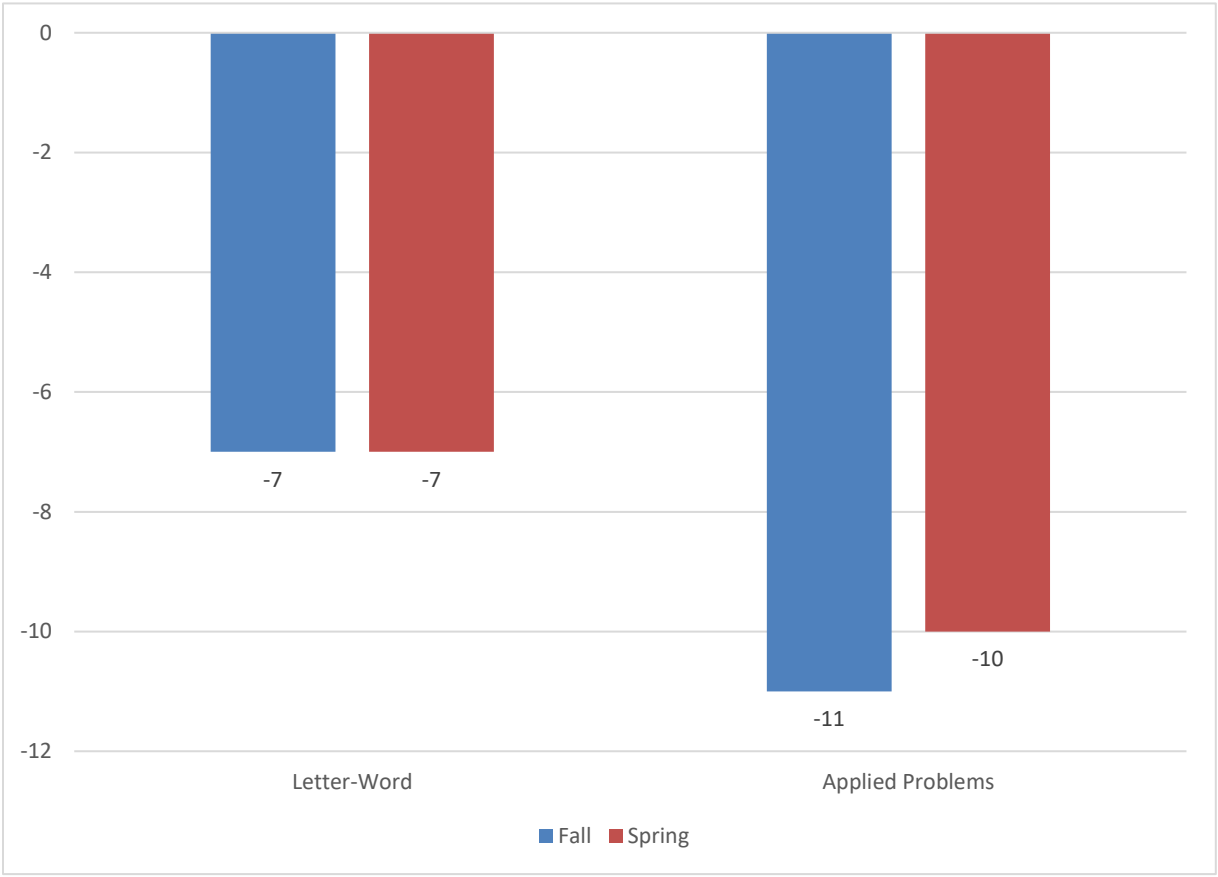
Note: Letter-Word measures early literacy skills, and Applied Problems measures early numeracy skills.

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 a random sample of 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 6 months of growth in learning in 6 months, and for early numeracy, approximately 7 months of growth in learning during 6 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 a random sample of Gardendale children experience accelerated learning to help narrow achievement gaps in early literacy and early numeracy?) indicated that accelerated learning did not occur for either outcome. For early literacy, children were on average 7 months below the norms in the fall, and in the spring, children remained 7 months below the norms (see Figure 5). For early numeracy, children were on average 11 months below the norms in the fall, and in the spring, children were 10 months below the norms. Therefore, the gap between Gardendale children and the national norm was decreased by 1 month only for one of the subtests. (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.

