NIEER
NATIONAL INSTITUTE FOR EARLY EDUCATION RESEARCH

# EVALUATION OF THE PHILADELPHIA PREK PROGRAM

## Year 8 Report

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

About the Authors

**Erin Harmeyer, Ph.D.** Dr. Harmeyer is an Assistant Research Professor at the National Institute for Early Education Research (NIEER) at Rutgers University. Dr. Harmeyer conducts research at NIEER related to early childhood programs and evaluation, supporting projects in New Jersey, Rhode Island, and Pennsylvania.

**Milagros Nores, Ph.D.** Dr. Nores is Co-Director of Research and Associate Research Professor at The National Institute for Early Education Research (NIEER) at Rutgers University. Dr. Nores conducts research at NIEER on early childhood policy, programs, and evaluation, both nationally and internationally.

**Charlie Whitman, M.Ed.** Whitman is Project Coordinator at the National Institute for Early Education Research (NIEER). He leads NIEER's field work on the PHLpreK Evaluation Study and related work in New Jersey. He has contributed to NIEER's research in Philadelphia and New Jersey.

Correspondence regarding this report should be addressed to Erin Harmeyer at the National Institute for Early Education Research. Email: eharmeyer@nieer.org.



Suggested citation: Harmeyer, E., Nores, M., & Whitman, C. (2024). *Evaluation of the Philadelphia PreK program. Year 8 report.* National Institute for Early Education Research.

## Table of Contents

Table of Contents ..... ii
Introduction ..... 3
Study Methods ..... 4
1. Sample ..... 5
2. Measures and Procedures ..... 5
Results ..... 6
1. Classroom Observations ..... 6
*CLASS Pre-K Results* ..... 6
*CLASS Pre-K – 3rd Domains* ..... 9
*CLASS Pre-K Comparison of Programs* ..... 11
*2023-2024 School Year Pilot: Summary of Results* ..... 12
Discussion of Findings ..... 14
Acknowledgments ..... 16
References ..... 17
Appendix A. Measures ..... 20

# Introduction

Philadelphia's Preschool Program (PHLpreK) has recently concluded its eighth year of programming. The program was initiated after a May 2015 vote where city voters approved the creation of the Philadelphia Commission on Universal Pre-kindergarten. The commission was given the responsibility of proposing a universal pre-K program to provide high-quality, affordable, and accessible services to the city's three- and four-year-olds. The National Institute for Early Education Research (NIEER) at Rutgers is in its eighth year of conducting annual evaluations assessing program quality. This report outlines findings from this year's evaluation.

Previous reports have summarized the importance of high-quality preschool education to reduce persistent achievement gaps in kindergarten and throughout primary (e.g., Barnett et al., 2018; Harmeyer et al., 2023; Nores et al., 2017, 2018, 2019, 2020, 2021, 2022). We have highlighted research that has shown that high-quality preschool education programs can produce lasting effects on school success and achievement and reduce achievement gaps at kindergarten entry and beyond (e.g., Barnett, 2008; Barnett & Jung, 2021; Barnett.& Nores, 2015; Ceci & Papierno, 2005; Duncan & Murnane, 2011; Gray-Lobe, 2023; Johnson et al., 2023). Strengthening and supporting preschool systems to achieve and sustain high quality requires continuous systems of improvement that include measurement and assessment, training and technical assistance and use of data to align system weaknesses and strengths with the initiative to increase quality over time (Barnett & Frede, 2017; Nores & Harmeyer, 2023). This includes understanding the quality of classroom processes and interactions (Pianta & Hamre, 2009; Hamre et al., 2014).

This report summarizes the results of the year eight evaluation of Philadelphia's PreK Program (PHLpreK). The report provides an overview of the environment and teaching interactions in these classrooms. The present report builds from classroom quality observations in the program since 2017, with the goal of supporting a data-driven continuous improvement approach to support improvements in quality in the city's program, alongside its expansion.

Findings indicate that PHLpreK classrooms continue to consistently demonstrate high to moderate levels of quality in the Emotional Support and Classroom Organization domains, meaning teachers are consistently interacting in a positive and sensitive way with children, and are structuring the classroom day and environment in a way that supports learning. Additionally, scores on the Emotional Support domain were significantly higher in 2023-2024 as compared to the prior school year, moving to 5.94 from 5.81. This occurred despite continued program expansion: In the 2023-2024 school year, PHLpreK incorporated 48 new classrooms at 46 new sites. Classroom Organization scores also increased, although at a slightly smaller rate, another positive trend highlighted in this year's report. In contrast, classroom scores for Instructional Support are low and this trend has persisted over time (i.e., lower than 3.0, on average), including a slight decrease in scores from the last school year in which we conducted the evaluation. We explore quality separately for several subgroups of interest, including Star level, lead teacher credentials, PHLpreK partner agency, and new and returning sites. Very small (but not significant) differences were found between subgroups, and are reported.

Importantly, this was the first year in which the PHLpreK program engaged in a small observation and coaching pilot, in which 100 classroom teachers participated in observations in the fall and the spring, and received additional coaching supports throughout the year. The sample of classrooms selected was composed of classrooms new to the program, those with previous low scores on the Instructional Support and Emotional Support domains, as well as a

small group of teachers who were not observed in the previous school year. Results show that, on average, teachers in the pilot improved in all three domains from the fall of 2023 to the spring of 2024, and that some subgroups of teachers, including those who engaged in the supports provided by SPARK, demonstrated greater growth in classroom observation scores over the school year. While there is a full, separate report on the pilot included as part of this evaluation, a summary of their results is included in this report.

## Study Methods

The PHLpreK Evaluation started as 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 2022, these efforts have evolved into annual observations of classroom quality to inform the program and its technical assistance partners. This report therefore focuses on an eighth year of classroom quality for the program. Data collection included classroom observations across all providers to inform the following research questions:

1. What is the observed quality of children's classroom experiences and how does it compare relative to the prior years of PHLpreK?
2. Are there differences in the quality of children's classroom experiences based on classroom characteristics, including star level, hub, program type, whether the program is new or returning, or teacher qualifications?
3. How does quality in PHLpreK classrooms compare to quality in other programs in the country?
4. What is the observed quality of PHLpreK classrooms participating in a pilot program offering additional coaching and supports to teachers? Does this differ for teachers who participated in the coaching supports?

The PHLpreK evaluation efforts were designed to assess program progress and quality over time, with the goal of informing a continuous improvement approach to quality. In Year 1, the research team measured classroom quality. In Years 2 and 3, the research team assessed children's learning and development at the beginning and end of the school year and repeated the observations of classroom quality. In Year 4, the research team collected some classroom and child-level data, but study procedures were interrupted by the onset of the COVID-19 pandemic. In year 5 (2020-21), teachers completed a self-report measure of classroom quality, and directors participated in focus groups discussing the impact of the pandemic on their programming. In year 6, data was collected from all classrooms enrolled in the PHLpreK program in order to measure quality. In 7, quality was measured in all PHLpreK classrooms and a sample of other classrooms across the city, and learning gains for a subset of these children were assessed. In the current year's evaluation (year 8; 2023-2024), we observed quality in all PHLpreK classrooms during the winter/spring months of 2024. In addition, we observed quality in a subset of 100 classrooms who participated in a pilot program in which they were offered additional classroom supports. These programs were observed in the fall and spring in order to measure change in quality over the school year.

