NIEER
NATIONAL INSTITUTE FOR EARLY EDUCATION RESEARCH

EVALUATION OF THE PHILADELPHIA PREK PROGRAM

Year 6 Report

August 2022

Milagros Nores, PhD, Erin Harmeyer, PhD, and Carmen Espinosa, MEd. The National Institute for Early Education Research

About the Authors

Milagros Nores, Ph.D. is Co-Director of Research and Associate Research Professor and Associate Research Professor 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.

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 West Virginia, Indiana, New Jersey, and Pennsylvania.

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

Correspondence regarding this report should be addressed to Milagros Nores at the National Institute for Early Education Research. Email: mnores@nieer.org.

Suggested citation: Nores, M., Harmeyer, E., & Espinosa, C. (2022). Evaluation of the Philadelphia PreK Program. Year 6 Report. New Brunswick, NJ: National Institute for Early Education Research.

Table of Contents

Introduction

Philadelphia's Preschool Program (PHLpreK) has recently concluded its sixth year of programming. The program was initiated after a May 2015 vote where Philadelphia 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 children in the city, ages 3 and 4. For the past six years, The National Institute for Early Education Research (NIEER) at Rutgers University has been conducting a multi-year, multi-site evaluation assessing program components, program quality, and children's learning and development.

In our past five years of reports, we have demonstrated the importance of high-quality preschool education, and highlighted it as a mechanism for preparing children for kindergarten and beyond and reducing persistent gaps in achievement between lower-income and higher-income children (e.g., Barnett et al., 2018; Nores et al., 2017; Nores et al., 2018 & 2019; Yoshikawa et al., 2013). This report summarizes classroom quality for students in PHLpreK classrooms during the 2021-22 school year, a year in which classroom instruction and practices were still largely impacted by the COVID-19 pandemic, including the Delta variant that was prevalent in the summer of 2021, and the Omicron variant that was prevalent in the winter of 2021-22. These variants led to surges in case numbers nationwide and in Philadelphia that had impacts on classroom practice including numerous quarantines for children ineligible for vaccinations through the school year, and social distancing policies in the classroom.

Despite the continued disruptions from the pandemic, findings suggest PHLpreK classrooms are averaging high to moderate levels of quality in the Emotional Support and Classroom Organization domains, and these scores have increased relative to the scores that were collected in our limited sample of data collection in 2020. Scores continue to be lower for the Instructional Support domain; however, they have increased significantly relative to scores collected in 2020 prior to the onset of the pandemic. The last time we were able to collect data on a larger sample was in 2019; while scores for the Instructional Support domain this year were higher than in 2019, scores on the Emotional Support and Classroom Organization domain were slightly lower than what was recorded in 2019 (although both domains have mean scores that are still in the good quality range). We explored quality separately for several subgroups of interest, including STAR level, lead teacher credentials, lead teacher area of study, PHLpreK partner agency, curriculum, and whether or not sites were new in the PHLpreK program. Small differences were found between subgroups and are reported. The report includes some recommendations on areas to strengthen quality, based on the data collected.

Study 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. This report focuses on the sixth year of the study: a year in which programs continued to be impacted by the COVID-19 pandemic. Children under the age of 5 were still ineligible for vaccinations throughout the duration of this data collection year. The focus of the evaluation this year was therefore on classroom quality observations. We sought to answer the following questions in this project:

1. What is the observed quality of children's classroom experiences and how does it compare relative to the prior years of the PHLpreK program?
2. How does the observed quality of children's classroom experiences vary by STAR level, lead teacher credentials, lead teacher area of study, curriculum usage and whether or not classrooms are in new sites?

The PHLpreK evaluation is an effort to measure program development in its early years in terms of quality and children's learning and development. 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, we collected data from all classrooms enrolled in the PHLpreK program. Procedures and measures are described in detail below. Classroom observations were conducted to assess teacher-child interactions and took place between February and June 2022. As in previous years, quality was assessed using well-known observation protocols during one visit of about two and a half to three hours.

1. Sample

In the 2021-22 school year, NIEER assessed classroom quality in all classrooms across all PHLpreK sites, which totaled 270 classrooms (24 which were home-based providers) across 153 sites. Classroom quality was assessed with the CLASS Pre-K in all 270 classrooms in the program.

2. Measures and Procedures

Classroom quality was captured using the *Classroom Assessment Scoring System Pre-K* (*CLASS Pre-K*; Pianta, La Paro, & Hamre, 2008). The CLASS measures teacher-child interactions and classroom processes. More detail on this measure is provided in Appendix A.