Letter-Word High-Achieving Subgroup

In this section, we will present the second group, which is a subgroup of children demonstrating high achievement. We adapted research question 3A (What percentage of a random sample of Gardendale children performed at or above their age level in early literacy?) to focus on a high-achieving subgroup ($n < 10$) of Gardendale children. Results showed the majority of the high-achieving subgroup tested at or above their age level at both assessment times throughout the school year.

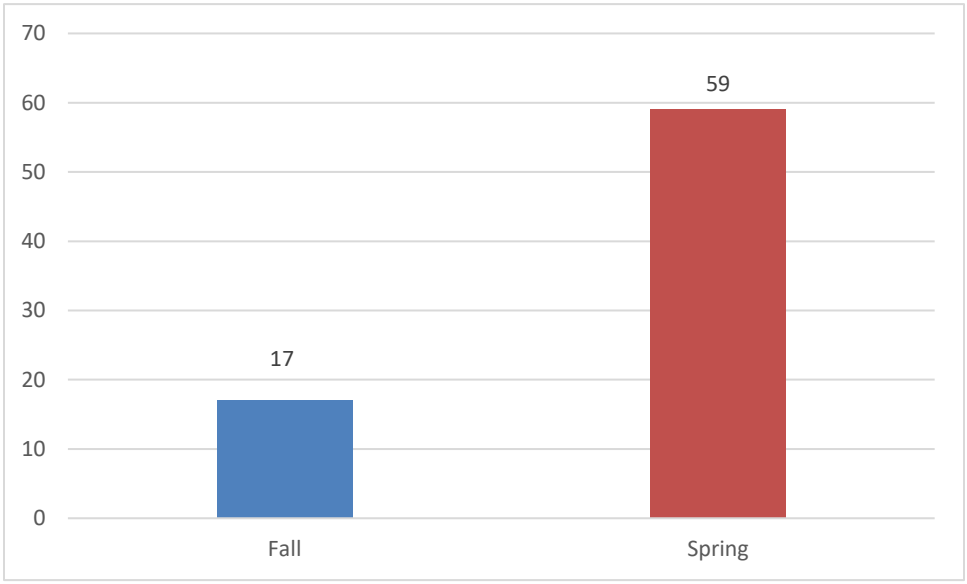
When applying the second part of research question 3A (To what extent did the percentage change?) to this high-achieving subgroup, results showed no significant difference in children performing at or above their age level in early literacy in the spring compared to the fall. (For more detailed information and results by grade level, see Appendix D, Table D-4.)

Reframing research question 3B (Did a random sample of Gardendale children demonstrate significant improvement in early literacy?) to focus on this high-achieving subgroup revealed there was significant improvement. For early literacy, there was approximately 4 years of growth in learning during 6 months of time. Therefore, this finding suggests children gained significant

knowledge of early literacy from fall to spring: approximately eight times the actual amount of time between assessments. (For more detailed information and results by grade level, see Appendix D, Table D-5.)

Similarly, reframing research question 3C (Did a random sample of Gardendale children experience accelerated learning to help narrow achievement gaps in early literacy?) to focus instead on this high-achieving subgroup indicated significant accelerated learning did occur in early literacy. In the fall, children were 17 total months ahead of the norms on average, and in spring, children were 59 total months ahead of the norms (see Figure 6). This implies Gardendale children were outperforming the national norms at both assessment points and further surpassed the norms in the spring. Therefore, the existing positive difference between Gardendale children and the national norm was surpassed even more by 3 years and 6 months. (For more detailed information and results by grade level, see Appendix D, Table D-6.)

Figure 6. Size of early literacy achievement difference in months between high-achieving Gardendale subgroup and normative sample by assessment time point



Peabody Picture Vocabulary Test (PPVT)

To measure children's receptive vocabulary, Westat analyzed data collected by Pre-K 4 SA from a random sample (n = 65) on the PPVT. For this assessment, children were presented pictorial images of words and asked to select the image that matched the word said by the examiner. To evaluate children's understanding, their scores were converted 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, analyses of vocabulary growth were conducted 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 learning across the year.

Results for research question 4A (What were the receptive vocabulary performance levels of a random sample of Gardendale children?) demonstrated the majority of children were performing in

the expected range in both fall and spring (for more detailed information and results by grade level, see Appendix E, Table E-1).

