NIEER
NATIONAL INSTITUTE FOR EARLY EDUCATION RESEARCH

EVALUATION OF THE PHILADELPHIA PREK PROGRAM

2023-2024 Pilot Final Report

Erin Harmeyer, PhD, Milagros Nores, PhD, Charles Whitman, M.Ed. The National Institute for Early Education Research

Suggested citation: Harmeyer, E., Nores, M., & Whitman, C. (2025). *Evaluation of the Philadelphia PreK program. 2023-2024 Pilot final report.* National Institute for Early Education Research.

Introduction

In recent years, the City of Philadelphia put into motion a series of quality improvement efforts for its PHLpreK program. In particular, the City contracted with SPARK Learning Labs to support teachers through professional development and coaching efforts. This report summarizes the results of the 2023-2024 pilot taking place in the year 8 evaluation of the PHLpreK program. The purpose of this pilot is to support SPARK Learning Labs, the coaching organization contracted by the City of Philadelphia, in tailoring coaching in PHLpreK classrooms that were identified as being in need of additional supports. For this pilot, 100 classrooms were selected across 89 sites in the PHLpreK program to take part in a fall observation, completed between approximately October – December 2023 (4 programs were observed in January due to repeated conflicts with scheduling the observations). A follow-up spring observation also took place between March – June 2024. This brief summarizes characteristics of the pilot sample including rationale for being selected, teacher demographics, and results of the pilot, including how CLASS scores changed from fall to spring of the school year.

Methods

The PHLpreK Evaluation is a multi-year, multi-site study encompassing several components to provide a comprehensive perspective of the program's design, its quality, and its impact on children over time. Since 2023, the evaluation has focused on yearly classroom quality observations at the request of the City of Philadelphia. This report focuses on a pilot taking place in the eighth year of the program. Data collection included assessing a subset of 100 classrooms to address the following questions:

1. What is the observed quality of children's classroom experiences for a sample of 100 classrooms in the PHLpreK program participating in a pilot program?
2. Does quality in these classrooms differ as a function of why they were selected for the pilot (i.e., as a new program, as a classroom with scores below a set cut point in 2023, or as a program that previously did not complete scheduled observations)?
3. How do CLASS scores in the spring of 2024 compare to fall 2023 scores for the pilot participants? Does the change in scores differ as a function of whether the providers participated in the supports provided by SPARK?

Classroom observations were conducted to assess teacher-child interactions and quantify children's experiences during a typical learning day. Fall classroom observations for the pilot program took place between October and December 2023, and 4 classrooms had observations rescheduled to January due to illness and other issues in the classroom. The full universe of PHLpreK classrooms were observed from January – June 2024. In addition, pilot classrooms were followed up with beginning in March and continuing through the end of the school year. As in previous years, quality was assessed using well-known observation protocols during one visit of about two and a half hours. The fall and spring observations for classrooms in the pilot were on average 6.0 months apart (minimum 3.0 & maximum 8.1).

1. Sample

In the 2023-24 school year, NIEER assessed classroom quality with the CLASS Second Edition (pre-K – 3rd). The CLASS was used in 100 PHLpreK classrooms (center and home-based) in the fall pilot, and 264 classrooms over the winter. In addition, we followed up with fall pilot classrooms in the spring, although we were unable to reobserve three pilot classrooms in the spring. Of the classrooms with data at both time points, in 8 classrooms, the teacher observed in the spring was different than the teacher observed in the fall due to teacher turnover.

Before the observation, data collectors asked teachers a few questions about their highest educational attainment, area of study, and demographics. In terms of race/ethnicity, in the fall, 4.0% of teachers observed identified as Asian, 74.0% identified as Black/African-American, 8.3% identified as Hispanic/Latino, 12.1% identified as White, 5.0% identified as Multi-racial, 1.0% identified as Other, and 1.0% did not respond (numbers do not add up to 100% because teachers could select multiple options). In the spring, 5.2% of teachers observed identified as Asian, 73.2% identified as Black/African-American, 8.0% identified as Hispanic/Latino, 13.5% identified as White, 7.3% identified as Multi-racial, and 2.1% identified as Other.

