NIEER
NATIONAL INSTITUTE FOR EARLY EDUCATION RESEARCH

EVALUATION OF THE PHILADELPHIA PREK PROGRAM

Year 3 Report

October 2019

Milagros Nores, PhD, W. Steven Barnett, PhD, Zijia Li, PhD, Mariel Acevedo, BA and Charlie Whitman, BA. The National Institute for Early Education Research

About the Authors

Milagros Nores, Ph.D. Dr. Nores is a Co-Director of Research at The National Institute for Early Education Research (NIEER) at Rutgers University. Dr. Nores conducts research at NIEER on issues related to early childhood policy, programs, and evaluation, both nationally and internationally. She is also on staff with the Center for Enhancing Early Learning Outcomes (CEELO), a federally funded comprehensive center that provides technical assistance to state agencies around early childhood.

W. Steve Barnett, Ph.D. Dr. Barnett is a Senior Co-Director of the National Institute for Early Education Research (NIEER) and a Board of Governors Professor at Rutgers University. He is also Principal Investigator of the Center for Enhancing Early Learning Outcomes (CEELO). His research includes studies of the economics of early care and education including costs and benefits, the long-term effects of preschool programs on children's learning and development, the economics of human development, practical policies for translating research findings into effective public investments and the distribution of educational opportunities.

Zijia Li, Ph.D. Dr. Li is an Assistant Research Professor at the National Institute of Early Education Research (NIEER) at Rutgers University. Dr. Li is an experienced psychometrician and statistician. She has led and participated leading and conducting rigorous reliability and validity research studies for multiple subjects, including Peabody Picture Vocabulary Test III and IV, the Hawaii Early Learning Profile®, High/Scope Child Observation Record.

Mariel Acevedo, **BA.** Acevedo is Project Coordinator II at the National Institute for Early Education Research (NIEER). She leads NIEER' field work on the PHLpreK Evaluation Study and related work in New Jersey. She has contributed to NIEER's research on Philadelphia, New Jersey, and West Virginia.

Charles Whitman, BA. Whitman is Assistant Project Coordinator at the National Institute for Early Education Research (NIEER). He leads NIEER's field work on a New Jersey study and has also contributed to NIEER's research on Philadelphia, New Jersey, and West Virginia

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

Permission is granted to reprint this material if you acknowledge NIEER and the authors. For more information, call the Communications contact at (848) 932-4350, or visit NIEER at nieer.org.

Suggested citation: 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.

Table of Contents

Introduction

Philadelphia's Preschool Program (PHLpreK) has recently concluded its third year of programming. In the initial two years, the program partnered with about 80 programs and close to 140 classrooms and home providers. 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 children in the city, ages 3 and 4. The National Institute for Early Education Research (NIEER) has been for three years now conducting a multi-year, multi-site evaluation assessing program components, program quality, and children's learning and development.

The two previous reports for this evaluation have summarized the importance of high-quality preschool education to reduce persistent achievement gaps in kindergarten and throughout primary (Nores, Francis & Barnett, 2017; Nores et al., 2018). 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.[1] Strengthening and supporting preschool systems for them 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. This includes understanding the quality of classroom processes, space, and use of time.[2]

This report summarizes the third year of the Philadelphia's PreK Program (PHLpreK) evaluation, conducted by the National Institute of Early Education Research (NIEER). The overall evaluation is focused on research questions on the quality and performance of the PHLpreK program on children's learning outcomes and the overall economy, as well as classroom quality and program design. This report summarizes classroom quality for students in PHLpreK classrooms, provides a thorough description of the environment and teaching practices in these classrooms and summarizes the gains of children in the program. In addition, and unlike previous years, this report also summarizes quality in comparable classrooms in the city. The present report is one of the various components of this evaluation meant to support a data-driven continuous improvement approach to support improvements in quality in the city's program.

Findings suggest PHLpreK classrooms are averaging high to moderate levels of quality in the emotional support and classroom organization domains and have increased in the last year. Classrooms scores are low for instructional supports yet have notably increased relative to last year. That is, scores for all three domains of the CLASS measure increased in the previous year. Observations showed classrooms largely balanced in their use of a variety of activity settings. Under a fifth of the day is spent in transitions with reduced opportunities to learn. There is a strong correlation between the time spent in transitions and the time spent without any content area. In particular, while children are engaged in various content areas for different portions of the day, no learning content was observed for over a third of the day. Classrooms exhibit low integration across subject areas, a predominance of didactic over scaffolded interactions and metacognitive approaches remain quite absent. We explored quality separately for several subgroups of interest, including Star level, lead teacher credentials, area of study, PHLpreK

[1] Ceci & Papierno, 2005; Barnett, 2008; Duncan & Murnane, 2011; Nores & Barnett, 2015; Camilli et al., 2010, Friedman-Kraus, et al., 2016; Yoshikawa et al., 2013.
[2] Pianta & Hamre, 2009; Hamre et al. 2014.

partner agency, curriculum, and success by age 6. Small differences were found between subgroups and are reported. Higher quality classrooms exhibit less time in transition, more time in whole group and choice settings, more time with content areas (math and social sciences specifically), more scaffolded learning, less time without content and integrate content more often.

In 2018–19, as in the previous year, we assessed children's developmental gains over the school year. We find chronic absenteeism in the school year was at 39%. We report overall gains and if and how they differ among subgroups of children. We also assessed how centers and teaching and learning characteristics relate to child gains.

Hispanic, Black and DLL children and children with an IEP started the school year performing comparatively lower. We find gains are larger for Black children in receptive vocabulary but smaller in literacy and math, relative to their White peers. Unlike last year, this year we find stronger gains for dual language learners and children with IEPs in some developmental areas. These overall larger gains for Blacks and Hispanics are positive signs, even if not strong enough to be sustained when accounting for differences in other child factors and center quality and characteristics, in relation to literacy, math and executive functions.

We do not find many consistent and systematic effects across children's learning for the various teacher and center characteristics. Higher rated centers seem to be positively associated with children's learning. Child absenteeism is negatively related with child outcomes. We find evidence of the importance of classroom organization as it relates to various outcomes, consistent with last year. We also found instructional supports to be associated with children's outcomes when above 3.5. In addition, more time spent in word identification was related to some child outcomes. Teachers reported various mechanisms of professional development and technical assistance and agree that most existing supports to be useful. However, coaching was reported at a low rate (mostly for less then a day) and teachers expressed a need for stronger supports on behavior management and socio-emotional development, and on classroom quality. The report includes some recommendations on areas to strengthen quality.

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 third year of the study. Data collection included assessing child gains (fall and spring assessments), classroom observations, and teacher and director surveys 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?
2. How does quality in PHLpreK classrooms compare to quality in other classrooms in the City of Philadelphia?
3. What are the learning gains of children in vocabulary, literacy, math, executive functions, and social-emotional development through 2018–19, and how did gains relate to classroom quality and children's background characteristics? How do these compare relative to prior years?

The PHLpreK evaluation was designed to assess the 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. Procedures and measures are described in detail below. Children were assessed early in the Fall of 2018, and again at the end of the school year in the Spring of 2019. Classroom observations were conducted to assess teacher-child interactions and quantify children experiences during a typical learning day. Classroom observations took place between February and June 2019. Like previous years, quality was assessed using well-known observation protocols during two visits of about two and a half to three hours each.