Results for research question 4B (Did a random sample of Gardendale children demonstrate significant improvement in receptive vocabulary?) indicated children overall experienced significant improvement and gained additional vocabulary (approximately four growth scale points) when comparing their fall and spring scores across the year. (For more detailed information and results by grade level, see Appendix E, Table E-2.) When comparing children's growth across the year to the normative sample, there was no significant difference between the two groups. This indicates Gardendale children were progressing at a rate that is typical of children of the same age.

Results for research question children 4C (What types of receptive vocabulary improvement did a random sample of Gardendale children demonstrate?) showed five distinct types of improvement. Most children learned new vocabulary. Moreover, for about one-third (30.8 percent of the total sample), the gap between their performance and the performance of typically developing children of the same age had narrowed, indicating that they experienced accelerated learning in their vocabulary. (For more detailed information and results by grade level, see Appendix E, Table E-3.)

Measures of Academic Progress (MAP)

To measure children's mathematics, reading, and science[8] knowledge, Westat analyzed MAP data collected by Gardendale. Data were collected for children in kindergarten through second grade in the fall, winter, and spring of the school year. Children were included in analyses[9] if they had outcome data for all three assessment times. As data were not collected on a comparison or control group, comparisons were conducted using the nationally representative normed data (Thum & Kuhfeld, 2020).[10]

Mathematics Results

About two-thirds of kindergarten children (66.3 percent; n = 61), two-thirds of first-grade children (69.5 percent; n = 57), and a little over three-quarters of second-grade children (82.3 percent; n = 51) were included in analyses. Results for the first MAP test discussed in research question 5A (How did Gardendale children in kindergarten through second grade compare to the normative sample on MAP mathematics and reading?) varied by grade level. Kindergarten children were on par[11] with the normative sample in the fall and spring. Their initial gap shrunk by 42.4 percent from fall to spring, scoring below the normative sample by 2.4 points in the fall and by 1.4 points in the spring. First-grade children were significantly below the normative sample in the fall and significantly surpassed the normative sample in the spring, with a medium effect size (Hedges' g = 0.5). Second-grade children were significantly below the normative sample in the fall and spring. Their initial gap decreased by 9.8 percent from fall to spring, scoring below the normative sample by 11.9 points in the fall and by 10.7 points in the spring. Findings for all three grade levels indicated children

---

8 The science assessment is administered for second-grade children.

9 As children were not randomly sampled, demographic tests of differences were conducted to determine if the sample of children included in and excluded from analyses were similar. (See Appendix A, Analytic Approach for more detailed information.)

10 Gardendale children assessed in English were compared to the normative sample based on grade level (Thum & Kuhfeld, 2020). Gardendale children assessed in Spanish were not compared to the normative sample because the Spanish normative sample is not nationally representative.

11 While Gardendale children's scores were different from the normative sample, this difference was not statistically significant.

experienced accelerated learning in mathematics. (For more detailed information, see Appendix F, Tables F-1, F-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 18.6 points; first-grade children's mathematics knowledge grew by 33.2 points; and second-grade children's mathematics knowledge grew by 15.6 points. (For more detailed information, see Appendix F, Table F-4.)

Reading Results

Almost half of kindergarten children (44.6 percent; *n* = 41), a little over half of first-grade children (57.3 percent; *n* = 47), and about two-thirds of second-grade children (64.5 percent; *n* = 40) were included in analyses. Results for the second MAP test discussed in research question 5A (How did Gardendale children in kindergarten through second grade compare to the normative sample on MAP mathematics and reading?) varied by grade level. Kindergarten children were on par with the normative sample in the fall and were significantly below the normative sample in the spring. Their initial gap grew by 4.7 points, from scoring below the normative sample by 1.1 points in the fall to scoring below by 5.8 points in the spring, suggesting they experienced less improvement in reading. First-grade children scored significantly below the normative sample in the fall (below by 9.3 points) and closed the gap in the spring by scoring similar to the normative sample (above by 3.5 points), demonstrating they experienced accelerated learning. Second-grade children scored significantly below the normative sample in the fall and spring. Second-grade children's initial gap grew by 1.5 points from fall to spring, scoring below the normative sample by 8.8 points in the fall to below by 10.3 points in the spring, suggesting they experienced less improvement in reading. (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 11.7 points; first-grade children's reading knowledge grew by 28.3 points; and second-grade children's reading knowledge grew by 11.7 points. (For more detailed information, see Appendix F, Table F-8.)