## 1. Sample

In the 2023-2024 school year, classroom quality was assessed with the CLASS Second Edition (pre-K – 3rd). The CLASS was used in 362 total PHLpreK classrooms (center and home-based) in winter/spring 2024. For the pilot subsample, 97 classrooms were observed two times (in the fall of 2023 and the spring of 2024) Three pilot classrooms were observed only in the fall. The remaining 265 PHLpreK classrooms were observed only in the winter/spring of 2024. We asked teachers to respond to questions about their highest level of education, area of study for highest degree, and race/ethnicity, and supplemented this information with registry data from PHMC when the information was missing. The final sample of 362 teachers had the following characteristics:

- 58 teachers reported a master's degree (16.4% of respondents), 138 a bachelor's degree (39%), 136 an associate's degree (38.4%), 17 some college (4.8%), and 5 no degree (1.4%). Eight teachers did not respond and had no information in the registry.
- The majority of teachers who responded (186 teachers, 67.6%) reported Early Childhood Education as their highest degree area; 19 teachers (6.9%) reported Elementary Education; 23 teachers reported Education: Other (e.g., Special Education, 8.4%); 16 teachers reported Psychology (5.8%); and 31 reported Other (11.3%). 87 teachers chose not to respond to this question.
- In terms of demographics, 22 respondents (6.1%) identified as Asian, 222 (61.3%) identified as Black/African American, 37 (10.2%) identified as Hispanic/Latino, 80 (22.1%) identified as White, 24 (6.6%) identified as Multi-Racial, and 4 (1.1%) identified as Other. Seven teachers chose not to answer and had no registry data. (*Respondents could select more than one option).*

## 2. Measures and Procedures

Classroom quality was captured using one observational instrument: *The Classroom Assessment Scoring System Pre-K- 3rd Second Edition* (*CLASS*; *Classroom Assessment Scoring System 2nd Edition,* 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 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 (*CLASS, 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. In addition, the revised tool covers the range of preschool through third grade classrooms.

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.). In addition, observers were required to pass a calibration assessment about mid-point through the data collection period.

# Results

The following sections report results of the classroom observations. Results are reported for the full sample of classrooms observed in the winter (i.e., all 362 PHLpreK classrooms that completed a winter observation). Additional results for the subsample of classrooms that participated in the pilot are also reported.

## 1. Classroom Observations

### *CLASS Pre-K Results*

Average CLASS scores for PHLpreK classrooms across all domains and dimensions are reported in Table 1. In terms of the domain scores, patterns are consistent with the field and previous years, with Instructional Support scoring lower than other domains. Emotional Support (ES) scores are higher than those recorded in year 7 (5.81 in 2023, compared to 5.94 in 2024), and these scores are significantly higher than those recorded in year 7. This is also the case for Classroom Organization (CO) (5.42 in 2023, compared to 5.55 in 2024), and these scores are higher than those recorded in year 7. However, scores on the Instructional Support domain are just slightly lower in 2024 than they were in 2023 (2.45 in 2023, compared to 2.40 in 2024). Scores are discussed in further detail below, including scores on dimensions.

Observed increases in this eighth year in ES were of 0.17 SD (standard deviations),[1] and in CO these increases were just slightly smaller, of 0.13 SD. In contrast, for IS there was a decrease although this was minimal: of 0.06 SD.

Table 1. PreK CLASS Dimension and Domain Means and Ranges.

| CLASS Dimensions and Domains | 2017 Mean (Range) N=139 | 2018 Mean (Range) N=137 | 2019 Mean (Range) N=147 | 2020 Mean (Range) N=102** | 2022 Mean (Range) N=270 | 2023 Mean (Range) N=283 | 2024 Mean (Range) N=364 |
|---|---|---|---|---|---|---|---|
| **Emotional Support Domain (ES)** | **5.85 (2.85-6.90)** | **5.64 (3.20-6.95)** | **6.01 (3.05-7.00)** | **5.74 (3.55-6.80)** | **5.86 (2.75-7.00)** | **5.81 (3.15-7.00)** | **5.94[a] (2.40-7.00)** |
| 1. Positive Climate | 5.90 (1.60-7.00) | 5.73 (3.20-7.00) | 6.13 (2.40-7.00) | 5.77 (3.20-7.00) | 5.95 (2.40-7.00) | 6.07 (3.20-7.00) | 6.19 (2.60-7.00) |
| 2. Negative Climate* | 6.77 (5.00-7.00) | 6.67 (4.00-7.00) | 6.91 (5.40-7.00) | 6.74 (4.2-7) | 6.78 (3.80-7.00) | 6.79 (3.40-7.00) | 6.82 (4.00-7.00) |
| 3. Teacher Sensitivity | 5.69 (2.20-7.00) | 5.52 (2.80-7.00) | 5.89 (1.60-7.00) | 5.58 (3.20-7.00) | 5.54 (1.80-7.00) | 5.69 (1.20-7.00) | 5.91[b] (1.40-7.00) |
| 4. Regard for Student Perspectives | 5.03 (2.00-6.80) | 4.65 (2.40-7.00) | 5.11 (1.60-7.00) | 4.88 (2.8-6.8) | 5.19 (2.00-7.00) | 4.70 (1.00-7.00) | 4.85[c] (1.40-7.00) |
| | **5.34** | **5.28** | **5.60** | **5.26** | **5.40** | **5.42** | **5.55** |

[1]Standard deviation is a measure of variation in the data. It measures how close together or spread apart the classrooms are relative to the mean. The larger the value, the farther apart from the mean classrooms are, and the smaller the value, the closer to the mean classrooms are, in a specific indicator, such as classroom size. It also helps to understand change, by dividing change by the standard deviation of the previous year. This helps understand how much of a standard deviation a distribution has changed.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Classroom Organization Domain (CO)** | **(1.87-6.93)** | **(2.80-6.93)** | **(2.40-7.00)** | **(3.20-6.80)** | **(1.87-6.93)** | **(2.07-7.00)** | **(1.73-7.00)** |
| 5. Behavior Management | 5.49 (1.60-7.00) | 5.48 (2.80-7.00) | 5.81 (2.40-7.00) | 5.54 (3.00-7.00) | 5.44 (2.00-7.00) | 5.67 (1.80-7.00) | 5.64 (1.40-7.00) |
| 6. Productivity | 5.76 (1.80-7.00) | 5.65 (2.80-7.00) | 5.72 (2.40-7.00) | 5.54 (3.40-7.00) | 5.76 (1.20-7.00) | 5.66 (1.80-7.00) | 5.90[b] (1.80-7.00) |
| 7. Instructional Learning Formats | 4.77 (1.60-7.00) | 4.72 (1.80-6.80) | 5.27 (2.00-7.00) | 4.68 (2.40-6.60) | 5.00 (1.80-7.00) | 4.93 (1.20-7.00) | 5.10 (2.00-7.00) |
| **Instructional Support Domain (IS)** | **2.41 (1.00-5.00)** | **2.05[a] (1.00-4.60)** | **2.54 (1.00-5.33)** | **2.30 (1.33-4.13)** | **2.75 (1.00-5.80)** | **2.45 (1.00-6.40)** | **2.40 (1.00-4.73)** |
| 8. Concept Development | 2.09 (1.00-4.80) | 1.84 (1.00-4.00) | 2.27 (1.00-5.60) | 2.10 (1.00-4.00) | 2.50 (1.00-6.60) | 2.10 (1.00-6.60) | 2.04 (1.00-4.60) |
| 9. Quality of Feedback | 2.23 (1.00-5.00) | 1.91 (1.00-4.40) | 2.53 (1.00-5.20) | 2.10 (1.00-4.20) | 2.65 (1.00-6.00) | 2.44 (1.00-6.60) | 2.18[b] (1.00-5.20) |
| 10. Language Modeling | 2.91 (1.00-5.20) | 2.41 (1.00-5.60) | 2.80 (1.00-5.80) | 2.70 (1.40-4.40) | 3.11 (1.00-6.00) | 2.80 (1.00-6.00) | 2.99[b] (1.00-5.80) |