1. Sample

In the 2018–19 school year, NIEER assessed 503 children in 149 PHLpreK classrooms (16 which were home-based providers) at pre- and post-test. To recruit children, consent forms were distributed to families as part of the PHLpreK enrollment process. A total of 585 children were assessed at pre-test with family consent, and 82 children (14%) were lost due to their attrition from the program or parental withdrawal of consent. We randomly selected four children per classroom. The final sample of children was 63% African American, 14% Hispanic, 14% White, and 9% Asian, mixed-race, or other. This is closely comparable to the K-12 PHL school district demographics of 53% African American, 19% Latino, 14% White, and 13% other.[3]

Classroom quality was assessed with two separate instruments used during two different days: CLASS Pre-K and EduSnap (described below). The CLASS was used in 147 classrooms (center and home-based). The EduSnap was only used in the 129 center-based classrooms.

2. Measures and Procedures

Classroom quality was captured using two observational instruments: *The Classroom Assessment Scoring System Pre-K* (*CLASS Pre-K*; Pianta, La Paro, & Hamre, 2008) and the *EduSnap Classroom Observation Measure* (*EduSnap;* Ritchie, Weiser, Mason, & Holland, 2015). The CLASS measures teacher-child interactions and classroom processes. Home-based providers were observed only using the CLASS. The EduSnap quantifies the experiences of children in various activity settings, teaching and learning strategies, and content areas. More detail on classroom observation measures is provided in Appendix A.

Children were assessed with a measure of receptive language (the *Peabody Picture Vocabulary Test—Fourth Edition or PPVT-IV*; Dunn & Dunn, 2007), emerging literacy (the letter-word identification subtest from the *Woodcock-Johnson Psycho-Educational Battery—Fourth Edition or WJ-IV;* Schrank, Mather & McGrew, 2014) and mathematics (the applied problems subtest from the WJ-*IV*). In addition, children were assessed with two measures of executive functions, which capture children's inhibitory control, short term memory, and attention. These are the *Dimensional Change Card Sort Task* (DCCS; Zelazo, 2006) and the *Peg Tapping Test* (PT; Diamond & Taylor, 1996). Socio-Emotional development was measured

[3] https://dashboards.philasd.org/extensions/philadelphia/index.html#/

using the *Caregiver-Teacher Report Form* (C-TRF: Achenbach, 2009). More detail on child measures is provided in Appendix A.

Observers were trained to reliability before conducting observations of classroom quality. EduSnap observers were trained by the developer and subsequently completed kappa reliability with pre-coded online videos. CLASS observers were trained by a CLASS Affiliate Trainer from NIEER, completed the online reliability 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 first for classroom observations (the CLASS and then for EduSnap), followed by a comparison to a set of comparable classrooms and centers in the city of Philadelphia. A third section reports children's gains across child and center characteristics and in relation to observed classroom quality. A final section describes professional development and technical assistance as reported by teachers and directors. We conclude with a summary of the findings and recommendations.

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 the 2018 and 2019 (from 5.64 to 6.01). This is also the case for Classroom Organization (CO) (increasing from 5.28 to 5.60) and for Instructional Support (IS) (increasing from 2.05 to 2.54). The increase in scores is statistically significant for CLASS ES, CLASS CO, and CLASS IS. Results for each domain are discussed further below.

Observed increases in this third year in ES were of 0.51 SD (standard deviations),[4] in CO these were of 0.38 SD and in IS these were of 0.74 SD. These are significant increases in quality.

[4] 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 has a distribution 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
Emotional Support Domain (ES)	**5.85 (2.85-6.90)**	**5.64[a] (3.20-6.95)**	**6.01[b] (3.05-7.00)**
1. Positive Climate	5.90 (1.60-7.00)	5.73 (3.20-7.00)	6.13 (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)
3. Teacher Sensitivity	5.69 (2.20-7.00)	5.52 (2.80-7.00)	5.89 (1.60-7.00)
4. Regard for Student Perspectives	5.03 (2.00-6.80)	4.65 (2.40-7.00)	5.11 (1.60-7.00)
Classroom Organization Domain (CO)	**5.34 (1.87-6.93)**	**5.28 (2.80-6.93)**	**5.60[b] (2.40-7.00)**
5. Behavior Management	5.49 (1.60-7.00)	5.48 (2.80-7.00)	5.81 (2.40-7.00)
6. Productivity	5.76 (1.80-7.00)	5.65 (2.80-7.00)	5.72 (2.40-7.00)
7. Instructional Learning Formats	4.77 (1.60-7.00)	4.72 (1.80-6.80)	5.27 (2.00-7.00)
Instructional Support Domain (IS)	**2.41 (1.00-5.00)**	**2.05[a] (1.00-4.60)**	**2.54[b] (1.00-5.33)**
8. Concept Development	2.09 (1.00-4.80)	1.84 (1.00-4.00)	2.27 (1.00-5.60)
9. Quality of Feedback	2.23 (1.00-5.00)	1.91 (1.00-4.40)	2.53 (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)

*The Negative Climate dimension is reverse scored so that a high score represents "good." [a]Statistically significant difference between 2017 and 2018. [b]Statistically significant difference between 2018 and 2019 distributions of scores.

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 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. Overall, 93% of classrooms were found to have ES levels above 5 (up from 81%). For CLASS CO, an improvement is observed on average scores and a reduction in the lowest scores attained (a shorter tail); 78% of classrooms were found to have CO scores above 5 (up from 66%). The distribution for CLASS IS also shifted to the right this year, with a greater number of classrooms attaining higher scores and with some classrooms attaining higher levels than observed in previous years; 26% were above the threshold of 3 in IS (up from 9% in year 2 and 19% in year 1).

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

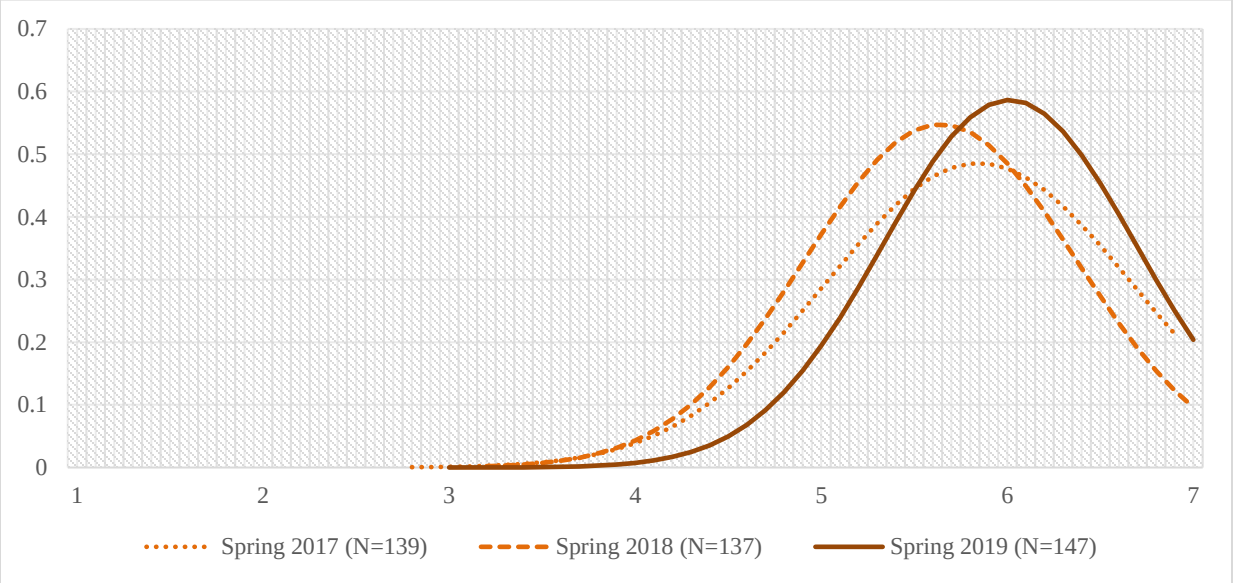


Figure 2. Distribution of CLASS Classroom Organization scores for 2017, 2018 & 2019.