Science Results

A little over 40 percent of second-grade children (41.9 percent; *n* = 26) were included in the analyses. Results for research question 5C (How did second-grade Gardendale children compare to the normative sample on MAP science?) demonstrated children were significantly below the normative sample in the fall and spring. Second-grade children's initial gap shrunk by 1.3 percent from fall, when children scored below the normative sample by 7.7 points, to spring, when children scored below the normative sample by 7.6 points. (For more detailed information, see Appendix F, Table F-9.)

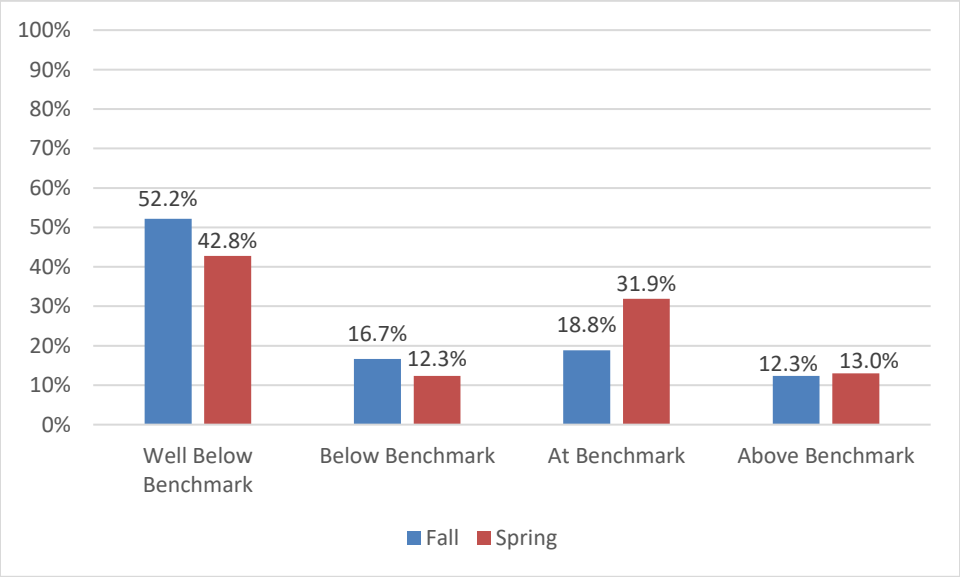
Results for research question 5D (Did second-grade Gardendale children demonstrate significant improvement on MAP science?) indicated significant growth in science knowledge from fall to spring: Second-grade children's science knowledge grew by 10.3 points. (For more detailed information, see Appendix F, Table F-10.)

mCLASS

To measure children's early literacy, Westat analyzed data collected by Gardendale for children in kindergarten through second grade (79.7 percent of total sample; *n* = 188) on the mCLASS, administered in the fall, winter, and spring of the school year. Children were included in the analyses[12] if they had outcome data for all three assessment times. To evaluate children's understanding, their scores were converted into 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).

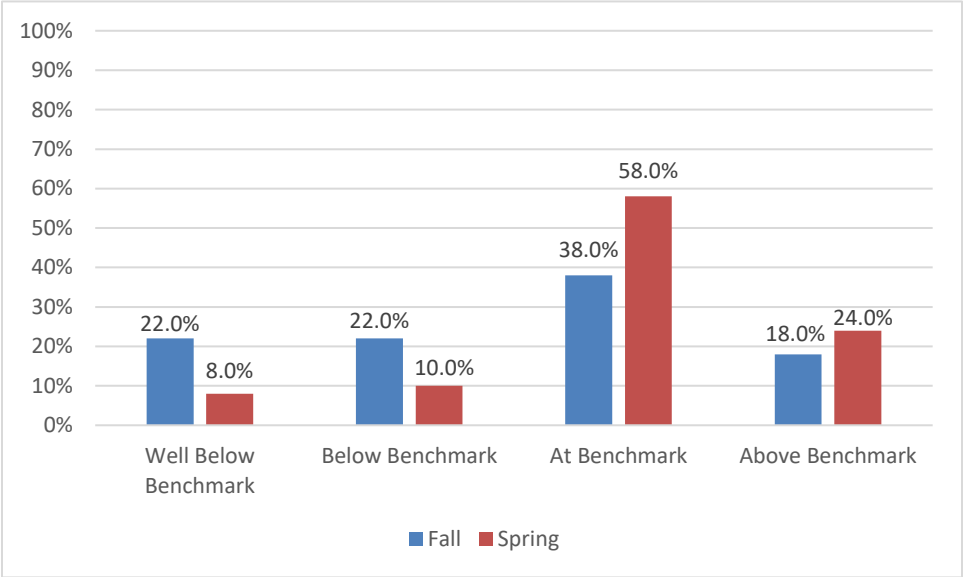
Results for research question 6A (What were the performance levels of Gardendale children in kindergarten through second grade in mCLASS literacy?) demonstrated significant positive movement for children assessed in English and Spanish. For children assessed in English, a majority (52.2 percent) tested at the lowest level (well below benchmark) in the fall. However, in the spring, less than half of the assessed children (42.8 percent) were testing at the lowest level. Similarly, more children tested in the highest two levels (at benchmark and above benchmark) from fall to spring (see Figure 7; for more detailed information and results by grade level, see Appendix F, Table F-11). For children assessed in Spanish, 38.0 percent of the assessed children tested at the third level (at benchmark) in the fall, and 58.0 percent of the assessed children tested at the third 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 Figure 8; for more detailed information and results by grade level, see Appendix F, Table F-12).