In terms of teacher credentials, in the fall 3.0% of teachers had completed no degree, 4.0% had completed some college, 22.0% had completed an associate's degree, 50.0% had completed a bachelor's degree, and 21.0% had completed master's degree. In terms of area of study, 63.9% identified their highest degree as being in Early Childhood Education (ECE), 5.2% identified Elementary Education, 9.3% identified Education: Other (e.g., Special Education, Art Education), 9.3% identified Psychology, and 12.4% identified Other.

In the spring, 1% had no degree, 5.2% had completed some college, 22.7% had completed an associate's degree, 49.5% had completed a bachelor's degree, and 21.6% had completed a master's degree. Finally, in terms of area of study, in the spring, 63.6% identified ECE, 5.1% identified Elementary Education, 9.1 % identified Psychology, 10.1% identified Education: Other, and 12.4% identified Other.[1]

In terms of class sizes, in the fall classrooms had an average of 9.57 children present (this represents the average number of children present during the observation across all cycles observed), with a minimum of 1 and a maximum of 22. In terms of class size in the spring, classrooms had an average of 9.33 children present, with a minimum of 2 and a maximum of 21. Breaking this down by site type, child care centers had an average of 10.52 children present in the fall, and 10.14 in the spring. Family child care/group family child care had an average of 5.80 children in the fall, and 6.09 children in the spring. Finally, with regards to the type of sites engaged in the pilot, 80 of the classrooms were in child care centers, and 20 classrooms were in family homes (with 11 in family child care homes and 9 in group child care homes).

2. Measures and Procedures

Classroom quality was captured using one observational instrument: *The Classroom Assessment Scoring System Pre-K- 3rd Second Edition* (*CLASS 2.0*; Pianta & Hamre, 2022). The CLASS measures teacher-child interactions and classroom processes; this was the second year using the second edition of the observation tool. According to the measure developers, CLASS Second

[1] The small changes from fall to spring are driven by the change in number of classrooms observed (100 in the fall and 97 in the spring) and the change of classroom teacher in 8 classrooms during the school year.

Edition was developed using more equitable and inclusive measures of effective interactions, and includes increased representation of children and teachers in training materials. The developers state it allows for consideration of possible variations in effective interactions due to context, and aims to help observers confront bias in their own observations (Pianta & Hamre, 2022). Notably, the updated tool does not make any changes to the dimensions or domains that are scored; revisions focus on broadening the description of effective interactions.

Observers were trained to reliability before conducting observations of classroom quality. CLASS observers were trained using the Teachstone® virtual training platform, completed the online reliability certification test required by Teachstone® and met their requirement (80%) for observer certification. Observers were also trained in practices and procedures for conduct and required to complete background checks, as well as training in human subjects research (human subject protections, ethical issues, etc.).

Results

Results are presented for classroom observations from the pilot below. These include a breakdown of scores across all 10 dimensions and three domains. For comparison, average winter/spring scores from the full universe of classrooms in PHLpreK is included.

Classroom Observations

CLASS Pre-K Results: Fall and Spring Pilot Sample

Average CLASS scores for PHLpreK classrooms for all domains and dimensions in the pilot sample are reported in Table 1, broken down by fall and spring scores. Patterns are relatively consistent with the field and previous years, with Instructional Support scoring lower than other domains. Emotional Support (ES) scores for the spring pilot sample are 5.88 (compared to 5.94 for the full universe). Classroom Organization (CO) scores are 5.48 (compared to 5.55 for the full universe). Finally, scores on the Instructional Support domain are 2.45 (compared to 2.40 for the full universe). Across the three domains, the pilot group showed improvement in average scores. Using a paired-samples t-test we tested whether the domain and dimension scores between fall and spring statistically differed for the pilot classrooms. Scores were statistically significantly higher in all domains, as well as multiple dimensions in the spring as compared to the fall.