*The Negative Climate dimension is reverse scored so that a high score represents "good." [a]Independent samples t-tests assuming equal variance showed statistically significant differences in domain score between 2023 and 2024. **No scores recorded for 2021 due to the COVID-19 pandemic; which also limited data collection in 2020. [b]Independent samples t-test showed statistically significant difference in dimension score between 2023 and 2024 (i.e., $p < .05$). [c]Independent samples t-test showed statistically significant difference in dimension score between 2023 and 2024 at the trend level (i.e., $p < .10$).

The changes in the distribution of ES, CO, and IS scores across the years are shown in Figures 1, 2, and 3, respectively. Some research appears to support thresholds for ES and CO above 5 and IS above 3 as necessary to evidence a relation between quality and children's outcomes (other research defines these as slightly higher, at 5.5 and 3.5) (Burchinal et al., 2009; Burchinal et al., 2014; Hatfield, et al., 2016). Emotional Support scores have, on average, increased with a higher number of classrooms reaching scores of 6 and 7. The number of classrooms scoring at 5+ in ES (Figure 1) has consistently increased in the past 3 years – from 86% of classrooms in 2022, to 88% in 2023, to 94% in 2024. For CLASS CO, an improvement in classrooms scoring above the 5 threshold is also observed across the past three years, with 67% of classrooms meeting this bar in 2022, compared to 77% in 2023, and 79% in 2024 (Figure 2). The distribution for CLASS IS has shifted to the left again this year, however, with 18% of classrooms scoring above 3 in 2024, compared to 23% of classrooms meeting this bar in 2023, and 37% in 2022 (Figure 3).

Figure 1. Distribution of CLASS Emotional Support scores for 2017 – 2024.

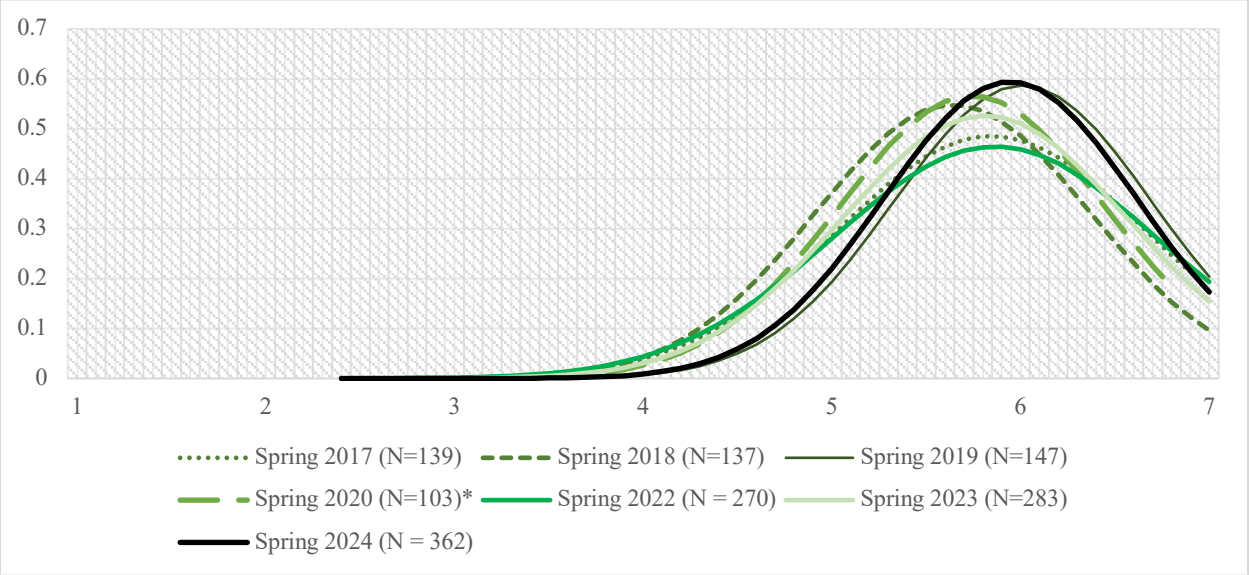



Figure 2. Distribution of CLASS Classroom Organization scores for 2017 – 2024.