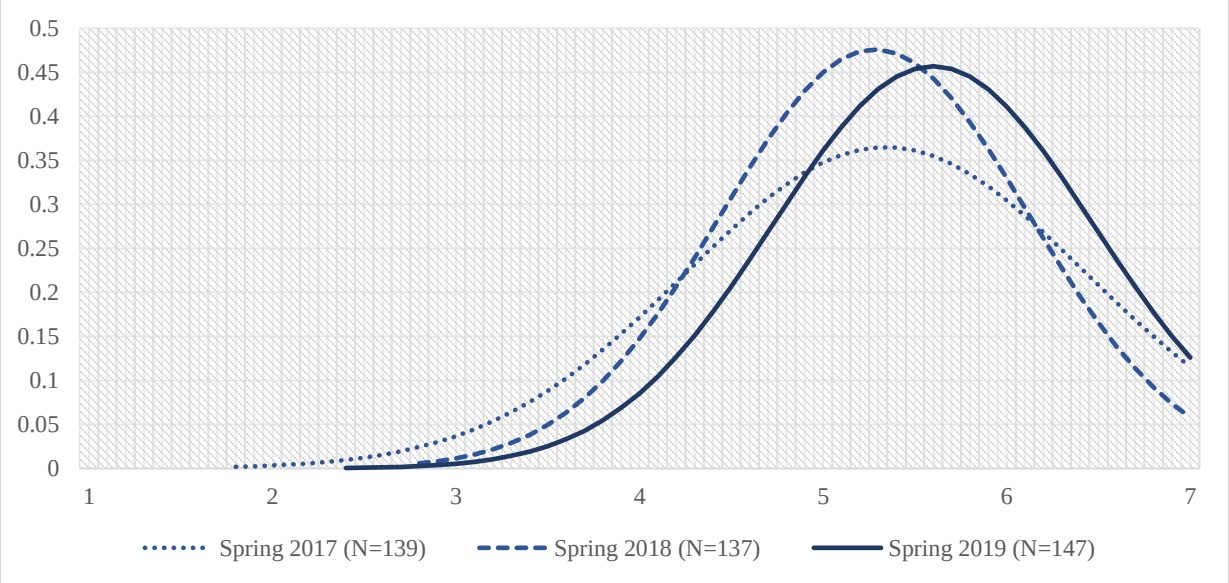
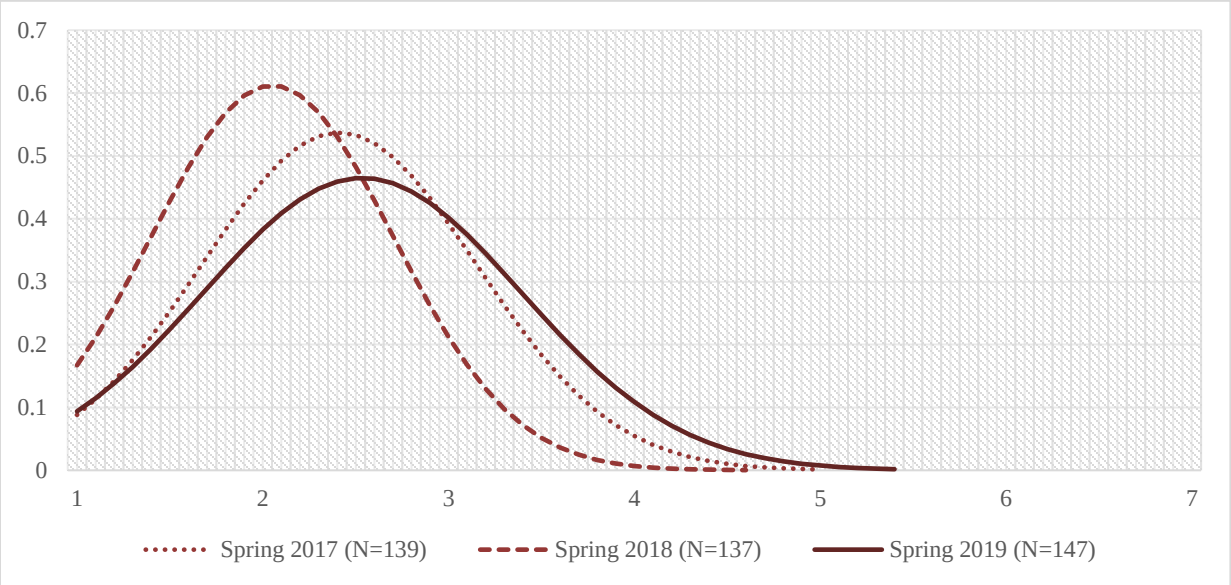


Figure 3. Distribution of CLASS Instructional Support scores for 2017, 2018 & 2019.



CLASS Pre-K Domains

The Emotional Support (ES) domain focuses on teaching behaviors that support the development of supportive relationships between teachers and children, and that help children enjoy the learning process and be comfortable in the classroom. The overall mean score for ES is 6.01 (SD 0.68), putting it in the high-quality range. The minimum score is 3.05, which indicates no classrooms with low levels of emotional support. The highest scoring dimension is Negative Climate (6.91) indicating that on average classrooms exhibited extremely few negative interactions between teachers and children and among children. The lowest scoring dimension is Regard for Student Perspectives (5.11). Increasing this dimension requires that teachers become flexible and follow children's lead, provide choice in what children are doing, and encourage student responsibility. Additional opportunities for children to express their ideas and to be involved in activities that will allow them to be active, would further increase this score.