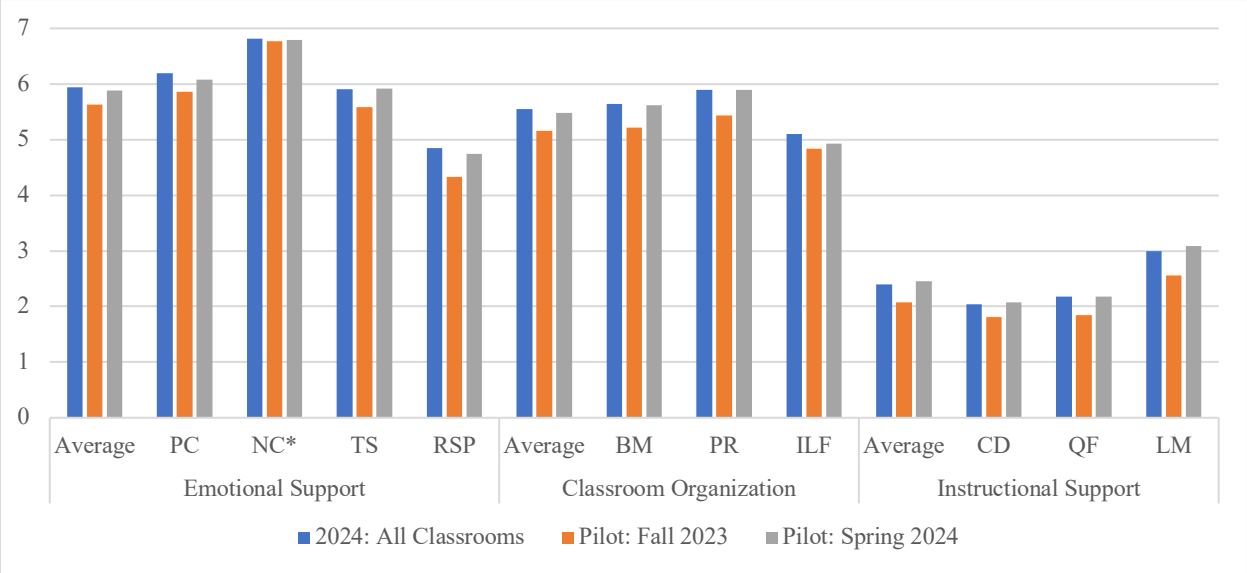
Table 1. Average scores for pilot classrooms, fall 2023 and spring 2024.

CLASS Dimensions and Domains	2024 PHLpreK All Classrooms Mean (Range) N=364	2023 Fall Pilot Sample Mean (Range) N=100	2024 Spring Pilot Sample Mean (Range) N=97
Emotional Support Domain (ES)	**5.94**	**5.63**	**5.88***
	(2.40-7.00)	**(2.50-6.95)**	**(2.40-6.95)**
1. Positive Climate (PC)	6.19	5.86	6.08
	(2.60-7.00)	(1.80-7.00)	(2.60-7.00)
2. Negative Climate* (NC)	6.82	6.77	6.79
	(4.00-7.00)	(2.40-7.00)	(4.00-7.00)
3. Teacher Sensitivity (TS)	5.91	5.58	5.92*