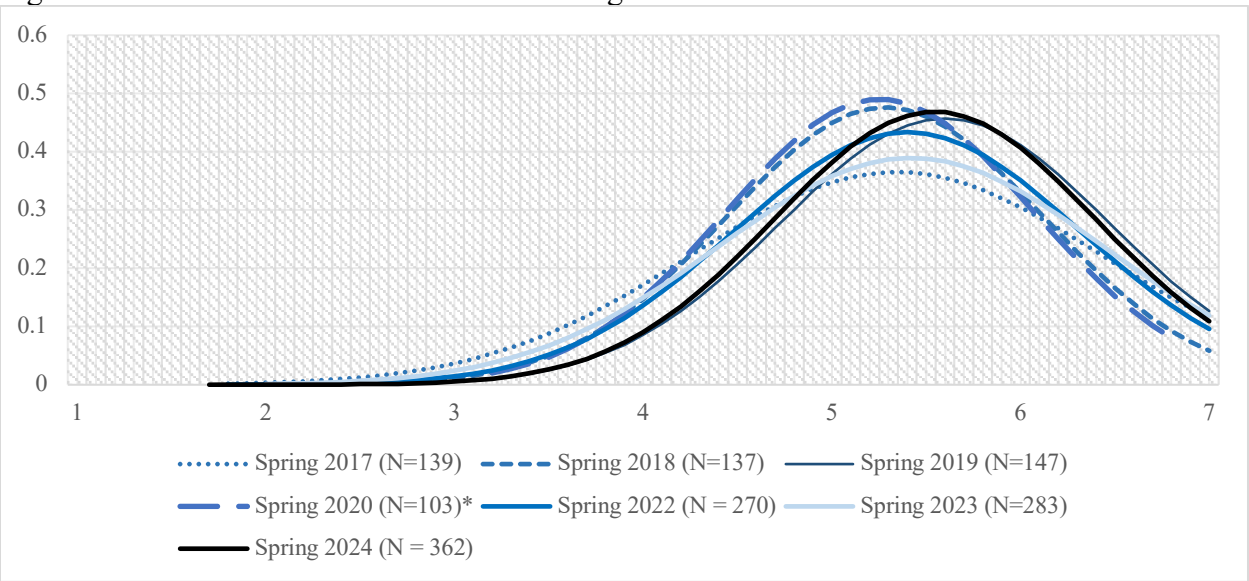
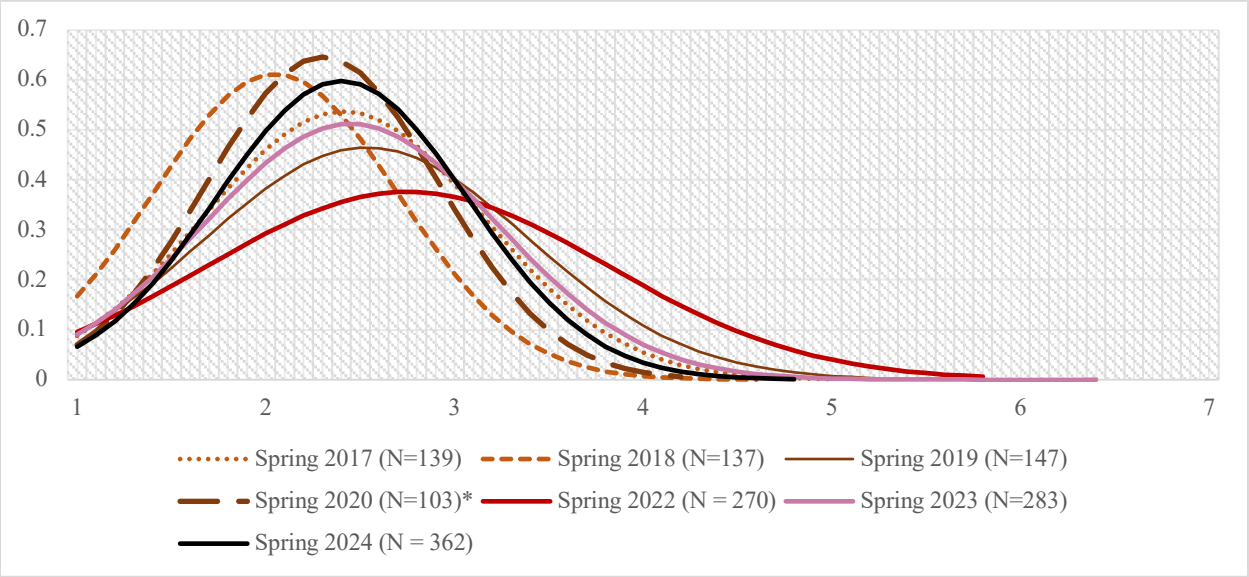


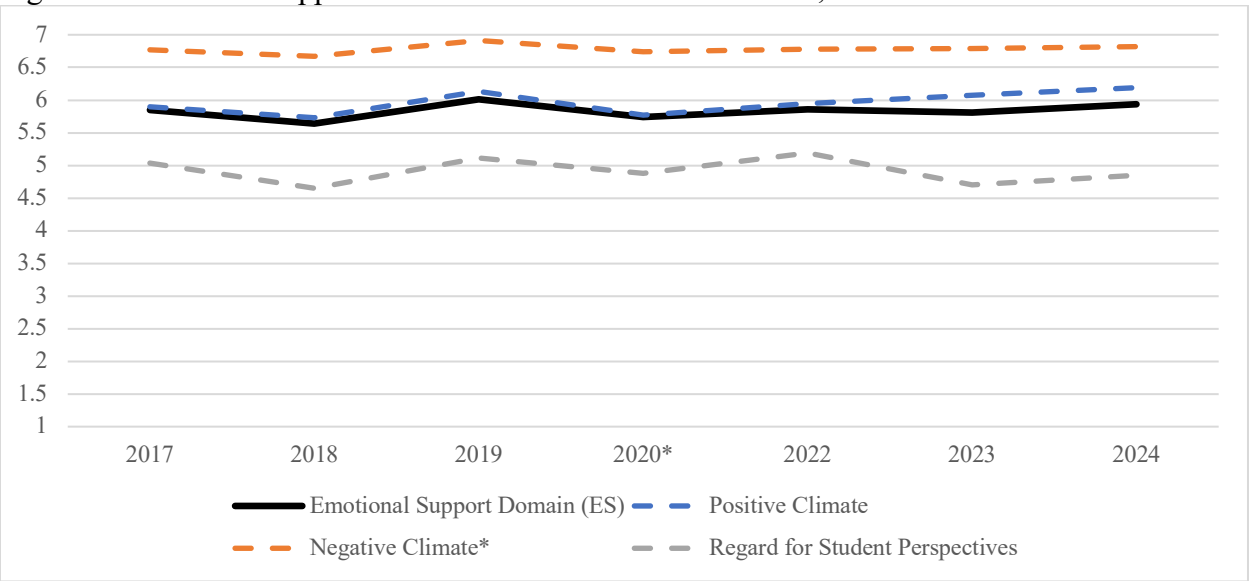
Figure 3. Distribution of CLASS Instructional Support scores for 2017 – 2024.




### *CLASS Pre-K – 3rd Domains*

As seen in Figure 4, Emotional Support scores have in general been increasing over time in PHLpreK classrooms. The overall mean score for ES is 5.94 (SD 0.67) – nearly the highest score recorded on this domain in PHLpreK classrooms (the mean in 2019 was just slightly higher, at 6.01). This difference was significantly higher than the ES scores in 2023. As also depicted in Figure 4, Regard for Student Perspectives remains the lowest scoring dimension in this domain, while Negative Climate (reverse scored for ease of interpretation) remains the highest.