Observers were trained to reliability before conducting observations of classroom quality. CLASS observers were trained by a CLASS Affiliate Trainer from NIEER or through an online training provided by Teachstone®, completed the online reliability required by Teachstone® and met their requirement (80%) for observer certification. Observers were also trained in practices and procedures for appropriate conduct in PHLpreK classrooms (including following COVID-19 protocols) and were required to complete background checks, as well as training in human subjects research (human subject protections, ethical issues, etc.).

Results

We report results first for the whole sample, and then provide analyses of results for different subgroups. Recommendations based on the results concludes this year's report.

1. Classroom Observations

CLASS Pre-K Results

Average CLASS scores for PHLpreK classrooms for all domains and dimensions are reported in Table 1. Patterns are consistent with the field and previous years, with Instructional Support scoring lower than other domains. Emotional Support (ES) scores show improvements between 2020 and 2022 (from 5.74 to 5.86). This is also the case for Classroom Organization (CO) (increasing from 5.26 to 5.40) and for Instructional Support (IS) (increasing from 2.30 to 2.75). The increase in scores from 2020 to 2022 is statistically significant for CLASS IS. Results for each domain are discussed further below. Observed increases between 2020 to 2022 in ES were of 0.17 SD (standard deviations),[1] in CO these were of 0.18 SD and in IS these were of 0.72 SD. For IS particularly, this is a significant increase in quality as compared to 2020.

[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.

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=103	2022 Mean (Range) N=270
Emotional Support (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)**
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)
2. Negative Climate**	6.77 (5.00-7.00)	6.67 (4.00-7.00)	6.91 (5.40-7.00)	6.74 (4.20-7.00)	6.78 (3.80-7.00)
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)
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 (2.80-6.80)	5.19 (2.00-7.00)
Classroom Organization (CO)	**5.34 (1.87-6.93)**	**5.28 (2.80-6.93)**	**5.60 (2.40-7.00)**	**5.26 (3.20-6.80)**	**5.40[b] (1.87-6.93)**
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)
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)
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)
Instructional Support (IS)	**2.41 (1.00-5.00)**	**2.05 (1.00-4.60)**	**2.54 (1.00-5.33)**	**2.30 (1.33-4.13)**	**2.75[a] (1.00-5.80)**
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)
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)
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)

*The sample of classrooms collected in 2020 was limited due to the COVID-19 pandemic disrupting data collection. **The Negative Climate dimension is reverse scored so that a high score represents "good." [a] Statistically significant difference between 2020 and 2022, and 2019 and 2022. [b] Statistically significant difference between 2019 and 2022.

The changes in the distribution of ES, CO, and IS scores across the years are observable 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 support 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). Overall, the distribution of scores for CLASS ES and CLASS CO has not changed significantly over the last few years of data collection. The number of classrooms reporting scores above 5 in ES was 86% this year (compared to 93% in 2019, and 87% in 2020).

For CLASS CO, a slight improvement from 2020 is observed on average scores and an increase in the percentage of classrooms scoring in the excellent range of 6.5 – 7 (3% in 2020, and 12% in 2022). In addition, a total of 67% of classrooms have a score above 5 (up slightly from 65% in 2020, but lower than the 78% found in 2019). However, the score is significantly

lower than it was in 2019 (5.40 in 2022, down from 5.60 in 2019) which is the last year in which all classrooms in the PHLpreK program were observed.

Finally, as shown in Figure 3, the distribution for IS scores shifted to the right in 2022, with a greater proportion of teachers scoring above the threshold of 3 in this domain. While 26% were above the threshold of 3 in 2019 and 14% in 2020, a total of 37% of classrooms have a score above 3 this year. The scores on IS are statistically significantly higher in 2022 than they were in 2019 and than they were in 2020, with scores of 2.54 in 2019, 2.30 in 2020, and 2.75 in 2022.

Figure 1. Distribution of CLASS Emotional Support scores for 2017, 2018, 2019, 2020 & 2022.