	(1.40-7.00)	(1.40-7.00)	(1.40-7.00)
4. Regard for Student Perspectives (RSP)	4.85	4.33	4.74*
	(1.40-7.00)	(1.40-6.80)	(1.40-6.80)
Classroom Organization Domain (CO)	**5.55**	**5.16**	**5.48***
	(1.73-7.00)	**(1.33-7.00)**	**(1.73-6.87)**
5. Behavior Management (BM)	5.64	5.22	5.62*
	(1.40-7.00)	(1.60-7.00)	(1.40-7.00)
6. Productivity (PR)	5.90	5.44	5.89*
	(1.80-7.00)	(1.20-7.00)	(1.80-7.00)
7. Instructional Learning Formats (ILF)	5.10	4.84	4.93
	(2.00-7.00)	(1.20-7.00)	(2.00-7.00)
Instructional Support Domain (IS)	**2.40**	**2.07**	**2.45***
	(1.00-4.75)	**(1.00-4.13)**	**(1.00-4.73)**
8. Concept Development (CD)	2.04	1.81	2.07*
	(1.00-4.60)	(1.00-3.80)	(1.00-4.20)
9. Quality of Feedback (QF)	2.18	1.84	2.18*
	(1.00-5.20)	(1.00-4.40)	(1.00-4.60)
10. Language Modeling (LM)	2.99	2.56	3.09*
	(1.00-5.80)	(1.00-5.20)	(1.00-5.40)

Note: *Paired samples t-tests show a statistically significant difference ($p<.05$) between fall and spring scores on these dimensions/domains. Note: Teacher Sensitivity drops to significant only at the trend level (i.e., $p<.10$) when the 8 classrooms with a teacher switch are excluded from the analysis. Averages here show the full pilot group, including classrooms with a teacher switch during the 2023-2024 school year.

Figure 1. CLASS scores for all classrooms, and in fall and spring for the pilot subgroup.



The following table breaks down CLASS scores in the fall as a function of the reason programs were selected for the pilot. In consultation with the City, NIEER selected 100 classrooms to be part of the pilot sample. This included all classrooms new to PHLprek in 2023-24 (n=48), classrooms that missed scheduled observations in 2023 (n=11), and classrooms that scored lower than 5.25 on Emotional Support *and* lower than or equal to 2.2 on the Instructional Support domain in 2023 (n=41). Of these 41 programs, 28 were initially selected for having Emotional Support scores lower than 5 *and* IS scores lower than 2.0, additional classrooms were added with slightly higher scores to reach a target sample of 100, and to replace classrooms that were

removed from the pilot (i.e., because classrooms were merged into one or did not open due to low enrollment).

Table 2. Pilot Sample Breakdown, Fall 2023.

CLASS Dimensions and Domains	New classrooms	Not observed year prior	Chosen for prior year scores
	Mean (Range) N=48	Mean (Range) N=11	Mean (Range) N=41
Emotional Support Domain (ES)	**5.77*** **(2.50-6.95)**	**5.86** **(4.95-6.45)**	**5.41*** **(2.50-6.69)**
1. Positive Climate	5.90 (1.80-7.00)	6.16 (5.20-7.00)	5.72 (3.60-7.00)
2. Negative Climate*	6.83 (4.80-7.00)	6.98 (6.80-7.00)	6.65 (2.40-7.00)
3. Teacher Sensitivity	5.80[b] (1.40-7.00)	6.09 (4.60-7.00)	5.19[b] (1.60-7.00)
4. Regard for Student Perspectives	4.58* (1.80-6.80)	4.20 (2.80-5.60)	4.06* (1.40-6.25)
Classroom Organization Domain (CO)	**5.39[a]** **(1.33-7.00)**	**5.62** **(4.47-6.73)**	**4.78[a]** **(2.40-6.73)**
5. Behavior Management	5.40[b] (1.60-7.00)	5.95 (5.00-6.80)	4.80[b] (2.40-7.00)
6. Productivity	5.65[b] (1.20-7.00)	5.84 (4.60-6.80)	5.08[b] (1.40-7.00)
7. Instructional Learning Formats	5.12[b] (1.20-7.00)	5.09 (3.40-6.60)	4.45[b] (2.00-6.60)
Instructional Support Domain (IS)	**2.21[a]** **(1.00-4.13)**	**2.03** **(1.40-2.80)**	**1.91[a]** **(1.27-3.27)**
8. Concept Development	1.98[b] (1.00-3.40)	1.62 (1.00-3.00)	1.65[b] (1.00-3.80)
9. Quality of Feedback	1.95[b] (1.00-4.40)	1.85 (1.00-2.60)	1.71[b] (1.00-2.80)
10. Language Modeling	2.70* (1.00-5.20)	2.62 (1.60-3.60)	2.38* (1.20-4.60)