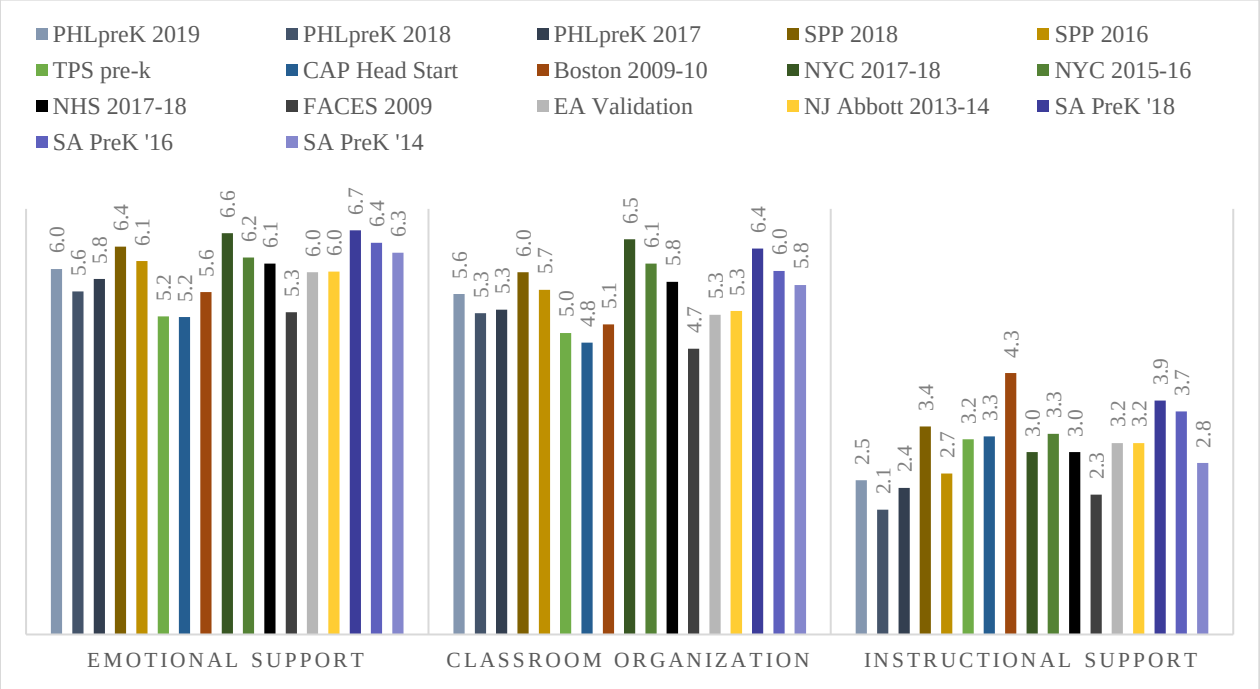
The Classroom Organization (CO) domain focuses on teaching behaviors that use effective methods to manage behavior expectations, instructional time and routines, and the provision of activities that maximize children's interests and engagement. The average mean score for the Classroom Organization Domain is 5.60 (SD 0.87). The high scores indicate that teachers show effective methods to both prevent and redirect misbehavior, and most student behaviors observed being consequently compliant and appropriate. High scores also indicate that teachers are organized and plan ahead, provide clear instructions, and quickly take care of managerial tasks. Within this domain, Instructional Learning Formats scored lower than the other two dimensions (5.27). Increasing this dimension requires consistent use of interesting and creative materials, explicitly orienting children towards learning objectives and use of effective questioning that expands children's involvement. To further increase this domain, teacher involvement in learning activities and exposure to opportunities that allow children to use different modalities, including hands-on activities, are required.

Instructional Support measures the interactions by which teachers facilitate and encourage higher-order thinking skills, expand understanding and learning, and promote language development. This domain has consistently scored lower across preschool evaluations and systems. However, it is a critically central domain to further children's learning and development. The average IS score is 2.54 (SD 0.68) with averages ranging from 1 to 5 on a 7-point scale. Concept Development and Quality of Feedback are the two dimensions scoring lower. Concept Development refers to teaching behaviors that focus on facilitating children's thought process. Effective use of concept development strategies includes problem-solving, experimentation, brainstorming, as well as encouraging children's creative processes. Concept Development strategies also include linking concepts and ideas to children's lives and the real world. Quality of Feedback measures ways in which teachers scaffold learning, engage in back-and-forth exchanges, ask follow-up questions to provide specific information, utilize metacognitive approaches to expand on children's thinking processes, and recognize children's efforts to increase student involvement and persistence. Scores in Language Modeling subdomain are higher but still average below the threshold of 3. Consistent and intentional use of strategies is critical to increasing the IS dimensions further.

CLASS Pre-K Comparison of Programs

Figure 4 reports score patterns for the PHLpreK in relation to those of other cities and states. The PHLpreK CLASS from 2019, 2018 and 2017 are reported by domain together with scores from various other programs in the U.S. in Figure 2. This includes high-quality city programs and shows how other programs have shifted in their first few years as they grow in access. For example, the PreK4All program in San Antonio, SPP in Seattle and Pre-K for All in NYC have all mostly evidenced increases in CLASS scores in their first few years.

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 higher star levels score higher on IS. This is also the case for teachers with either a masters' degree or a CDA/ECE training. Concerning partner agency, classrooms in sites in collaboration with PHMC evidence higher scores across all domains. Those using the Mother Goose curriculum perform lower on IS. Success by 6 classrooms have higher ES and CO scores, but lower IS scores. Classrooms in sites that started in PHLpreK only this last year perform at similar levels for ES and IS.

Table 2. CLASS domains scores by subgroups, N = 147.

		CLASS Mean Scores		
		Emotional Support	Classroom Organization	Instructional Support
STAR Level	1-2 (n=3)	6.58	5.42	2.40
	3 (n=72)	5.96	5.54	2.56
	4 (n=72)	6.04	5.67	2.52
Lead Teacher Credential	CDA/ECE Course (n=8)	6.16	5.50	2.81
	AA (n=53)	6.03	5.54	2.52
	BA (n=48)	5.97	5.64	2.52
	MA, Med, MSc (n=23)	6.17	5.91	2.82
	Other, Studying, Missing (n=15)	5.77	5.26	2.05
PHLpreK Partner Agency	UAC (n=45)	6.02	5.53	2.35
	PHMC (n=73)	6.03	5.71	2.76
	1199c (n=14)	5.95	5.51	2.28
	SDP (n=15)	5.95	5.36	2.26
Curriculum	Creative (n=94)	6.07	5.64	2.55
	Creative + (n=42)	5.91	5.48	2.51
	Mother Goose (n=10)	5.87	5.83	2.47
	Other (n=1)	-	-	-
Success by 6	Yes (n=5)	6.35	6.09	2.43
	No (n=142)	6.00	5.58	2.54
New Site	Yes (n=11)	6.01	5.45	2.56
	No (n=136)	6.01	5.61	2.53

CLASS Pre-K Domains for PHLpreK and Control classrooms

This year we assessed classroom quality using the CLASS in 58 randomly selected classrooms.[5] In addition, we surveyed programs on their funding sources. The sample of control classrooms is 21% Head Start providers, 54% in centers, and 25% are home-based providers. One classroom

[5] A list of providers for the City of Philadelphia was used. The list was then restricted to the zip codes in which PHLpreK is currently providing services. Providers were categorized as Head Start., Center-based non-Head Start, and home providers. A random list was created for a target of N=60 classrooms, with a percentage of each of these groups targeted that resembled the current PHLpreK sites distribution across these three groups. Random alternative lists were used to recruit additional providers if providers from the initial random sample either refused participation, did not have at least 3 children in the corresponding 3-5 age group, or did not respond to the invitation to participate.

was randomly selected for the centers with more than one classroom of the corresponding age group.

A comparison of overall scores for PHLpreK providers and the sample of control providers is summarized in Table 3. On average, PHLpreK classrooms and home providers exhibited higher scores across all CLASS domains. These differences are of 0.23 SD for ES, 0.14 SD for CO, and of 1.02 for IS. PHLpreK settings on average evidence higher variation. Statistically significant differences are marked with an asterisk. This is the case for all dimensions under the instructional support domain.