Figure 7. Percentage of children within each performance level based on mCLASS literacy (English) by assessment time point



[12] As children were not randomly sampled, demographic tests of differences were conducted to determine if the sample of children included in and excluded from analyses were similar. (See Appendix A, Analytic Approach for more detailed information.)

Figure 8. 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. Kindergarten children assessed in English improved by 122.6 points, and children assessed in Spanish improved by 133.8 points. First-grade children assessed in English improved by 121.5 points, and children assessed in Spanish improved by 101.9 points. Second-grade children assessed in English improved by 106.0 points, and children assessed in Spanish improved by 90.9 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 on the Devereux Early Childhood Assessment (DECA) and Devereux Student Strengths Assessment (DESSA) in the fall and spring.[13] 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. 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 Personal Responsibility, Optimistic Thinking, Goal-Directed Behavior, Social Awareness, Decision Making, Relationship Skills, Self-Awareness, and Self-Management. When taken together, these outcomes form the Overall Total.

[13] Two different assessments were selected because they are tailored to be developmentally appropriate based on grade level: the DECA is administered to pre-K children, and the DESSA is administered to children in kindergarten through second grade.

Devereux Early Childhood Assessment (DECA)

About 85 percent of pre-K children (84.4 percent; *n* = 27) were included in DECA analyses[14] if they had outcome data for both assessment times. 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 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 66.7 percent difference), Attachment (a maximum of a 77.7 percent difference), and combined score for Total Protective Factors (a 48.1 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 maximum of a −14.8 percent difference[15]), Self-control (a −14.8 percent difference), and Total Protective Factors (a maximum of a −7.4 percent difference[16]). Self-control showed an increasing percentage of children scoring at the second-highest level (Typical) between the fall and spring (a 33.3 percent difference), but a decreased percentage of children scoring at the highest level (Strengths) between the fall and spring (a −18.5 percent difference). Behavioral Concerns showed no change in children demonstrating need. Therefore, these findings indicate significant positive change between levels of understanding across all outcomes except Behavioral Concerns. (For more detailed information, see Appendix G, Table G-1.)

[14] As children were not randomly sampled, demographic tests of differences were conducted to determine if the sample of children included in and excluded from analyses were similar, (See Appendix A, Analytic Approach for more detailed information.)

[15] Results are rounded to protect confidentiality.

[16] Results are rounded to protect confidentiality.

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 Self-control and Behavioral Concerns. The mean between the fall and spring grew 12.3 points for Initiative, 8.8 points for Attachment, and 8.3 points for Total Protective Factors. For Self-control and Behavioral Concerns, there was no significant change between the fall and spring: for Self-control, the mean increased by 0.8 points, and for Behavioral Concerns, the mean decreased by 0.04 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 were instructed 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.[17] The mini-DESSA is a shortened form of the full DESSA and measures social-emotional competency in an overall score. The full DESSA provides a deeper understanding and measures eight aspects of social-emotional competency.[18] However, in the fall, data were not collected for second graders. Therefore, we used children's spring 2023 scores from the 2022–23 school year (when they were first graders) in place of fall 2023 scores. Because the data collection time points differ, we conducted the analyses separately by age group: children in kindergarten through first grade were combined into a group, and second-grade children were a separate group. Children in kindergarten through first grade (47.1 percent of total sample; n = 82) and children in second grade (40.3 percent of total sample; n = 25) were included in the mini-DESSA analyses[19] if they had outcome data for both assessment time points.