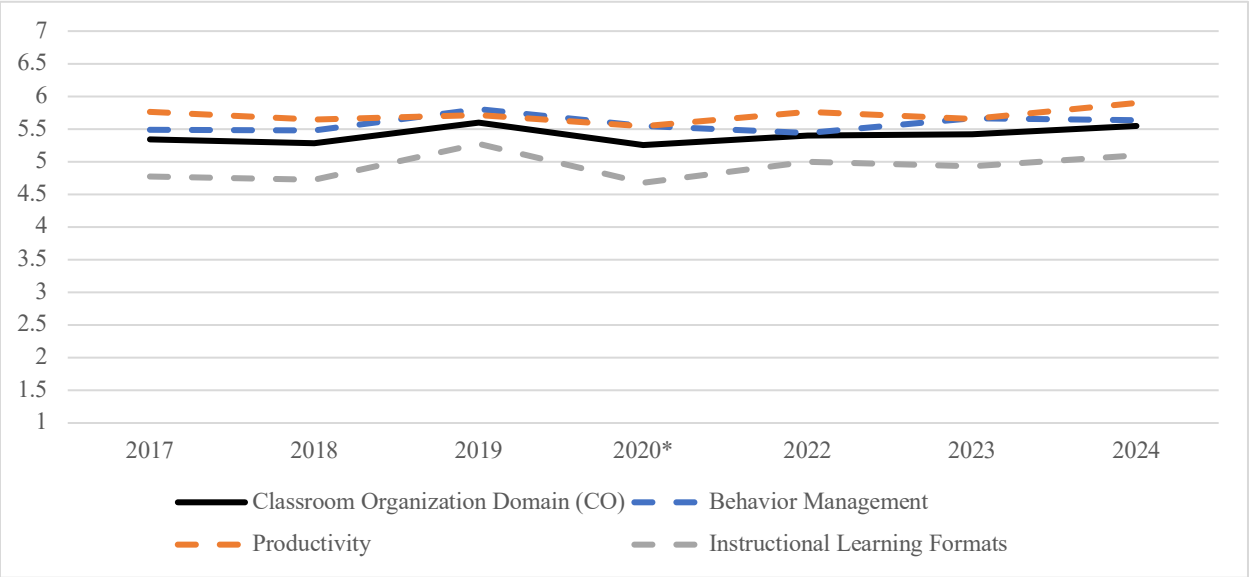
Figure 4. Emotional Support domain and associated dimensions, trends from 2017-2024.




As seen in Figure 5, scores on Classroom Organization have, in general, shown a small but positively increasing trend over time. The average mean score for the Classroom Organization Domain this year was 5.55 (SD 0.85). In terms of differences in dimension scores from 2023-2024, scores on Productivity were significantly higher in 2024 (5.90) as compared to 2023

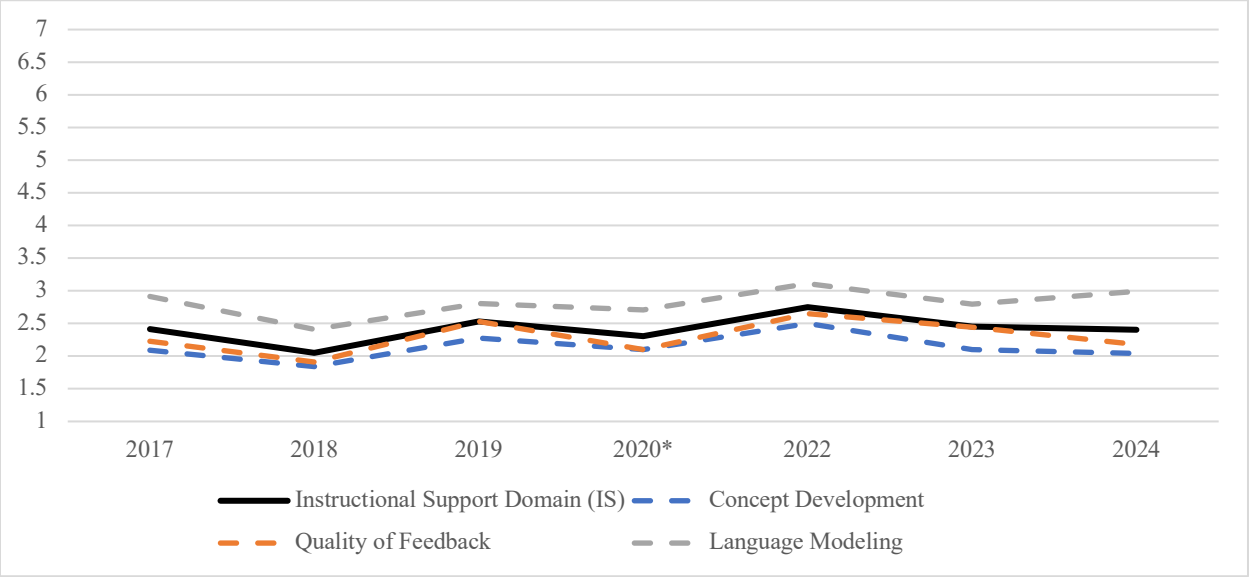
(5.66). As shown in Figure 5, Instructional Learning Formats scored lower than the other two dimensions (5.10), which is consistent with prior years and has remained the lowest-scoring dimension over time.

Figure 5. Classroom Organization domain and associated dimensions, trends from 2017-2024.




Instructional Support measures the ways in which teachers encourage analysis and reasoning, prompt children to think more deeply through high-quality feedback, and encourage and advance language development. This domain has consistently scored lower across preschool evaluations and systems, including in PHLpreK, as showing in Figure 6.

Figure 6. Instructional Support domain and associated dimensions, 2017-2024.



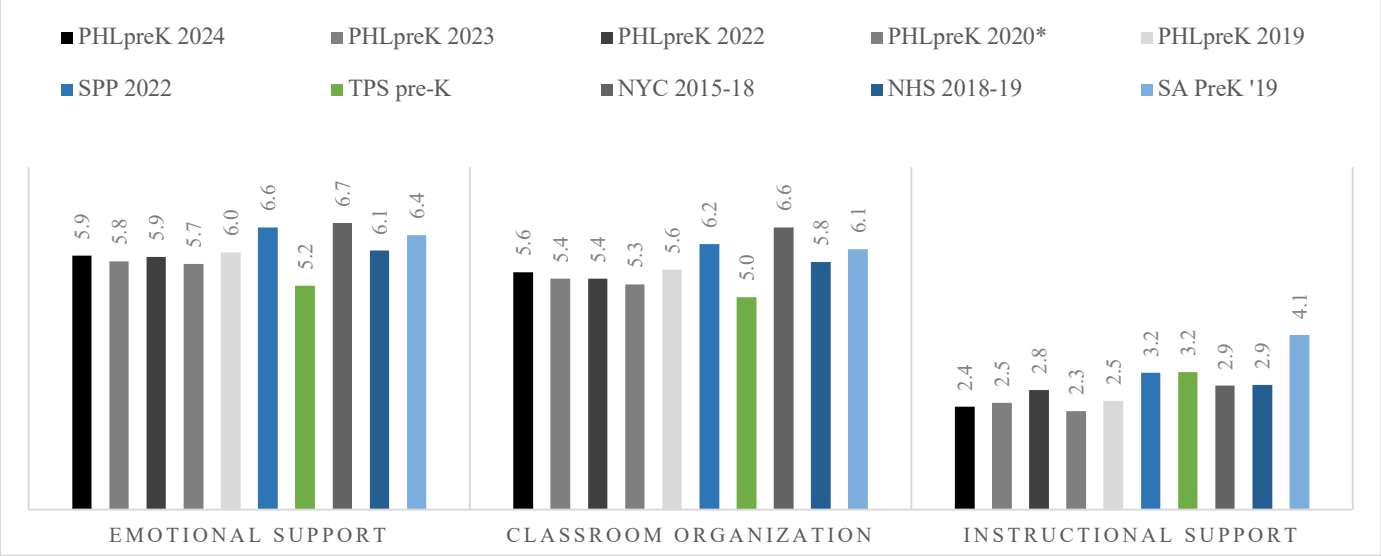

The average IS score is 2.40 (SD 0.67) with averages ranging from 1 to 4.73 on a 7-point scale. Consistent with prior years' evaluations, Concept Development and Quality of Feedback were

the lowest scoring dimensions, both in this domain and on the tool as a whole, with scores of 2.04 and 2.18, respectively. In considering changes from last year to this year within the Instructional Support domain, while scores on the domain looked very similar to last year (2.45 in 2023, 2.40 in 2024), there were changes in the dimensions that were significant. On the positive end of changes, the scores on Language Modeling are significantly higher in 2024 than they were in 2023 (2.99 compared to 2.80). Across the eight years of the PHLpreK evaluation, this is one of the highest recorded scores for Language Modeling; teachers should continue to be encouraged to use strategies that support advanced language with students. Conversely, scores on Quality of Feedback are significantly lower in 2024 as compared to 2023 (2.44 and 2.18, respectively).