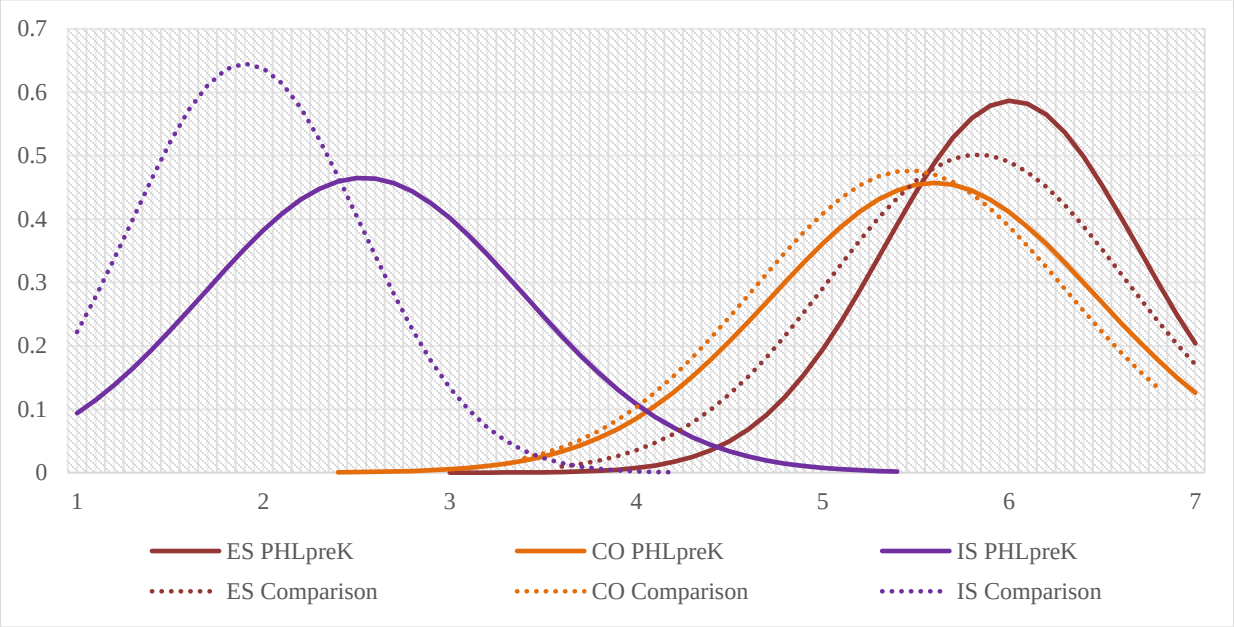
Table 3. CLASS domains & dimension scores for PHLpreK and comparison providers in PHL.

Domains and Dimensions	PHLpreK 2019 (N=147)				Comparison group 2019 (N=58)			
	Mean	**(SD)**	**Min.**	**Max.**	**Mean**	**(SD)**	**Min.**	**Max.**
Emotional Support	**6.01**	**(0.68)**	**3.05**	**7.00**	**5.83**	**(0.80)**	**3.65**	**7.00**
1. Positive Climate	6.13	(0.86)	2.40	7.00	6.00	(0.96)	3.40	7.00
2. Negative Climate[a]	6.91	(0.24)	5.40	7.00	6.89	(0.23)	6.00	7.00
3. Teacher Sensitivity	5.89	(0.92)	1.60	7.00	5.69	(1.05)	3.20	7.00
4. Regard for Student Perspectives*	5.11	(1.19)	1.60	7.00	4.74	(1.41)	1.00	7.00
Classroom Organization	**5.60**	**(0.87)**	**2.40**	**7.00**	**5.46**	**(0.84)**	**3.47**	**6.80**
5. Behavior Management	5.81	(0.94)	2.40	7.00	5.82	(0.89)	3.40	7.00
6. Productivity	5.72	(0.91)	2.40	7.00	5.64	(0.88)	3.80	6.80
7. Instructional Learning Formats*	5.27	(1.02)	2.00	7.00	4.93	(1.11)	2.60	7.00
Instructional Support*	**2.54**	**(0.86)**	**1.00**	**5.33**	**1.90**	**(0.62)**	**1.07**	**4.13**
8. Concept Development*	2.27	(0.97)	1.00	5.60	1.65	(0.67)	1.00	4.00
9. Quality of Feedback*	2.53	(0.93)	1.00	5.20	1.91	(0.68)	1.00	3.80
10. Language Modeling*	2.80	(0.90)	1.00	5.80	2.15	(0.69)	1.00	4.60

[a]Inversely coded for ease of interpretation.

Figure 5 illustrates the distributions for PHLpreK classrooms (solid line) and comparison sites (dotted line). PHLpreK distributions are further to the right (exhibiting a larger fraction of classrooms at higher quality levels), particularly for IS and ES, although the mean score for IS is still below the threshold score of 3.

Figure 5. CLASS domain distributions for PHLpreK and comparison classrooms.



EduSnap results

The EduSnap includes measures for activity settings, content areas and teaching and learning approaches. We present these for all classrooms (excluding home-based providers) and we follow this with summaries by selected program-level characteristics.

Activity Settings

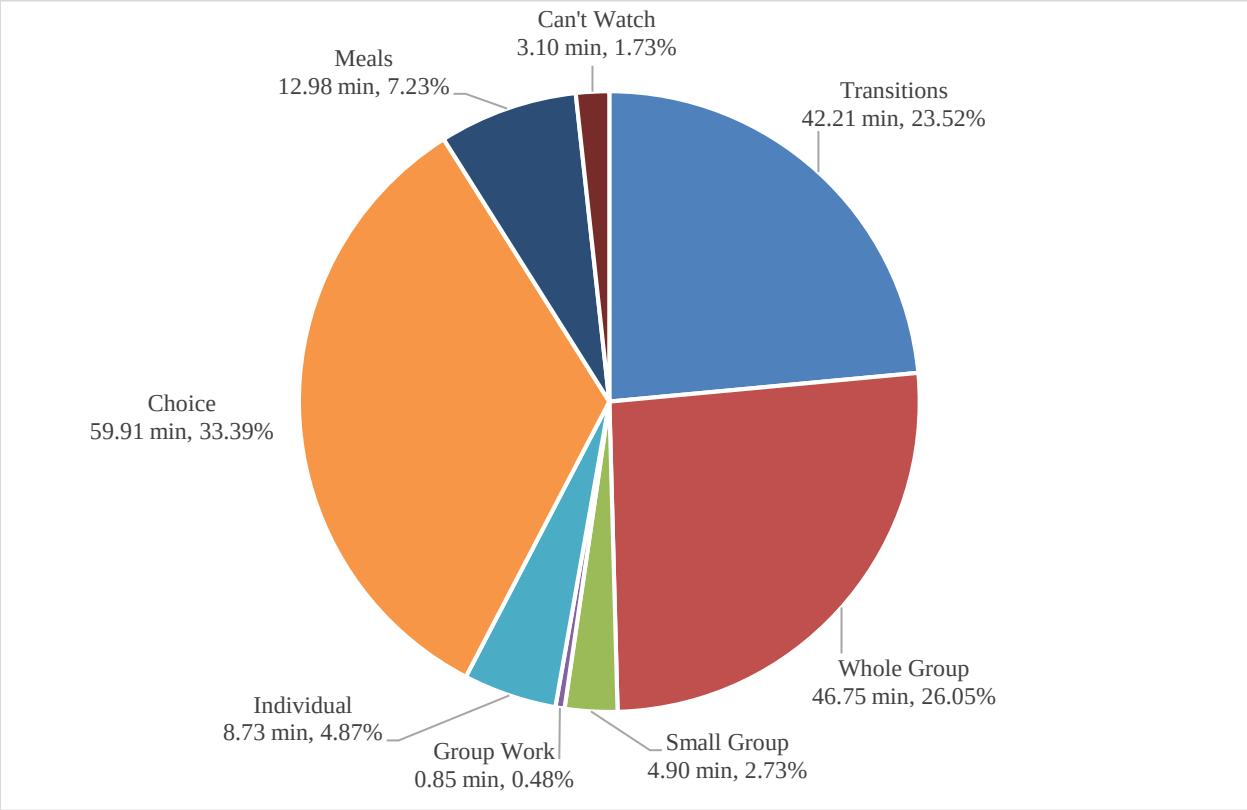
Activity settings capture transitions, whole group teacher-led activities, small group teacher-led activities, group work, and choice, as well as work on individual assignments and meals. In the EduSnap, 100 percent of the three hours observed are coded by activity settings. This implies that each of these captures the percentage of time that four randomly selected children are engaged in a particular setting. Random selection of children implies that these experiences are representative of the average child experience in the classroom.