Note: We ran t-tests to compare scores between new classrooms and classrooms chosen for previous scores.
[a]Domain score significantly lower for classrooms chosen for previous scores than for new classrooms ($p<.05$).
[b]Dimension score significantly lower for classrooms chosen for previous scores than for new classrooms ($p<.05$).
*Scores significantly different between classrooms chosen for previous scores and new classrooms at the trend level ($p<.10$).

As shown in Table 2, scores were the lowest across the board for classrooms that were selected for the sample based on their 2022-2023 observation score. As a whole, scores on Emotional Support for the universe of classrooms in spring 2023 (5.81) were similar for new sites (5.77) and sites not observed in the prior school year (5.86), with classrooms chosen for their prior scores also averaging lower scores in the fall (5.41). Similarly, scores on Classroom Organization in spring 2023 for the full universe (5.42) were similar to those for new sites (5.39) and those not observed the year prior (5.62). Classrooms chosen for previous scores, however, scored lower than the other groups (4.78). Finally, in terms of Instructional Support – as in previous years, scores were low for all groups. While the full universe of classrooms scored an average of 2.45 in spring 2023, new sites averaged 2.21, sites not observed the prior year averaged 2.03, and sites chosen for their prior observation scores averaged 1.91.

We conducted independent samples t-tests between new classrooms and classrooms selected for their previous scores, to understand whether these groups differed significantly from one another in the fall. Multiple statistically significant differences emerged: In terms of domains, classrooms selected due to their previous observation scored statistically significantly lower than new classrooms on both Instructional Support and Classroom Organization, and their scores were also lower on Emotional Support (significant at the trend level, $p<.10$). They also scored significantly lower on Teacher Sensitivity, Productivity, Instructional Learning Formats, Concept Development, and Quality of Feedback. Scores were also lower at the trend level in Language Modeling and Regard for Student Perspectives.

Table 3 breaks down scores in spring 2024 as a function of why classrooms were selected for the pilot. Classrooms chosen for their prior observation are higher on ES and CO than new classrooms. In terms of Instructional Support, classrooms chosen for their previous observation look similar to the new classrooms (2.48 for new and 2.45 for classrooms chosen for their prior observation), and scored higher than the classrooms that were not observed the prior year. And when conducting independent samples t-tests between classrooms chosen for their prior scores and new classrooms, only scores on Productivity look significantly different between groups, such that classrooms selected due to their prior observation scored statistically significantly higher than new classrooms at the trend level. Thus, scores between the two groups converged over the school year, with many of the bigger differences between groups either disappearing or shrinking. In the next section, the change in scores from fall to spring is discussed further.

Table 3. Pilot Sample Breakdown, Spring 2024.

CLASS Dimensions and Domains	New classrooms Mean (Range) N=47	Not observed year prior Mean (Range) N=10	Chosen for previous scores Mean (Range) N=40
Emotional Support Domain (ES)	**5.82 (3.35-6.90)**	**6.26 (5.50-6.95)**	**5.86 (2.40-6.70)**
1. Positive Climate	6.00 (2.60-7.00)	6.36 (5.40-7.00)	6.10 (2.80-7.00)
2. Negative Climate*	6.78 (5.40-7.00)	6.92 (6.60-7.00)	6.76 (4.00-7.00)
3. Teacher Sensitivity	5.84 (2.40-7.00)	6.50 (5.80-7.00)	5.86 (1.40-7.00)
4. Regard for Student Perspectives	4.64 (2.40-7.00)	5.24 (3.60-6.80)	4.73 (1.40-6.20)
Classroom Organization Domain (CO)	**5.34 (2.87-6.87)**	**5.81 (4.27-6.73)**	**5.57 (1.73-6.67)**
5. Behavior Management	5.58 (2.40-7.00)	6.18 (5.00-7.00)	5.54 (1.40-7.00)
6. Productivity	5.69* (3.20-7.00)	6.10 (4.00-6.80)	6.07* (1.80-7.00)
7. Instructional Learning Formats	4.74 (2.20-6.60)	5.16 (3.20-6.80)	5.09 (2.00-7.00)
Instructional Support Domain (IS)	**2.48 (1.00-4.73)**	**2.26 (1.47-3.00)**	**2.45 (1.13-3.80)**
8. Concept Development	2.09 (1.00-4.20)	1.86 (1.20-2.80)	2.10 (1.00-4.00)
9. Quality of Feedback	2.26 (1.00-4.60)	1.82 (1.00-3.20)	2.18 (1.00-3.80)
10. Language Modeling	3.10 (1.00-5.40)	3.10 (2.00-5.00)	3.08 (1.40-4.80)