### *CLASS Pre-K Comparison of Programs*

Figure 4 reports score patterns for PHLpreK in relation to those of other cities and states in which the CLASS has been used. The PHLpreK CLASS scores from 2019, 2020, 2022, 2023, and 2024 are reported by domain together with scores from various other programs in the U.S. This includes high-quality city programs such as the Seattle Preschool Program (SPP), the NYC Pre-K for All program and San Antonio's program.

Figure 4. Comparison of PHLpreK CLASS scores with other programs.




Note: SPP is Seattle's preschool program, reported in Pierson, et. al (2022); TPS is Tulsa's preschool program, reported in Phillips et. al (2009); NYC is reported in NYC Department of Education( 2018), and averages scores from 2015-2018; NHS is National Head Start Overview (2019); . SA is San Antonio's pre-K program reported in Decker-Woodrow et.al (2019).

### *CLASS Pre-K Domains for Selected Center Characteristics*

Table 2 shows CLASS domain scores for selected program-level characteristics. Classrooms with lower star levels (e.g., 1-3) score lower on all domains, although these differences were not significant, and most sites in the PHLpreK program were rated STAR 4. In terms of teacher credentials, classrooms with teachers with no degree/some college score just marginally higher on ES and CO, but this group is so small relative to the rest of the group (n=22), that these results should be interpreted with caution. In IS, teachers with a Master's Degree score the highest, but again, these differences between teachers based on credentials were quite small. In

relation to partner agency, classrooms in sites in collaboration with PHMC scored higher on all domains than did school district programs, but as with STAR level, the sample size of school district programs is much smaller (26 classrooms total) than PHMC affiliated programs, and so any differences should also be interpreted with caution. We also analyzed scores for child care centers and family/group child care homes (combining both types of home-based programs). These differences were also slight, with family child care providers (FCCs) scoring slightly higher on all domains, although these differences were not significant and the sample size of FCCs/group FCCs was quite small (34 total sites). Finally, there were no statistically significant differences noted between new and returning programs, although returning programs scored slightly higher on all domains, moreso in Emotional Support and Classroom Organization.

Table 2. CLASS domains scores by subgroups, N = 362.

| | | CLASS Mean Scores | | |
|---|---|---|---|---|
| | | Emotional Support | Classroom Organization | Instructional Support |
| **STAR Level** | No Star, Star 1-3 (n=56) | 5.88 | 5.41 | 2.28 |
| | 4 (n=306) | 5.96 | 5.57 | 2.42 |
| **Lead Teacher Credential** | No Degree/Some College (n=22) | 6.06 | 5.62 | 2.42 |
| | AA (n=136) | 6.01 | 5.60 | 2.38 |
| | BA (n=138) | 5.90 | 5.51 | 2.40 |
| | Master's Degree (n=58) | 5.90 | 5.54 | 2.44 |
| **New or Returning Site*** | New (n=54) | 5.83 | 5.35 | 2.39 |
| | Returning (n=308) | 5.97 | 5.58 | 2.40 |
| **PHLpreK Partner Agency** | SDP (n=26) | 5.89 | 5.41 | 2.30 |
| | PHMC (n=336) | 5.95 | 5.56 | 2.41 |
| **Program Type** | FCC Home/Group Home (n=34) | 6.03 | 5.73 | 2.58 |
| | Child Care Center (n=328) | 5.94 | 5.53 | 2.38 |
| **Lead Teacher Area of Study (Highest Degree)** | ECE (n=186) | 5.90 | 5.53 | 2.40 |
| | Elementary Education (n=19) | 5.83 | 5.49 | 2.29 |
| | Education, Other (n=23) | 5.90 | 5.46 | 2.44 |
| | Psychology (n=16) | 6.05 | 5.59 | 2.47 |
| | Other (n=31) | 5.98 | 5.60 | 2.65 |

*Two programs were listed as new because they did not have a PHLpreK contract last year; we coded them as returning since they had previously been a PHLpreK site. We did not have information on highest level of education attained for 8 teachers and on area of study for 87 teachers; those teachers are excluded from these subgroup analyses.

### *2023-2024 School Year Pilot: Summary of Results*

Finally, we observed in a subset of 100 classrooms this school year, in which teachers were provided with additional supports and coaching, and observed once in the fall and once in the spring. The summary in Table 3 shows the change in scores from the fall to the spring. Across all domains, scores were significantly higher in the spring. In some dimensions in which there was little room for growth (e.g., Negative Climate), scores did not change much. However, scores were significantly higher on multiple dimensions: Teacher Sensitivity, Regard for Student Perspectives, Behavior Management, Productivity, Concept Development, Quality of Feedback, and Language Modeling. This trend toward higher scores in the spring than the fall seems to indicate that the additional coaching supports are meaningful for classroom practice.

Table 3. Average scores for pilot classrooms, fall 2023 and spring 2024.

| CLASS Dimensions and Domains | 2023 Fall Pilot Sample Mean (Range) N=100 | 2024 Spring Pilot Sample Mean (Range) N=97 |
|---|---|---|
| **Emotional Support Domain (ES)** | **5.63** | **5.88*** |
| | **(2.50-6.95)** | **(2.40-6.95)** |
| **1. Positive Climate** | 5.86 | 6.08 |
| | (1.80-7.00) | (2.60-7.00) |
| **2. Negative Climate*** | 6.77 | 6.79 |
| | (2.40-7.00) | (4.00-7.00) |
| **3. Teacher Sensitivity** | 5.58 | 5.92* |
| | (1.40-7.00) | (1.40-7.00) |
| **4. Regard for Student Perspectives** | 4.33 | 4.74* |
| | (1.40-6.80) | (1.40-6.80) |
| **Classroom Organization Domain (CO)** | **5.16** | **5.48*** |
| | **(1.33-7.00)** | **(1.73-6.87)** |
| **5. Behavior Management** | 5.22 | 5.62* |
| | (1.60-7.00) | (1.40-7.00) |
| **6. Productivity** | 5.44 | 5.89* |
| | (1.20-7.00) | (1.80-7.00) |
| **7. Instructional Learning Formats** | 4.84 | 4.93 |
| | (1.20-7.00) | (2.00-7.00) |
| **Instructional Support Domain (IS)** | **2.07** | **2.45*** |
| | **(1.00-4.13)** | **(1.00-4.73)** |
| **8. Concept Development** | 1.81 | 2.07* |
| | (1.00-3.80) | (1.00-4.20) |
| **9. Quality of Feedback** | 1.84 | 2.18* |
| | (1.00-4.40) | (1.00-4.60) |
| **10. Language Modeling** | 2.56 | 3.09* |
| | (1.00-5.20) | (1.00-5.40) |

*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.