Kindergarten and First-Grade Results

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 a significant percentage of children in kindergarten and first grade scoring at the highest level (Strengths) between the fall and spring assessments (a 2.4 percent difference; 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 that children in kindergarten and first grade significantly increased their scores from fall to spring by a mean of 2.8 points. (For more detailed information, see Appendix G, Table G-4.)

Second-Grade Results

Results for research question 7A (What were the levels of Gardendale children's social-emotional competency, and to what extent did the levels change?) showed an increased percentage of children scoring at the highest level (Strengths) between spring 2023 and spring 2024 (less than a 40.0 percent difference[20]), but this was not a significant increase. Results for research question 7B (Did

[17] There were no children with data from a full DESSA in the fall and spring. Therefore, it was not possible to conduct analyses for the full DESSA.

[18] The eight aspects are the scales, which include: Self-Awareness, Social Awareness, Self-Management, Goal-Directed Behavior, Relationship Skills, Personal Responsibility, Decision Making, and Optimistic Thinking.

[19] As children were not randomly sampled, demographic tests of differences were conducted to determine if the sample of children included in and excluded from analyses were similar, (See Appendix A, Analytic Approach for more detailed information.)

[20] Results are rounded to protect confidentiality.

Gardendale children demonstrate significant improvement in social-emotional competence?) indicated second graders showed a mean 1.3 point increase in scores between spring 2023 and spring 2024, but this was not significant. (For more detailed information, see Appendix G, Tables G-5 and G-6.)

Conclusions and Looking Ahead

Overview of Findings

The evaluation results of the Gardendale partnership in 2023–24 reflect the unique post-pandemic environment observed and present four encouraging findings. First, the levels of teacher–child interaction quality suggest that the partnership continues to educate children through safe and supportive classroom environments that are organized, managed, and 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) increased compared to last year.[21] The Emotional Support domain average is now in the high range compared to the midrange, and the Instructional Support average is now in the midrange compared to the low range.

Second, children demonstrated significant growth across multiple outcomes and measures over the 2023–24 school year. These outcomes and measures include kindergarten and first-grade readiness (GOLD), early literacy and numeracy (WJ and Batería), receptive vocabulary (PPVT), mathematics, reading, and science (MAP), literacy (mCLASS), and social-emotional competency (DECA and DESSA). The results suggest children benefit from their educational experience at Gardendale, and they provide empirical evidence that the partnership is supporting children's achievement and learning. Given concerns among the broader education community about necessary learning supports in response to the pandemic, these results provide one empirically evaluated example of a partnership that supports and achieves children's learning.

Third, some of the results indicated children (regardless of grade level) began and finished the 2023–24 school year performing below their age level across multiple measures and outcomes, such as early literacy and numeracy (WJ and Batería) and literacy and oral language (GOLD), when compared to a national normative sample of children. However, other measures, such as mathematics (GOLD) and reading and mathematics (MAP), point to different achievement patterns based on grade level. This will be expounded on in the next section.

Fourth, this was the second year conducting receptive vocabulary (PPVT), mathematics (MAP), literacy (mCLASS), and social-emotional (DECA and DESSA) analyses. This allows for comparing trends over time for these outcomes and measures. Receptive vocabulary results for this year are similar to those for the 2022–23 school year. For MAP mathematics, results were similar across both years for second-grade children. Results for kindergarten and first-grade children show positive improvement this year compared to last year. This year, kindergarten children closed the gap by the spring, and first-grade children significantly surpassed the gap between Gardendale and the normative sample in the spring assessment; for the 2022–23 school year, a significant gap remained for both grades in the spring. For mCLASS literacy, comparable results were observed across both years, suggesting that children exhibited the same performance this year as last year. For DECA,

[21] The Emotional Support average increased from 5.77 to 6.11; the Classroom Organization increased from 5.28 to 5.74; and the Instructional Support average increased from 2.76 to 3.71.

comparable results were observed, and for mini-DESSA, the results showed an improvement this year compared to last year. This year, children demonstrated significant improvement in their social-emotional competency, whereas last year no significant growth was observed.

Comparing Assessment Results Across Measures

As was evidenced in the results section, for most measures and outcomes children demonstrated significant improvement. However, different conclusions arose when comparing early literacy and numeracy findings across multiple measures to the normative samples. This is consistent with previous research comparing GOLD and direct child assessments (Miller-Bains et al., 2017; Qiu et al., 2021; Russo et al., 2019).