Figure 3 illustrates children's experiences throughout the day. The time spent in choice (33%) and whole group (26%) are on average well balanced, although time in whole group has decreased from last year about 4% points. Time spent on transitions (24%) has slightly increased (2% points) and the time spent in small groups (3%) has remained equally low. Group work activities are more effectively used in later grades, which makes the low percentage (less than 1%) developmentally appropriate. No large changes are observed for any of the activity settings for the 2018–2019 period.

We also assessed the range of the use of different settings (not shown in the figure). There are classrooms for which less than 10% of the time in transitions was observed, and in the higher range, the maximum time spent in transitions observed was 60%. This has corresponding implications in choice and whole group. Time spent on choice activities ranged between 0% to 70%. Time spent in whole group activities ranged between 1 and 67%.

The conversion of percentage time to minutes aids in the interpretation of what these means throughout the day and allows thinking even at the yearly implications in time. One percent is on average equivalent to 1.8 observed minutes, five percent to 9 minutes, ten percent to 18 minutes, and so forth. Therefore, Figure 3 also reports minutes. The minutes are calculated based on the average time across observations, which was 180 minutes (three hours).

Figure 6. Percentages of Time and Minutes Spent in Activity Settings, n=129.



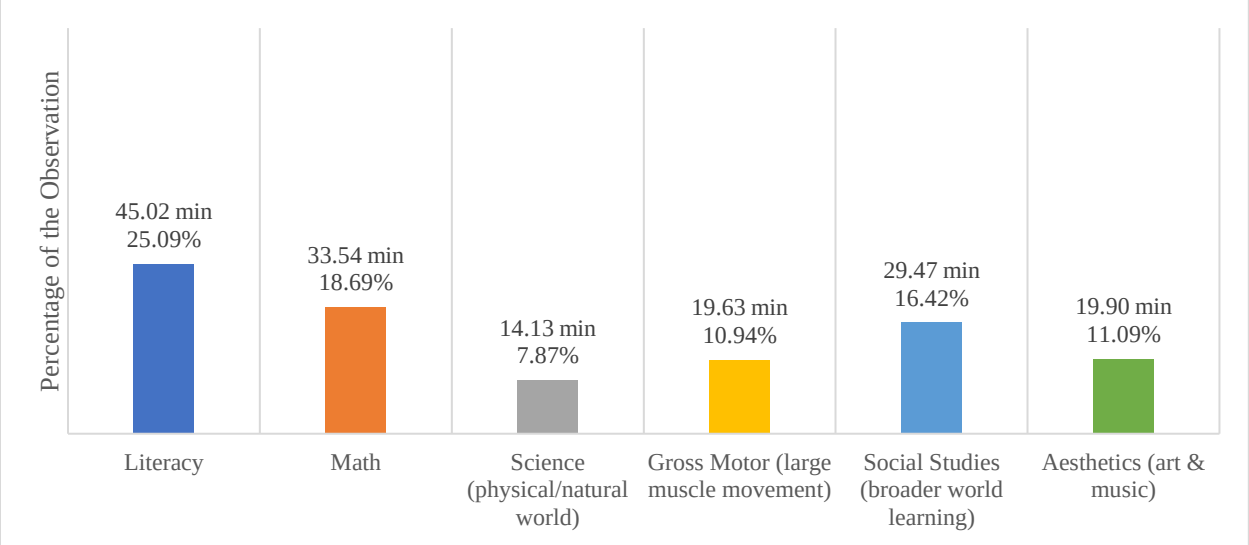
Patterns across subgroupings of programs by STAR ratings, curriculum, teacher qualifications, Success by 6 and seniority in PHLpreK, are reported in Table 4. No obvious patterns are observable. Transitions were observed to vary between 20 and 25% (lower for teachers with an MA/ME and lower in Success by 6 classrooms. The percentage of whole group activities ranged between 23 and 31% (lower in higher STAR levels, and for teachers with a CDA). Choice was observed for an average of about 25 to 39% and was higher in classrooms were teacher reported having an MA/ME and in higher STAR level classrooms.

Table 4. Percentages of time spent in Activity Settings by Subgroups, n=129.

		% of Time						
		Meals	Transitions	Whole Group	Small Group	Individual	Choice	Group Work
STAR Level	1-2 (n=3)	7.17	22.76	30.60	3.15	10.19	25.40	0.00
	3 (n=69)	7.46	24.64	25.97	3.00	5.12	31.39	0.59
	4 (n=57)	6.96	22.22	25.91	2.39	4.27	36.23	0.37
Lead Teacher Credential	CDA/ECE (n=6)	5.04	23.44	23.99	6.12	8.17	31.19	0.00
	AA (n=51)	7.22	23.88	25.96	2.18	4.78	33.79	0.60
	BA/BS (n=45)	8.38	24.74	27.04	2.95	5.36	28.81	0.51
	MA/ME (n=18)	5.99	20.94	26.11	2.89	3.76	38.73	0.46
	Missing (n=9)	5.54	20.59	22.89	2.22	2.90	44.75	0.00
PHLpreK Partner Agency	Phila SD (n=15)	7.74	21.00	30.07	1.89	5.67	32.13	0.11
	PHMC (n=69)	7.22	23.23	25.04	3.02	4.75	33.85	0.78
	UAC (n=45)	7.09	24.82	26.26	2.57	4.78	33.10	0.14
Curriculum	Creative (n=89)	7.27	22.69	25.04	3.22	4.76	34.77	0.56
	Creative + (n=30)	6.46	24.55	28.79	1.68	5.07	31.80	0.06
	Mother Goose (n=9)	9.07	27.96	26.76	1.78	5.70	24.99	1.10
	Other (n=1)	10.56	26.67	27.22	0.00	1.11	33.89	0.00
Success by 6	Yes (n=5)	6.43	17.94	33.31	1.78	4.20	35.11	0.00
	No (n=124)	7.27	23.75	25.76	2.77	4.89	33.32	0.50
New Site	Yes (n=9)	6.85	22.87	26.45	2.35	6.66	32.17	1.66
	No (n=120)	7.26	23.57	26.02	2.76	4.73	33.48	0.39

A second component of the EduSnap is the focus on content areas, including literacy, math, science, aesthetics, and gross motor, with an underlying concept on content area balance. Figure 6 depicts the percentage of time spent on the different content areas measured. While like 2017 and 2018 the percentage of time spent on literacy dominates over other content areas, it has decreased. Average total time spent in math has slightly increased. Time spent in Science and Gross motor activities remain the lowest (8 and 11%) and has shown no real change over the years. The percentage of time spent other areas has decreased: literacy from 35% to 25%, social studies from 20% to 16%, and aesthetics from 17% to 11%.