Finally, we also looked at pilot participants and measured change in scores from fall to spring as a function of whether the participants engaged in the supports provided by SPARK. A total of 19 classrooms opted not to engage in the provided supports (and one of these classrooms did not participate in the spring observation at all). While the scores for the classrooms that did not participate in the supports are higher on all domains and dimensions in the spring, the *growth* of scores from fall to spring is higher for pilot participants who did engage with the supports, as shown in Table 4. To compute these scores, we subtracted the fall 2023 score from the spring 2024 score. Thus, a positive number indicates a higher score in the spring, while a negative number indicates scores were lower in the spring.

Table 3. 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) |

## Discussion of Findings

This report summarizes the findings for the 2023-24 school year for Philadelphia's preschool program. The program has concluded its eighth year of operations and continues to grow since its inception through solidifying partnerships with community-based providers across the city. The purpose of the annual external classroom observations is to provide information that allows identifying strengths and weaknesses in the program through its expansion period in order to inform professional development and technical assistance efforts. This information also serves to inform continuous improvement strategies to support the program's maturation.

The 2023-24 scores in Emotional Support and Classroom Organization demonstrate that providers are building on previous year strengths in terms of developing a warm classroom climate, fostering positive relationships amongst children, and setting and maintaining high behavioral expectations. Notably, the statistically significant increase in Emotional Support scores from the 2023-24 school year as compared to the prior year demonstrates a positive trend and growth in the right direction for PHLprek. Supports for teachers on classroom quality should ensure this trend persists in future years, and emphasize efforts to address these two domains in low scoring classrooms. Additionally, the higher scores on the Classroom Organization domain as compared to year 7 are a positive change.

The Instructional Support domain is still low across classrooms, and scores on this domain were just slightly lower than last year and continue to evidence a need for strong supports. In summary, classrooms on average are nurturing and safe environments for children and are adequately structured and organized. Areas to strengthen include teachers' use of strategies to scaffold children's learning, incorporating conversational feedback loops that support children's understanding of concepts, increasing conversations to encourage children to use advanced language, questioning that supports the development of analytical thinking skills, linking concepts across activities so that children learn to apply their knowledge to the real world, providing opportunities to engage in problem-solving activities, and planning and production processes that incorporate and build upon children and their initiatives.

However, a focus on increasing classroom quality on Instructional Support specifically within PHLpreK will require particular focus around strengthening instructional supports (concept development, quality of feedback, language modeling, metacognition), and providing teachers with targeted coaching and supports for doing so across the whole system. The scores on Instructional Support indicate an absence of consistent, higher-level instructional practices that support children to develop more complex language skills, use feedback to expand skills,

and help them solve problems and reason. Teachers who score high on this domain frequently engage in practices such as asking open ended questions that extend conversation, expand on children's ideas by adding novel words or additions that make phrases more complex, and support children's play in a way that helps them get more out of the experience. For PHLpreK to see meaningful changes in the Instructional Support domain, a focus on supporting teachers to engage in these practices more consistently would benefit children's learning.

Finally, results from the pilot demonstrate that targeting a subgroup of teachers with supports can positively impact classroom practice, but that engaging with the supports is an important piece of improving this practice over time. PHLpreK should continue to find ways to support teachers to deliver classroom instruction through dynamic coaching and instructional supports. In addition, since about a fifth of the classrooms exhibit higher levels of instructional support, it is possible to generate pathways for cross-learning across the system, to further strengthen supports across the system.

## Acknowledgments

We are grateful for the partnership with the City of Philadelphia and PHMC, whose funding and support made this work possible. This partnership is instrumental in supporting the research and facilitating collaboration with program providers throughout the City across research components. Finally, we are thankful to SPARK and for all the programs and teachers that opened their doors for the work to be conducted.

## References


Barnett, W. S. (2008). Preschool education and its lasting effects: Research and policy implications. Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit. Retrieved September 20, 2017, from http://nepc.colorado.edu/files/PB-Barnett-EARLY-ED_FINAL.pdf

Barnett, W. S., & Frede, E. C. (2017). Long-term effects of a system of high-quality universal preschool education in the United States. *Childcare, early education and social inequality: An international perspective*, 152-172.

Barnett, W. S., & Jung, K. (2021). Effects of New Jersey's Abbott preschool program on children's achievement, grade retention, and special education through tenth grade. *Early Childhood Research Quarterly, 56,* 248–259.

Barnett, W. S., & Nores, M. (2015). Investment and productivity arguments for ECCE. Investing against Evidence, 73.

Barnett, W. S., Jung, K., Friedman-Krauss, A., Frede, E. C., Nores, M., Hustedt, J.T., Howes, C. & Daniel-Echols, M. (2018). State prekindergarten effects on early learning at kindergarten entry: An analysis of eight state programs. *AERA Open*, 4(2).

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A., (2009). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25,* 166-176.

Burchinal, M., Vernon-Feagans, L., Vitiello, V., & Greenberg, M. (2014). Thresholds in the Association between child care quality and child outcomes in rural preschool children. *Early Childhood Research Quarterly*, 29, 41–51. doi:10.1016/j.ecresq.2013.09.004

Ceci, S. J., & Papierno, P. B. (2005). The Rhetoric and Reality of Gap Closing: When the "Have-Nots" Gain but the "Haves" Gain Even More. American Psychologist, 60(2), 149-160.

*Classroom Assessment Scoring System 2nd Edition* (2022). Teachstone.

Decker-Woodrow, L., Diaz, E., & Lamey, G. (2019). *Pre-K 4 SA Evaluation Report.* Retrieved from: https://prek4sa.com/wp-content/uploads/2020/10/PreK4SA_Year6_EvaluationReport.pdf.

Duncan, G, & Murnane, R. (2011). Introduction: The American dream, then and now. In eds. Greg J. Duncan, Richard J. Murnane, (Eds.), *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances* (pp. 3–26). Russell Sage Foundation and Spencer Foundation.

Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2023). The long-term effects of universal preschool in Boston. *The Quarterly Journal of Economics*, *138*(1), 363-411.

Harmeyer, E., Nores, M., & Espinosa, C. (2023). *Evaluation of the Philadelphia Prek Program. Year 7 Report.* National Institute for Early Education. Retrieved from: https://nieer.org/sites/default/files/NIEER%20PHL%20Classroom%20Qual%20Report%202023_FINAL_1.pdf

Hamre, B., Hatfield B. E., Pianta R. C., & Jamil F. (2014). Evidence for general and domain-specific elements of teacher-child interactions: Associations with preschool children's development. *Child Development, 85,* 1257–1274.