Note: We ran independent samples t-tests to compare scores between new classrooms and those chosen for previous observations, in spring 2024. *Scores significantly different between new classrooms and those chosen for their previous observation at the trend level ($p<.10$).

Changes in CLASS Scores: 2023-2024 School Year

CLASS Pre-K Score Changes: Pilot Sample

To compute the change in CLASS scores over the school year, we subtracted the fall 2023 CLASS score from the spring 2024 CLASS score. Table 4 shows the average change in scores from fall to spring for all classrooms, and then the table is split as a function of the reason classrooms were selected into the pilot. As seen in the table, scores in all domains and dimensions were higher in the spring 2024 than they were in the fall 2023. There are some notable differences in these changes. Some dimensions showed relatively little change (i.e., Negative Climate increased on average by 0.02 points from fall to spring). Others, however, were larger, including the change on the Language Modeling dimension (increasing on average 0.53 points from fall to spring) and the Productivity and Behavior Management dimensions (increasing on average 0.45 and 0.42 points, respectively). All three domains showed positive

changes from fall to spring: For Emotional Support, this increase was 0.25 points; for Classroom Organization, this growth in scores was on average of 0.30 points, and for Instructional Support, this change was 0.37 points.

When looking at these changes as a function of pilot group selection criteria, differences emerge. With the exception of one dimension (Regard for Student Perspectives), the change in scores is highest for the group of classrooms selected due to prior year observation scores. These classrooms demonstrated greater improvements in scores than did the new classrooms as well as the group of classrooms that had not been observed the year prior. In some dimensions, these gains are quite pronounced, with these classrooms that had been selected into the pilot due to scores in the year prior improving on average 0.94 on the Productivity dimension, 0.70 on the Behavior Management dimension, and 0.76 on the Classroom Organization domain. In the fall, these classrooms averaged lower scores than the other categories; in the spring, these classrooms were scoring on par or higher than the other classrooms in most domains and dimensions.

Table 4. Pilot Sample Score Changes

CLASS Dimensions and Domains	2023-2024 All Pilot Classrooms Mean Change (Range) N=97	2023-2024 New Classrooms Mean Change (Range) N=47	2022-2023 Classrooms not observed year prior Mean Change (Range) N=10	2022-2023 Classrooms selected for prior year scores Mean Change (Range) N=40
Emotional Support Domain (ES)	**0.25 (-3.20-2.95)**	**0.05* (-2.45-2.95)**	**0.37 (-0.40-1.35)**	**0.45* (-3.20-2.80)**
1. Positive Climate	0.22 (-4.20-3.60)	0.10 (-4.20-3.60)	0.12 (-1.40-1.40)	0.39 (-4.20-3.00)
2. Negative Climate*	0.02 (-3.00-3.80)	-0.04 (-1.40-1.60)	-0.06 (-0.20-0.00)	0.11 (-3.00-3.80)
3. Teacher Sensitivity	0.33 (-4.60-4.40)	0.06* (-3.40-4.40)	0.36 (-0.80-2.00)	0.65* (-4.60-4.00)
4. Regard for Student Perspectives	0.41 (-2.40-4.20)	0.07[a] (-2.40-4.00)	1.06 (-0.40-3.20)	0.66[a] (-1.80-4.20)
Classroom Organization Domain (CO)	**0.30 (-2.40-3.67)**	**-0.05[a] (-2.40-2.87)**	**0.07 (-1.27-1.80)**	**0.76[a] (-2.40-3.67)**
5. Behavior Management	0.39 (-3.20-4.60)	0.18* (-2.20-3.20)	0.18 (-1.20-1.20)	0.70* (-3.20-4.60)
6. Productivity	0.43 (-3.20-4.00)	0.05[a] (-3.20-3.20)	0.14 (-1.80-2.00)	0.94[a] (-2.60-4.00)
7. Instructional Learning Formats	0.07 (-3.80-3.20)	-0.37[a] (-3.80-3.20)	-0.10 (-2.40-2.20)	0.62[a] (-2.40-3.20)