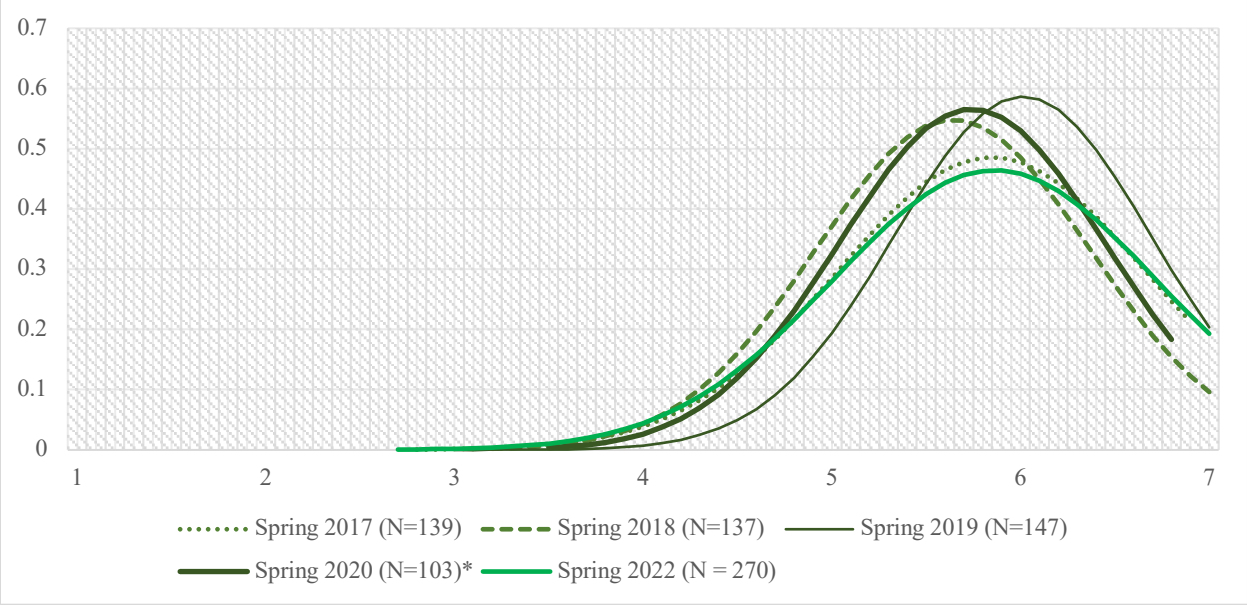


Figure 2. Distribution of CLASS Classroom Organization scores for 2017, 2018, 2019, 2020 & 2022.

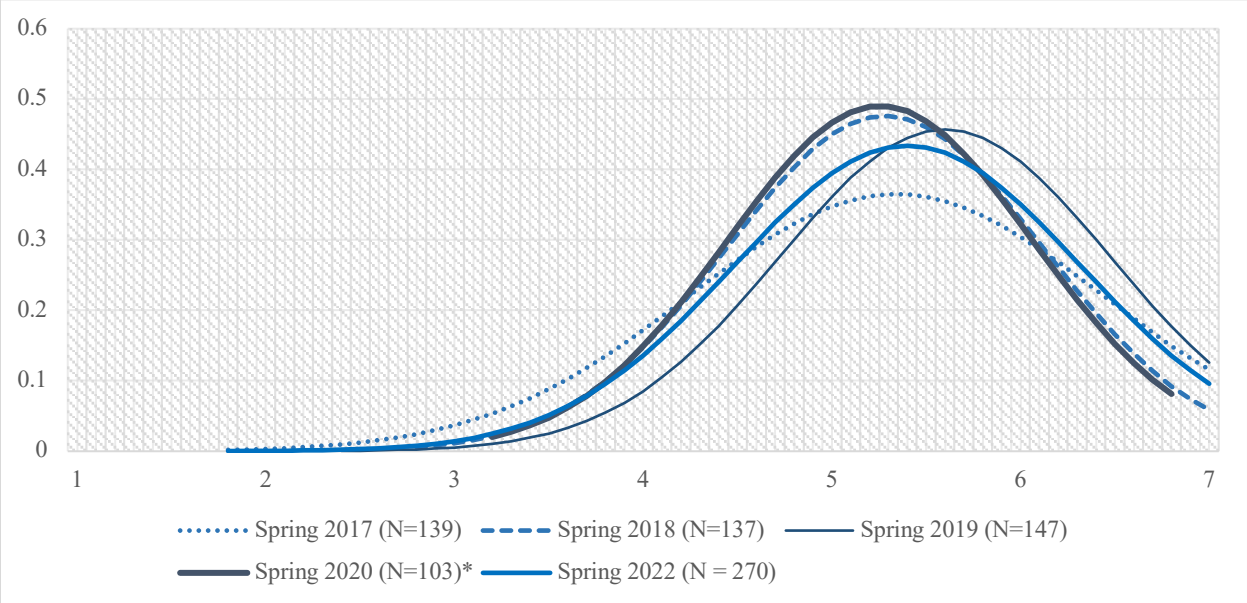
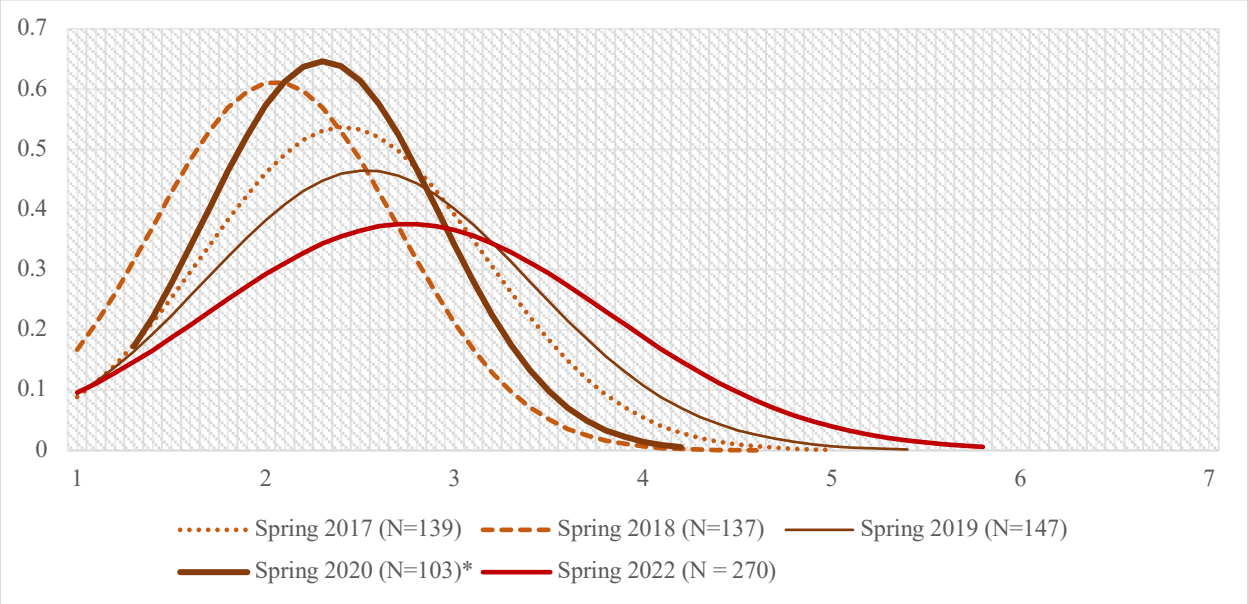