When comparing findings for early literacy as shown in Table 2, pre-K and kindergarten children either performed significantly below or on par with the normative sample in the fall and spring on literacy and oral language (GOLD), which agreed with Letter-Word findings[22] and MAP reading findings. When comparing findings based on levels of understanding according to mCLASS literacy findings, the majority of children performed below benchmark levels when assessed in English and above benchmark levels when assessed in Spanish. For receptive vocabulary (PPVT), the majority of children performed in the expected range of understanding.

[22] This is based on the first row of WJ findings representing the majority of children from the random sample.

Table 2. Summary of early literacy findings across assessments

Outcome	Assessment	Grade	Domain	Growth	Percentage change	Gap closure[a]	Benchmark and performance levels[b]		Norm comparisons[c]	
				Fall to Spring			Fall	Spring	Fall	Spring
Early Literacy	GOLD	PK	Literacy	↑					↓	↓
			Oral Language	↑					—	—
		K	Literacy	↑					↓	↓
			Oral Language	↑					↓	↓
	WJ[d]	PK-2nd	Letter-Word	↑	—	—			↓	↓
		PK-2nd	Letter-Word (high achieving)	↑	—	↑			↑	↑
	PPVT[d]	PK-2nd	Receptive 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 triangle that points down indicates a negative significant result. Analyses were conducted based on the assessment scoring methods indicated in the technical manuals. Columns and rows without icons indicate those analyses were not conducted.

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

[a] PPVT findings for gap closure are based on descriptive statistics; no inferential tests were conducted.

[b] These findings are based on descriptive statistics; no inferential tests were conducted.

[c] WJ findings for norm comparisons are based on descriptive statistics; no inferential tests were conducted. Across all assessments, the norm comparisons were created 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.

[d] Assessments conducted for a random sample of children.

When comparing findings for early numeracy as shown in Table 3, some findings were consistent, and others were not. Fall assessment results for MAP demonstrated comparable results as the fall Applied Problems findings. However, pre-K and kindergarten children were either significantly above or on par with the normative sample (GOLD). 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.

Table 3. Summary of early numeracy findings across assessments

Outcome	Assessment	Grade	Domain	Growth	Percentage change	Gap closure	Benchmark and performance levels		Norm comparisons[a]	
				Fall to Spring			Fall	Spring	Fall	Spring
Early Numeracy	GOLD	PK	Mathematics	↑					—	—
		K		↑					↑	—
	WJ[b]	PK-2nd	Applied Problems	↑	—	—			↓	↓
	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 triangle that points down indicates a negative significant result. Analyses were conducted based on the assessment scoring methods indicated in the technical manuals. Columns and rows without icons indicate those analyses were not conducted.

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

[a] WJ findings for norm comparisons are based on descriptive statistics; no inferential tests were conducted. Across all assessments, the norm comparisons were created 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.

[b] Assessments conducted for a random sample of children.

When comparing findings for social-emotional competency as shown in Table 4, pre-K and kindergarten children were either significantly below or on par with the normative sample (GOLD). For DECA and DESSA, findings demonstrate that the majority of children were scoring at the Typical or Strengths levels. Given the different nature of these comparisons, it is not appropriate to compare them to one another.

Table 4. Summary of social-emotional findings across assessments

Outcome	Assessment	Grade	Domain	Growth	Percentage change	Gap closure	Benchmark and performance levels[a]		Norm comparisons[b]	
				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	—	—		↑	↑		
	DESSA (mini)	K-1st	Overall	↑	↑		↑	↑		
		2nd		—	—		↑	↑		

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 triangle that points down indicates a negative significant result. Analyses were conducted based on the assessment scoring methods indicated in the technical manuals. Columns and rows without icons indicate those analyses were not conducted.

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

[a] These findings are based on descriptive statistics; no inferential tests were conducted.

[b] Across all assessments, the norm comparisons were created 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.

When comparing these results across Tables 2, 3, and 4, there are four reasons why these assessments could provide different conclusions. First, they each used a different type of assessor. GOLD, DECA, and DESSA are teacher-reported assessments; WJ and PPVT are collected by an independent assessor; MAP is administered online; and mCLASS is hybrid (administered online for kindergarten and by reading booklet for first and second grade). 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.