Figure 6. Percentages of Time and Minutes Spent in Content Areas.



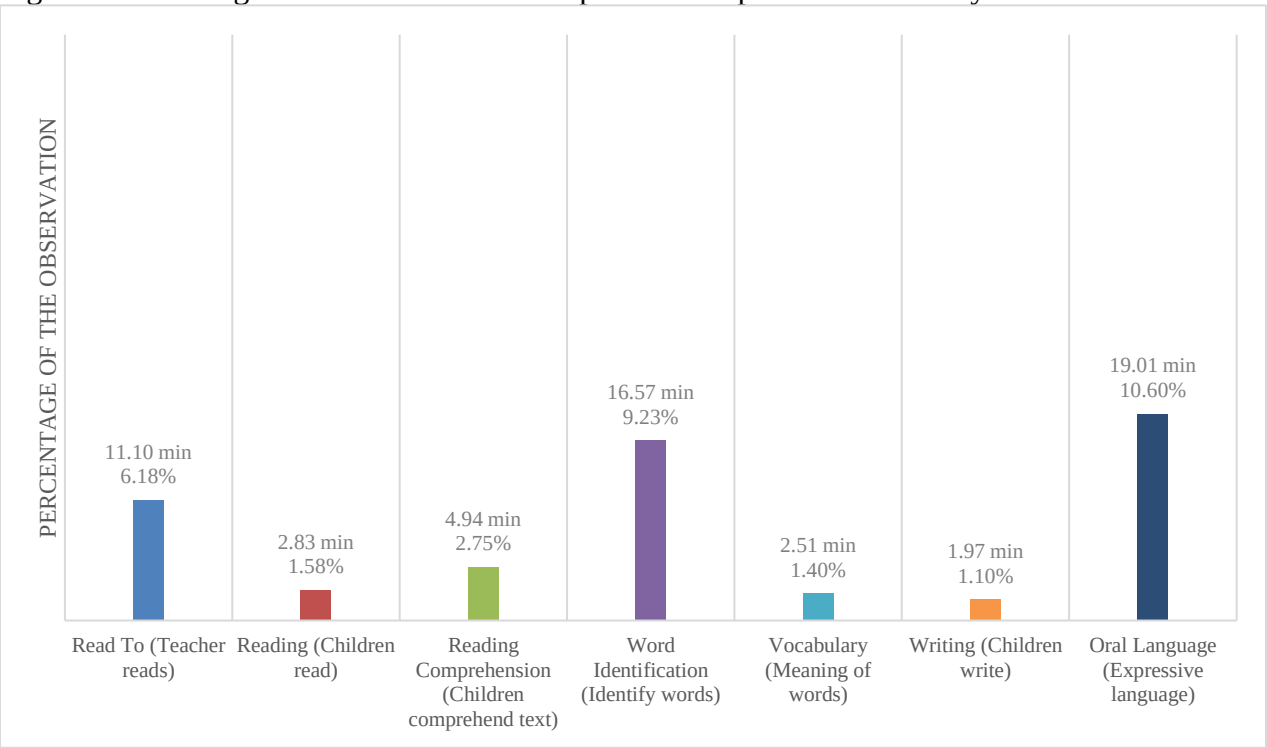
The percentage of time spent in the different content areas for subgroups of interest are reported in Table 7. New sites are observed having lower percentages of time spent in Literacy, but the opposite is the case for Science. Higher percentages of social studies are observed in classrooms where the lead teacher has an MA/ME. Mother Goose classroom spend the highest percentages of time in Science, a reverse from the previous year.

Table 7. Percentages of time spent in various content areas by subgroups, n=129.

		% of Time					
		Literacy	Math	Science	Gross Motor	Social Studies	Aesthetics
STAR Ratings	1-2 (n=3)	33.95	26.71	9.38	9.89	15.65	29.75
	3 (n=69)	25.46	18.94	8.41	10.47	16.02	10.18
	4 (n=57)	24.18	17.96	7.15	11.57	16.95	11.21
Lead Teacher Credential	CDA/ECE (n=6)	28.15	15.70	5.93	13.36	16.54	14.19
	AA (n=51)	23.25	17.04	8.87	11.47	14.94	10.83
	BA/BS (n=45)	26.50	19.76	6.74	9.93	16.19	10.03
	MA/ME (n=18)	25.32	21.03	9.18	11.92	18.32	12.76
	Missing (n=9)	25.97	19.99	6.54	9.43	22.13	12.46
Area of Study	ECE (n=70)	24.52	18.30	7.27	10.29	15.80	10.64
	Education (n=17)	25.31	20.64	10.63	12.13	18.11	13.59
	Other, Missing (n=42)	25.95	18.55	7.77	11.54	16.79	10.82
PHLpreK Partner Agency	Phila SD (n=15)	27.92	19.83	10.96	10.44	17.15	11.75
	PHMC (n=69)	24.42	18.72	7.87	11.18	16.63	11.19
	UAC (n=45)	25.16	18.26	6.85	10.74	15.86	10.72
Curriculum	Creative (n=89)	24.86	19.43	8.00	11.58	16.46	11.15
	Creative + (n=30)	25.28	17.46	7.19	10.79	16.65	11.57
	Mother Goose (n=9)	26.41	16.40	9.35	5.65	15.05	9.58
	Other (n=1)	27.78	10.56	3.33	6.67	18.33	5.00
Success By 6	Yes (n=5)	28.92	28.38	10.43	10.95	24.70	9.09
	No (n=124)	24.93	18.30	7.77	10.94	16.09	11.17
New Site	Yes (n=9)	19.54	20.35	13.09	9.42	13.48	15.33
	No (n=120)	25.50	18.56	7.48	11.06	16.65	10.77

In terms of literacy content, the EduSnap differentiates among components of literacy. Figure 7 illustrates the lack of balance among literacy components. Although the different components of literacy are not high, the (relatively) stronger areas are time spent on oral language (11%) and word identification (10%, increasing from 9.5% the previous year). Time spent on being read to (6.2%), vocabulary (1.4%), reading (1.6%), or writing (1%) has changed very little and remain particularly low.

Figure 7. Percentages of Time and Minutes Spent in Components of Literacy.