Figure 3. Distribution of CLASS Instructional Support scores for 2017, 2018, 2019, 2020 & 2022.



These distributions can also be seen in the following figures, which show the change in time in the number of classrooms scoring within half-point score ranges within the three CLASS domains. For example, for the domain Emotional Support, as show in Figure 4, the percentage of classrooms scoring in the excellent (6.5 – 7.0) range has changed over time, with 20% of classrooms scoring in this range in 2017, and 27% of classrooms scoring in this range in 2022. In Figure 5, the number of classrooms scoring in different score ranges in the Classroom Organization domain is depicted. As shown, the percentage of classrooms scoring in the excellent (6.5-7.0) range also increased, from 8% in 2017, to 12% in 2022. Figure 6 shows the number of classrooms in different score ranges from 2017 to 2022 in the Instructional Support domain. While 19% of classrooms scored above 3.0 in 2017, that number had increased to 37% in 2022. This domain is particularly important in the preschool years as children are getting ready for kindergarten; while there is still room for growth in scores, these changes over time in this domain are encouraging.

Figure 4. Distribution of CLASS scores for the domain Emotional Support in 2017 (left; n=139) and 2022 (right; n=270).

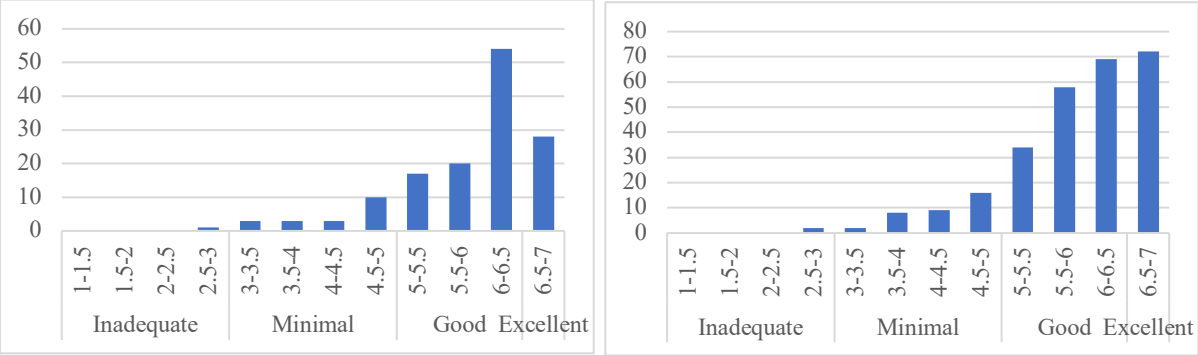


Figure 5. Distribution of CLASS scores for the domain Classroom Organization in 2017 (left; n=139) and 2022 (right; n=270).