Hatfield, B. E., Burchinal, M. R., Pianta, R. C., & Sideris, J. (2016). Thresholds in the association between quality of teacher–child interactions and preschool children's school readiness skills. *Early Childhood Research Quarterly, 36*, 561–571. doi:10.1016/j.ecresq.2015.09.005

Johnson, A. D., Partika, A., Martin, A., Horm, D., Phillips, D. A., & Tulsa SEED Study Team. (2023). A deeper dive, a wider pool: Preschool benefits sustain to first grade on a broader set of outcomes. *Child Development, 94,* 1298–1318. https://doi.org/10.1111/cdev.13928

New York City Department of Education (2021). *Pre-K program assessments Classroom Assessment Scoring System (CLASS) and Early Childhood Environmental Rating Scale – Revised (ECERS-R) release*. Retrieved from: https://infohub.nyced.org/docs/default-source/default-document-library/2019-20--program-assessment-results-summary.pdf.

Nores, M., Barnett, W.S., & Acevedo, M. (2018) Evaluation of the Philadelphia Prek Program. Year 2 Report. New Brunswick, NJ: National Institute for Early Education. Submitted report. Retrieved from: https://nieer.org/sites/default/files/NIEER%20PHL%20Classroom%20Qual%20Report%202018_FINAL.pdf

Nores, M., Barnett, W.S., Jung, K., Joseph, G. & Bachman, L. (2019). *Year 4 report: Seattle Preschool Program evaluation.* National Institute for Early Education Research & Cultivate Learning. Retrieved from: http://nieer.org/research-report/seattle-pre-k-program-evaluation

Nores, M., Barnett, W.S., Li, Z., Acevedo, M., & C. Whitman (2019). *Evaluation of the Philadelphia PreK Program. Year 3 Report.* National Institute for Early Education Research. Retrieved from: https://nieer.org/sites/default/files/NIEER%20PHL%20Classroom%20Qual%20Report%202019_FINAL.pdf

Nores, M., Francis, J. & Barnett, W.S. (2017). *Evaluation of the Philadelphia Pre-K program. Classroom quality report.* National Institute for Early Education Research.

Nores, M., & Harmeyer, E. (2023). Quality in New Jersey's Abbott Preschool Program: A closer look across the years. New Brunswick, NJ: National Institute for Early Education Research. Retrieved from: https://nieer.org/sites/default/files/2023-09/NIEER_Research_Report_AbbottovertheYears_Aug9_2023.pdf

Nores, M., Harmeyer, E., Li, Z., & Acevedo, M.(2021). *Evaluation of the Philadelphia PreK Program. Year 5 Report.* National Institute for Early Education Research. Retrieved from: https://nieer.org/sites/default/files/NIEER%20PHL%20Classroom%20Qual%20Report%202021_FINAL.pdf

Nores, M., Harmeyer, E., Li, Z., & Espinosa, C. (2022). *Evaluation of the Philadelphia PreK Program. Year 6 Report.* National Institute for Early Education Research. Retrieved from:

https://nieer.org/sites/default/files/NIEER%20PHL%20Classroom%20Qual%20Report%202022_FINAL.pdf

Nores, M., Li, Z., & Acevedo, M. (2020). *Evaluation of the Philadelphia PreK Program. Year 4 Report*. National Institute for Early Education Research. Retrieved from: https://nieer.org/sites/default/files/NIEER%20PHL%20Classroom%20Qual%20Report%202020_FINAL.pdf

Office of Head Start, U.S. Department of Health and Human Services, Administration for Children and Families. (2019). *A national overview of grantee CLASS scores in 2019.* Retrieved from: https://eclkc.ohs.acf.hhs.gov/data-ongoing-monitoring/article/national-overview-grantee-class-scores-2019

Philadelphia Commission on Universal Pre-kindergarten. (2016). *Final Recommendations Report*. Retrieved from: http://www.phila.gov/universalprek/Documents/Recommendations%20Report.pdf

Pianta, R. C. & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38 (2009), pp. 109-119, 10.3102/0013189X09332374

Pierson, A., Rooney, K., Serrano, D., & Gandhi, E. (2022, August). *Seattle Preschool Program Evaluation: Understanding trends in child and educator diversity, child outcomes, and classroom quality.* Retreieved from: https://www.seattle.gov/documents/Departments/DEEL/Results/Reports%20and%20Data/SPP%20Reports/2022-SPP-Evaluation-Technical-Report.pdf

# Appendix A. Measures

### *Classroom Assessment Scoring System 2.0 (CLASS; Classroom Assessment Scoring System 2nd Edition, 2022)*

The Classroom Assessment Scoring System (CLASS) 2.0 is an observational system that assesses classroom practices by measuring the interactions between students and teachers. CLASS measures interactions along ten distinct dimensions, which are grouped into three overarching domains. The Emotional Support (ES) domain is measured by four dimensions: Positive Climate, Negative Climate, Teacher Sensitivity, and Regard for Student Perspectives. The Classroom Organization (CO) domain is measured by three dimensions: Productivity, Behavior Management, and Instructional Learning Formats. The Instructional Support (IS) domain is measured by three dimensions: Concept Development, Quality of Feedback, and Language Modeling. Observations consist of five 20-minute cycles, with 10-minute coding periods between each cycle. Scores (codes) are assigned during various classroom activities and then averaged across all cycles for overall scores in three domains. Each dimension is scored on a 7-point Likert-type scale, for which a score of 1 or 2 indicates low quality, and a score of 6 or 7 indicates high quality.

Table A.1. CLASS Domains and Dimension Descriptions.

| Domain | Dimension | Description |
|---|---|---|
| **Emotional Support** | Positive Climate | Reflects the emotional connection between teachers and children and among children, and the warmth, respect, and enjoyment communicated by verbal and nonverbal interactions. |
| | Negative Climate | Reflects the overall level of expressed negativity in the classroom. The frequency, quality, and intensity of teacher and peer negativity are key to this dimension |
| | Teacher Sensitivity | Encompasses the teacher's awareness of and responsiveness to students' academic and emotional needs. |
| | Regard for Student Perspectives | Captures the degree to which the classroom activities and teacher's interactions with students place an emphasis on students' interests, motivations, and points of view and encourage student responsibility and autonomy. |
| **Classroom Organization** | Behavior Management | Encompasses the teacher's ability to provide clear behavior expectations and use effective methods to prevent and redirect misbehavior. |
| | Productivity | Considers how well the teacher manages instructional time and routines and provides activities for students so that they have the opportunity to be involved in learning activities. |
| | Instructional Learning Formats | Focuses on the ways in which teachers maximize students' interest, engagement, and abilities to learn from lessons and activities. |
| **Instructional Support** | Concept Development | Measures the teacher's use of instructional discussions and activities to promote students' higher-order thinking skills and cognition and the teacher's focus on understanding rather than on rote instruction. |
| | Quality of Feedback | Assesses the degree to which the teacher provides feedback that expands learning and understanding and encourages continued participation. |
| | Language Modeling | Captures the effectiveness and amount of teacher's use of language-stimulation and language-facilitation techniques. |