Instructional Support Domain (IS)	**0.37 (-2.60-2.27)**	**0.27 (-2.60-1.73)**	**0.17 (-0.80-1.33)**	**0.53 (-0.93-2.27)**
8. Concept Development	0.25 (-1.80-2.80)	0.10* (-1.80-2.00)	0.18 (-1.20-1.20)	0.44 (-1.20-2.80)
9. Quality of Feedback	0.33 (-3.40-3.40)	0.31 (-3.40-3.40)	-0.12 (-1.60-1.40)	0.47 (-1.40-2.20)
10. Language Modeling	0.52 (-2.60-2.60)	0.39* (-2.60-2.60)	0.44 (-1.20-2.40)	0.69* (-1.80-2.60)

Note: We ran t-tests to compare changes in scores between new classrooms and those selected for their prior observation. [a]Independent samples t-test between new classrooms and classrooms chosen for their previous observation revealed a statistically significant difference between these groups on change in scores. *Change in scores was significant at the trend level (p<.10) between new classrooms and those chosen based on previous scores.

CLASS Pre-K Scores: Differences in Classrooms Utilizing Supports

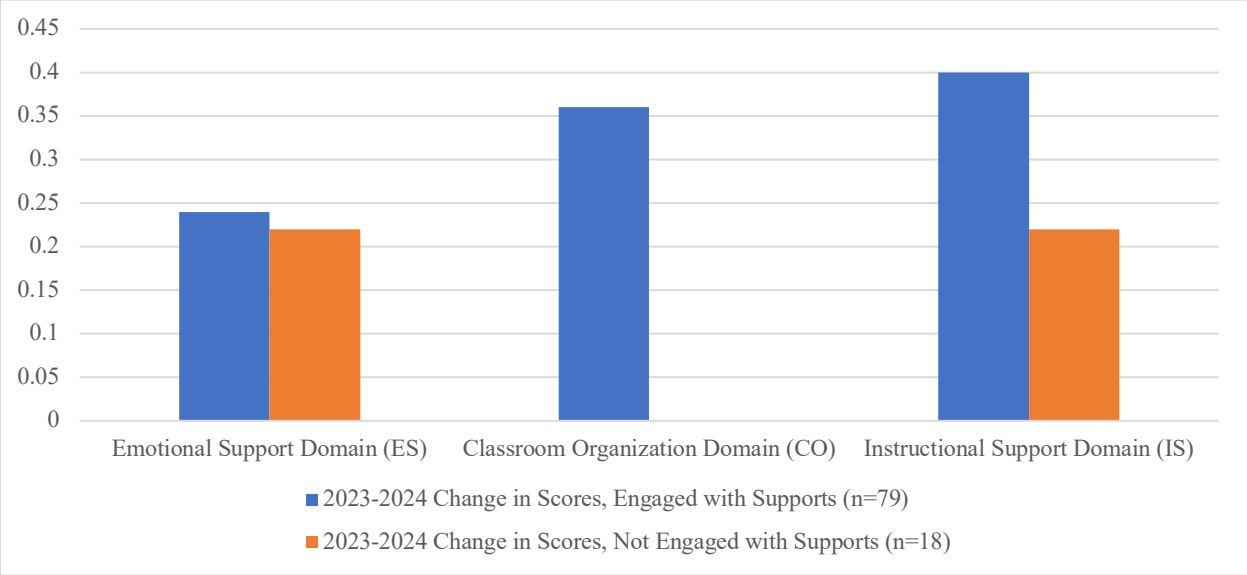
Although it was the intention that all classrooms included in the pilot would utilize the supports offered by SPARK, the early childhood education coaching organization contracted by the City, not all classrooms participated in the provided supports. We therefore also broke down the change in scores for those classrooms that SPARK reported participated in the supports, and for those that opted out. As seen in Table 5, those classrooms that opted to engage with the supports demonstrated higher average levels of growth than did the classrooms that did not engage in the supports. This was most pronounced for the Classroom Organization domain and also quite significant for the Instructional Support domain.