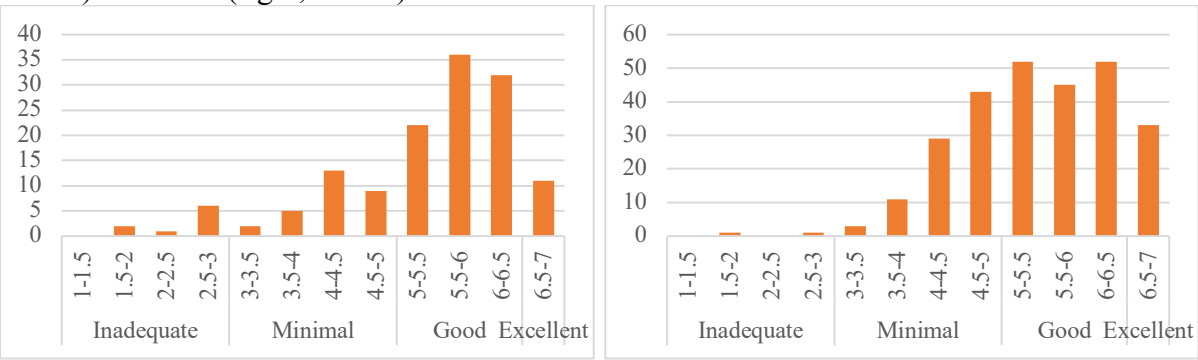


Figure 6. Distribution of CLASS scores for the domain Instructional Support in 2017 (left; n=139) and 2022 (right; n=270).

CLASS Pre-K Domains

The Emotional Support (ES) domain focuses on the types of teacher behaviors that support the development of relationships between teachers and students, along with behaviors that allow children to enjoy the learning process while feeling comfortable and supported within the classroom. The overall mean score for ES this year is 5.86 (SD 0.86), which is in the range of high quality.

The highest score in this domain has consistently been the dimension Negative Climate, with a score of 6.78 this year (compared to 6.74 in 2020). This indicates that classroom teachers very rarely engage in negative interactions such as yelling or displaying disrespect toward students. The lowest scoring dimension is Regard for Student Perspectives (5.19), although this score represents an increase compared to all other years of evaluation of the PHLpreK program. In order for classrooms to score high on this dimension, teachers demonstrate flexibility in following the lead of children, provide choice in activities, and encourage student responsibility, teaching behaviors that support the development of supportive relationships between teachers and children, and that help children enjoy the learning process and feel comfortable in the classroom.

The Classroom Organization (CO) domain focuses on teaching behaviors that include effectively managing behavioral expectations, instructional time and routines, and effectively facilitating instruction in a way that maximizes children's engagement and interests. The average mean score for the Classroom Organization domain is 5.40 (SD 0.92). Scores above 5.00 are in the good range and indicate that teachers are effective at preventing and redirecting misbehavior, and that the students in the classroom tend to be compliant and behave appropriately towards the teacher and one another. Teachers who score high in this domain are consistently organized with plans in place, provide clear instructions, and take care of managerial tasks efficiently. The lowest dimension within this domain was Instructional Learning Formats (5.00), which is consistent with past PHLpreK evaluations. Teachers who score high in this dimension ensure children are oriented towards learning objectives and use effective questioning with children that expands their interest in learning activities, along with using interesting and creative materials. Increasing scores in this domain involves teachers ensuring children are allowed to use different and interesting materials, including hands-on activities, that they orient children to learning activities, and that they use effective questioning methods.