Second, the GOLD oral language, literacy, and mathematics outcomes assess more content than WJ and PPVT. It is possible children score similarly on the GOLD content that mirrors the WJ and PPVT content but score higher on the content that is only captured on GOLD.

Third, there are different scoring methods for each assessment. GOLD results are based on comparing children to a single normative average, WJ and PPVT have age-specific (measured in months) normative averages, and DECA and DESSA are measured across three levels (so no normed comparisons are available). Therefore, it may be possible for children to show more nuanced understanding in WJ and PPVT than with GOLD literacy and oral language, which can lead to different conclusions.

Fourth, there is limited validity evidence in which these measures have been compared to determine how much content is similar and how much content is different across measures (e.g., Barghaus et al., 2022; Miller-Bains et al., 2017; Qiu et al., 2021; Russo et al., 2019). Therefore, these

measures may be measuring distinct aspects of early literacy, numeracy, and social-emotional competency.

Directions for Future Research

Taken together, these findings demonstrated children benefited from attending Gardendale. However, work is needed moving forward to further accelerate the learning of these children. The significant growth in early literacy and numeracy, vocabulary, and social-emotional competence provided empirical evidence of a step in the right direction. However, most results demonstrated children were performing below what would be expected for their grade based on national norms. 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; Diaz & Decker-Woodrow, 2021; Jung & Barnett, 2021; Socol, 2022; Weiland et al., 2021). Therefore, it is likely Gardendale children entered the 2023–24 school year with preexisting learning and achievement challenges. Taken together, these findings continue to shed light on Gardendale by not only providing results 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.

Limitations and Recommendations

We wish to highlight five limitations related to these findings First, it was not possible to collect data from a comparison school with which to compare the Gardendale children because of resource constraints. Therefore, normative samples were used for comparisons. This is important because (1) the normative sample data were collected prior to the COVID-19 pandemic and therefore do not take into account the pandemic-related disruption to learning and well-being that has occurred for children and families, and (2) the normative samples are most likely quite different from the Gardendale children and did not experience learning in the same context. Normative samples are created to be reflective of the demographic proportions similar to those found in the U.S. Census data and were constructed prior to the pandemic during a "typical" school year. 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.[23]

Second, the learning and growth children displayed during the school years following the COVID-19 pandemic is most likely different than the school years prior to the pandemic. 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). Given these disruptions and setbacks, it is likely most children are exhibiting different amounts of growth and learning in the 2023-24 school year compared to the growth and learning demonstrated in the normative comparisons. Given this, the normative comparisons are not ideal because children's current learning and their trajectories have been altered as a result of

[23] One way to form such a 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.

the COVID-19 pandemic. However, this is the best research evidence currently available for comparison. Therefore, these results should be interpreted 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 such analyses can be conducted.

Third, the early literacy 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. Moreover, most of the high-achieving children attended Gardendale in 2022–23 as well. Therefore, they provide some empirical evidence demonstrating that significant achievement occurred for Gardendale children who attended for multiple consecutive years.

Fourth, the receptive vocabulary assessment, PPVT, was only available in English. Therefore, it was not possible to assess children in Spanish. The Spanish version, Test de Vocabulario en Imagenes Peabody (Dunn et al., 1986), has been discontinued by the publisher.[24] In the 2024–25 school year, children will be assessed using the Receptive and Expressive One-Word Picture Vocabulary Tests in English and Spanish (Martin, 2013a, 2013b; Martin & Brownell, 2011a, 2011b).

Fifth, the GOLD and DECA findings are based on children for whom data were available and do not represent all Pre-K 4 SA children.[25] Pre-K 4 SA is aware of this limitation and is working to increase the data availability for the 2024–25 school year.

[24] The testing easels to administer the assessment in Spanish have been discontinued but the score sheets are still available. As the measure is being phased out, this led the team to assess children in English only and use a different measure with an accompanying version for Spanish speaking children for the 2024-25 school year.

[25] Demographic tests of differences were conducted to determine if the sample of children included in and excluded from analyses were similar (see Appendix A, Analytic Approach for more detailed information).

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.

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.

Bassok, D., Magouirk, P., & Markowitz, A. J. (2021). Systemwide quality improvement in early childhood education: Evidence from Louisiana. *AERA Open*, *7*, 1–19.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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

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

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.

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.

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

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

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

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

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.

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

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), 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.

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.

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

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.

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.