Table 5. Change in scores for pilot classrooms, as a function of engagement with supports.

CLASS Dimensions and Domains	2023-2024 Change in Scores, Engaged with Supports (Range) N=79	2023-2024 Change in Scores, Not Engaged with Supports (Range) N=18
Emotional Support Domain (ES)	0.25	0.22
	(-3.20-2.95)	(-0.90-1.45)
Classroom Organization Domain (CO)	0.36	0.00
	(-2.40-3.67)	(-2.20-1.93)
Instructional Support Domain (IS)	0.40	0.22
	(-2.60-2.27)	(-1.40-1.73)

Note: No statistically significant differences across groups.

Figure 2. Change in scores for pilot groups engaged and not engaged in supports.



Note*:* On average, scores did not change from fall to spring on the Classroom Organization domain for pilot classrooms not engaged with the supports.

Conclusion

This was the first year that PHLpreK engaged 100 selected classrooms in a pilot in which teachers were provided additional instructional supports. The results of this pilot are promising; the mean scores from fall to spring changed in a positive direction for all 10 dimensions and all domains. Although some changes were small (e.g., Negative Climate changed 0.02), others were quite large, such as in the Language Modeling dimension (mean scores changed from 2.56 in the fall to 3.09 in the spring) and in the Productivity dimension (mean scores changed from 5.44 in the fall to 5.89 in the spring).

Of note in these pilot results is that the reasons sites were selected for the pilot made a difference in the amount of change demonstrated in observed classroom quality over the school year. When comparing the classrooms that were selected for the pilot because they were new sites in the 2023-2024 school year with the classrooms that were selected based on prior observation scores, those classrooms chosen based on past observations demonstrated more growth on all domains and dimensions, and in some cases, these differences were significant. This was the case on the Classroom Organization domain, in which previously observed sites showed an average improvement of 0.76 points, as compared to -0.05 points at new sites (meaning the Classroom Organization scores were on average just slightly lower in the spring for new sites than they were in the fall). This was also the case for Regard for Student Perspectives (0.66 change in classrooms chosen for prior scores; 0.07 at new classrooms), Productivity (0.95 change at classrooms based on prior scores; 0.05 at new classrooms), and Instructional Learning Formats (0.62 change in classrooms chosen based on previous scores; -0.37 at new classrooms).

These differences in scores point to preliminary evidence that teachers enrolled in the pilot are receiving supports that positively impact their classroom practice. Data from pilot participants seems to support this – for example, the pilot participants that chose to engage in the

supports showed greater growth in their CLASS scores across all three dimensions. For the pilot for the 2024-2025 school year, PHLpreK should continue to focus on providers who are willing and able to engage with the provided supports, in order to see if this positive growth trend continues. This can provide further information to PHLpreK about the types of supports that best support teacher practice. In addition, understanding what is holding some providers from engaging, in order to find pathways to increase participation, would allow the system to potentially evidence higher levels of growth.