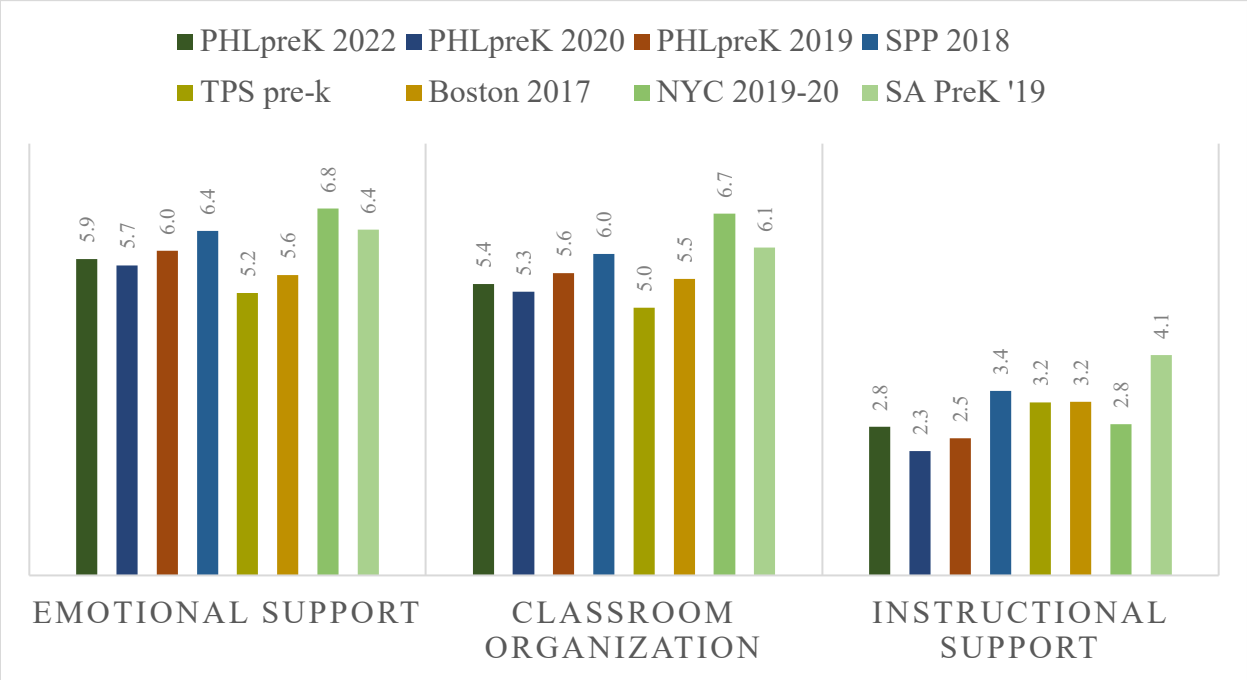
The Instructional Support domain measures how teachers facilitate and encourage higher-order thinking skills, expand understanding and learning and promote language development through their interactions. Scores on this domain are consistently lower across large-scale evaluations of preschool programs. However, this domain is critical to children's learning and development during the pre-K years. The average IS score is 2.75 (1.06), with scores ranging from 1 to 5.80. The dimensions Concept Development and Quality of Feedback both had lower scores within the domain (2.50 and 2.65, respectively), although these scores have consistently increased over time. Concept Development measures teacher interactions centered on facilitating children's higher-order thinking skills and cognition and a teacher's focus on understanding versus rote memorization. Higher scores in this domain would have teachers strongly facilitating children's thought processes through strategies including experimentation, problem-solving, brainstorming, and encouraging children's creative processes. Quality of Feedback is focused on teaching behaviors such as scaffolding learning, and the use of follow-up questions with specific information that encourage children to keep thinking. Higher scores in this dimension are

obtained when teachers use back-and-forth exchanges and metacognitive approaches to expand on children's thinking processes. Scores in Language Modeling are 3.11 (SD 1.11), reaching above the threshold of 3.0 for the first time since the PHLpreK program started. Language modeling is focused on interactions teachers use that model complex language usage and encourage children to talk; higher scores in this domain would occur with teachers engaging in back-and-forth exchanges with children, intentionally modeling advanced language, and encouraging children to talk.

CLASS Pre-K Comparison of Programs

Figure 4 reports score patterns for the PHLpreK program in relation to those of other city programs. The PHLpreK CLASS scores from 2019, 2020, and 2022 are reported by domain together with scores from various other programs in the U.S. This includes high-quality city programs in Boston, New York, Seattle, Tulsa, and San Antonio. As shown in the chart, PHLpreK scores on ES are roughly mid-range (i.e., higher than Tulsa, and Boston; lower than San Antonio, Seattle and NYC). Scores on CO are higher than Tulsa, but lower than all other programs. Finally, scores on IS are lower than all other programs.

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



CLASS Pre-K Domains for selected center characteristics

Table 2 shows CLASS domain scores for selected program-level characteristics. Classrooms with a 4 STAR level scored higher on ES and CO than those rated 3 STAR. Concerning partner agency, classrooms in sites in collaboration with PHMC showed slightly higher scores on ES and CO than SDP programs (both programs scored about the same on the IS domain) but these differences were small and not statistically significant. Those using the Creative Curriculum had lower scores than those using HighScope on ES and CO, but higher scores on IS (although the

vast majority of programs reported they use Creative Curriculum). These differences were not statistically significant.

Classrooms in sites that started in PHLpreK only this last year scored lower on all three domains than sites that are not new to the PHLpreK program. In the area of lead teacher degree, classrooms with teachers with a Master's degree scored higher on all three domains than those with teachers with any other educational credentials, and the difference in teachers with a Master's degree and those with Some College/Missing was statistically significant. Across the board, teachers with a Bachelor's degree scored next highest on all three domains (although not significantly higher than those with other credentials), and the lowest scores were recorded for teachers with Some College/Missing information. Finally, in the area of lead teacher area of study, teachers with a background in Elementary Education scored higher than teachers with other backgrounds in the Instructional Support domain, while those who studied Psychology scored the highest in the Classroom Organization domain. Teachers with a background in Education (other, general, secondary or special education backgrounds were included in this category) scored the highest on the Emotional Support domain. These differences were mostly small, and none were significant. The highest scores in Instructional Support belonged to those with a background in education (Early Childhood, Education, or Elementary Education).

Table 2. CLASS domains scores by subgroups, N = 270.

		CLASS Mean Scores		
		Emotional Support	Classroom Organization	Instructional Support
STAR Level	1-2 (n=4)	6.09	5.83	2.68
	3 (n=44)	5.82	5.27	2.78
	4 (n=222)	5.87	5.42	2.75
Lead Teacher Degree	AA (n=79)	5.80	5.35	2.73
	BA (n=94)	5.91	5.49	2.86
	Master's (n=47)	5.97	5.52	2.96[a]
	Some College, Missing (n=50)	5.78	5.21	2.40[a]
Lead Teacher Area of Study	Early Childhood Education (n=121)	5.78	5.34	2.73
	Education (Other) (n=26)	6.02	5.56	2.98
	Elementary Education (n=20)	5.96	5.54	3.14
	Psychology (n=12)	5.95	5.63	2.52
	Information Not Available (n=34)	5.74	5.22	2.42
	Other (n=57)	5.98	5.46	2.81
PHLpreK Partner Agency*	PHMC (n=255)	5.87	5.41	2.75
	SDP (n=15)	5.74	5.23	2.77
Curriculum	Teaching Strategies GOLD (n=258)	5.86	5.39	2.76
	HighScope+ (n=11)	5.92	5.48	2.51
	Montessori Education (n=1)	-	-	-
New Site	No (n=226)	5.87	5.42	2.78
	Yes (n=43)	5.83	5.31	2.57

Note: Two-tailed tests of differences in means were run between all groupings for each domain. Statistically significant differences were found in the category of Teacher Degree between Master's and Some College/Missing in the Instructional Support domain only. (*) In previous years these groupings differed.

Discussion of Findings

This report summarizes the findings for the 2021-22 school year for Philadelphia's PHLpreK program. The program has now finished its sixth year of operations and has continued to grow

year over year through partnerships with providers across the city. This component of the evaluation is meant to give information about program strengths and weaknesses that can inform professional development and technical assistance as the program continues to expand.

Pre-K classrooms in the PHLpreK program are averaging moderate to high levels of quality on the Emotional Support and Classroom Organization domains of the CLASS, and demonstrating improvement on the Instructional Support domain, though scores remain low. There are a number of trends in the CLASS scores overall that call for additional discussion here.

The first is related to trends in scores on CLASS over time. When comparing the scores of the full sample of classrooms observed this year (n=270) to the limited sample of classrooms we were able to observe in 2020 before the onset of the pandemic (n=103), scores increased on 9 out of 10 dimensions (all but the Behavior Management dimension of Classroom Organization), and scores on all three domains are higher than they were in 2020. In addition, the improvement in scores on Instructional Support from 2020 to 2022 is significant (improving from 2.30 to 2.75). This is particularly encouraging given the fact that many early childhood teachers have experienced higher levels of stress and lower levels of well-being at work than they experienced prior to the pandemic (e.g., Bigras et al., 2021; Souto-Manning & Melvin, 2022), which could negatively impact classroom instructional quality. The ability to continue to provide moderate to high quality early learning experiences, particularly in the areas of Emotional Support and Classroom Organization, is a programmatic strength.

When comparing the scores in 2022 to the scores in 2019, the findings are more nuanced. Scores are higher in 2022 on 4 out of 10 dimensions (Productivity in Classroom Organization, and all three dimensions in Instructional Support) than they were in 2019, and higher on the Instructional Support domain (2.75 in 2022, compared to 2.54 in 2019). However, scores are lower on Emotional Support (6.01 in 2019, and 5.86 in 2022) and Classroom Organization (5.60 in 2019 and 5.40 in 2022, a statistically significant difference). While these scores are all still in the range of "high" (i.e., above 5.00), this is a downward trend from the last time we were able to assess a larger sample. Two dimensions from each of these domains show larger decreases in scores than others. Negative Climate decreased from 6.91 in 2019 to 6.78 in 2022, and Behavior Management decreased from 5.81 in 2019 to 5.44 in 2022. In particular, with Behavior Management it seems teachers could use support in setting clear behavioral expectations, redirecting misbehavior, and being proactive in preventing misbehaviors.

Finally, the change over time in scores in the Instructional Support domain is encouraging, particularly in the finding that scores in 2022 on this domain are significantly higher than in 2020 or 2019. Scores on all three dimensions (Concept Development, Quality of Feedback, and Language Modeling) are higher when comparing 2022 data to 2019 and 2020 scores. In addition, the mean score on Language Modeling across all classrooms is 3.11, representing the first time scores on any dimension within this domain have been above 3.00. While these improvements are encouraging, there is still a lot of room for programs to enhance the practices teachers use in supporting children's understanding of new concepts. Structured supports around encouraging the use of advanced language, engaging in back-and-forth conversations with children, and in providing high quality feedback to children would benefit programs in the future.

Another finding worth discussion is how scores for classrooms in the PHLpreK program compare to those in other large city-funded programs. As shown in Figure 4, CLASS scores in PHLpreK in 2022 are lower than those for some other larger programs, but higher than others. On the ES domain, highlighted programs (Boston, Tulsa, Seattle, San Antonio and New York

City) had scores ranging from 5.2 to 6.8, with PHLpreK scoring in the middle of this range (5.86). On the CO domain, highlighted programs scored between 5.0 – 6.7, and PHLpreK (5.40) scored higher than Tulsa Public Schools, but lower than all other programs highlighted. Finally, on the IS domain, scores for other programs ranged from 2.75 – 4.1, and PHLpreK scored lower than all other programs highlighted. When taking into account the evaluations of other large city-funded programs, PHLpreK is performing similarly/slightly better in the Emotional Support domain, but has room for improvement in the Classroom Organization and Instructional Support domains.

Finally, we noted a few differences related to characteristics of providers and programs. In terms of STAR level, programs rated 4 STAR scored higher than 3 STAR programs on all domains (although these differences were relatively small). While providers rated 1-2 scored highest on ES and CO, there were just 4 of these programs, and this finding should thus be interpreted with caution. In terms of PHLpreK partner agency, providers contracted with PHMC scored higher on ES and CO, and providers with SDP scored slightly higher on IS, although the differences are very small (e.g., 2.75 for PHMC and 2.78 for SDP). Finally, providers that are a new site (n=43) scored lower across all three domains, although the differences were small. Ensuring there is targeted support available for these providers is an important takeaway from this year's findings. There were also some notable differences by lead teacher degree: Teachers with a Master's degree scored the highest across all three domains, and scored significantly higher than teachers with some college/missing college credentials on the Instructional Support domain.

The increase in scores in Instructional Support in 2022 is an encouraging result from this year's evaluation and is a positive development for children in the PHLpreK program, particularly as this domain is critical to children's learning and development in the pre-K years. However, the overall mean score for this domain still hovers below what some research indicates as the threshold needed to support a relationship between quality and children's outcomes. Continued progress toward improving scores, particularly in the Instructional Support domain, should be a focus of technical assistance and professional development efforts moving forward. In addition, while Classroom Organization scores are still high, they did fall below the 5.5 threshold they had been above in 2019. A focus on supporting teachers' behavior management strategies and ability to be proactive about behavior challenges would also be beneficial in future programming.

Acknowledgments

We are grateful to the William Penn Foundation who funded the work of NIEER that made this research possible (grant No. 16-17). We are also grateful for the partnership with the City of Philadelphia and PHMC. This partnership is instrumental in supporting the research and facilitate collaboration with program providers throughout the City across research components. Finally, we are thankful for all the programs and teachers that opened their doors for the work to be conducted.

References

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).

Bigras, N., Lemay, L., Lehrer, J., Charron, A., Duval, S., Robert-Mazaye, C., & Laurin, et I. (2021). Early childhood educators' perceptions of their emotional state, relationships with parents, challenges, and opportunities during the early stage of the pandemic. *Early Childhood Education Journal*, *49*(5), 775–787. https://doi.org/10.1007/s10643-021-01224-y

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

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

Nores, M., Barnett, W.S., Joseph, G., Stull, & Soderberg, J.S. (2018). Year 3 report: Seattle Pre-K program evaluation. New Brunswick, NJ: National Institute for Early Education Research & Seattle, WA: Cultivate Learning. 60 pp. 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. New Brunswick, NJ: National Institute for Early Education Research.

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

Philadelphia Commission on Universal Pre-kindergarten. (2016). *Final Recommendations Report*. Retrieved September 20, 2017, 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

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2006). CLASS: Classroom assessment scoring system manual preschool (Pre-K) version.

Souto-Manning, M., & Melvin, S. A. (2022). Early childhood teachers of color in New York City: Heightened stress, lower quality of life, declining health, and compromised sleep amidst COVID-19. *Early childhood research quarterly*, *60*, 34-48.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, 27(1), 122-154.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., & Gormley, W. T. (2013). Investing in our future: The evidence base on preschool education. Society for Research in Child Development.

Appendix A. Measures

Classroom Observation Measures

Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2006; Pianta & Hamre, 2009; Hamre, et al., 2014)

The Classroom Assessment Scoring System (CLASS) 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.