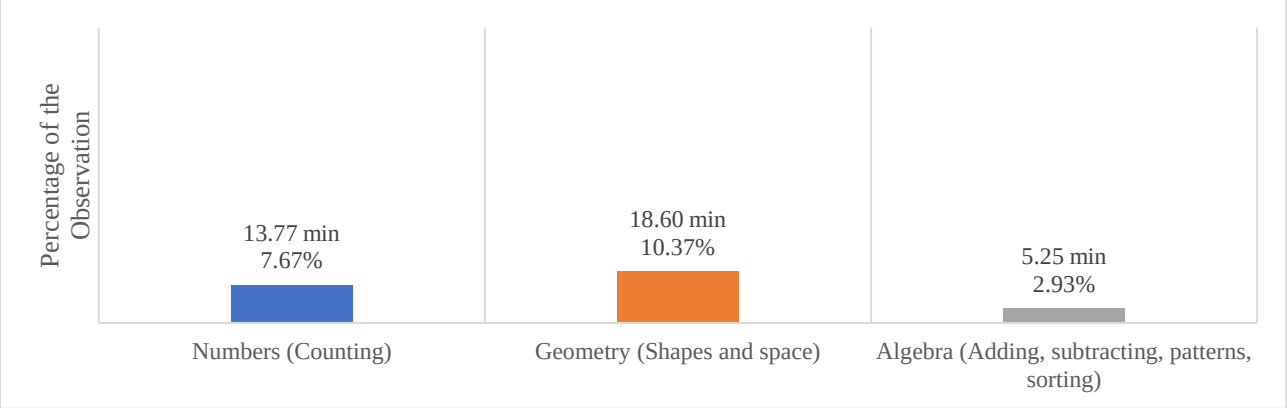
There is some small variation in observed time across literacy components and center/teacher characteristics. Table 5 reports this for selected subgroups across program level characteristics. Time spent in oral language varies between 8 percent and 17 percent, depending on the site. Time spent on word ID is lower in higher-rated classrooms and classrooms with a teacher with an MA/ME, both of do not simultaneously exhibit higher percentages in other literacy content areas. Teachers with an MA/ME do spend on average a slightly higher amount of time on all of the literacy content.

Table 5. Percentages of time spent in Components of Literacy by Subgroup, n=129.

		Literacy Activities						
		Read-to	Reading	Compreh.	Word ID	Vocabulary	Writing	Oral Language
STAR Ratings	1-2 (n=3)	7.56	0.93	3.31	13.22	2.01	0.00	17.03
	3 (n=69)	6.16	1.74	2.50	9.33	1.43	1.06	10.48
	4 (n=57)	6.14	1.41	3.03	8.91	1.33	1.20	10.40
Lead Teacher Credential	CDA/ECE (n=6)	4.76	1.51	2.97	10.88	0.74	1.48	12.98
	AA (n=51)	6.87	1.57	2.65	7.94	1.56	0.82	9.72
	BA/BS (n=45)	5.54	1.31	2.57	10.07	1.23	1.32	11.57
	MA/ME (n=18)	7.02	1.63	3.65	8.76	1.96	1.28	10.48
	Missing (n=9)	4.75	2.83	2.28	12.22	0.68	0.99	9.32
Area of Study	ECE (n=70)	6.03	1.85	2.49	8.97	1.18	1.41	9.67
	Education (n=17)	6.13	1.14	2.64	11.24	2.11	0.78	9.24
	Other, Missing (n=42)	6.47	1.29	3.24	8.87	1.48	0.72	12.69
PHLpreK Partner Agency	Phila SD (n=15)	7.88	1.82	3.21	8.73	1.55	1.72	12.39
	PHMC (n=69)	5.31	1.78	2.56	9.06	1.55	1.11	10.33
	UAC (n=45)	6.96	1.18	2.90	9.67	1.13	0.88	10.40
Curriculum	Creative (n=89)	5.82	1.50	2.69	9.52	1.39	1.11	10.31
	Creative + (n=30)	8.14	1.65	3.21	9.29	1.44	1.12	9.73
	Mother Goose (n=9)	3.32	1.59	1.72	6.69	1.22	0.86	15.98
	Other (n=1)	6.11	6.11	3.89	5.56	2.22	1.67	13.89
Success By 6	Yes (n=5)	3.22	0.33	1.44	9.74	2.11	3.32	15.96
	No (n=124)	6.30	1.63	2.81	9.21	1.37	1.01	10.38
New Site	Yes (n=9)	6.84	1.77	1.79	3.72	1.41	1.58	8.05
	No (n=120)	6.13	1.56	2.83	9.65	1.40	1.07	10.79

The EduSnap also captures the percentage of time spent in three specific math areas. These are numbers, geometry, and algebra (Figure 8). All components of math remain low. However, relative to the spring of 2018, an increase was observed in the time spent in all three of these.

Figure 8. Percentages of Time and Minutes Spent in Components of Math.



The percentages of time spent in the different components of math for the subgroups of interest are reported in Table 6. Generally, no clear differences emerge as what differences do exist are quite small. Lead teachers with an AA spend just slightly less time on numbers, while teachers with a BA/BS or MA/ME spend slightly more time on geometry and algebra. Time spent of the latter is generally low (3% or less).

Table 6. Percentages of time spent in components of math by subgroups, n=129.

		% of Time		
		Numbers	Geometry	Algebra
STAR Ratings	1-2 (n=3)	13.98	13.30	3.66
	3 (n=69)	7.92	10.28	3.27
	4 (n=57)	7.04	10.32	2.47
Lead Teacher Credential	CDA/ECE (n=6)	8.80	7.18	1.58
	AA (n=51)	6.95	9.53	2.40
	BA/BS (n=45)	8.77	10.46	3.29
	MA/ME (n=18)	7.45	12.73	3.76
	Missing (n=9)	5.92	12.03	3.33
PHLpreK Partner Agency	Phila SD (n=11)	7.37	12.30	2.63
	PHMC (n=69)	7.91	10.28	2.79
	UAC (n=45)	7.41	9.85	3.24
Curriculum	Creative (n=89)	8.49	10.54	3.00
	Creative + (n=30)	5.46	10.42	3.03
	Mother Goose (n=9)	7.19	9.02	2.15
	Other (n=1)	5.00	5.00	0.56
Success By 6	Yes (n=5)	9.42	21.18	2.55
	No (n=124)	7.60	9.93	2.94
New Site	Yes (n=11)	6.12	12.35	3.70
	No (n=136	7.79	10.22	2.87

With the EduSnap, it is possible to analyze curriculum integration across areas (Figure 9) using the codes for the different content areas explored above. Classrooms were observed with no content 35% of the time (an increase from 28% in 2018). This is the equivalent of 64 minutes. The correlation[6] between no content and transitions is of 0.68, which implies a strong association between the time spent on transitions and the time without content. However, on average, the time spent without content is higher than the time spent in transitions.

One content area (no integration) accounted for 43% of the time (about 77 minutes), without changes relative to 2018. Integration of two or more content areas was observed in about 22% of the time (a reduction from 28% in 2018).

[6] A correlation coefficient is a number between −1 and +1 that represents the linear dependence between two variables. Negative correlations imply negative associations between variables. Higher correlations (closer to 1 or −1) imply higher associations. A correlation of 1 would imply a perfect and positive association between two variables.

Figure 9. Percentages of Time and Minutes Spent in Curriculum Integration

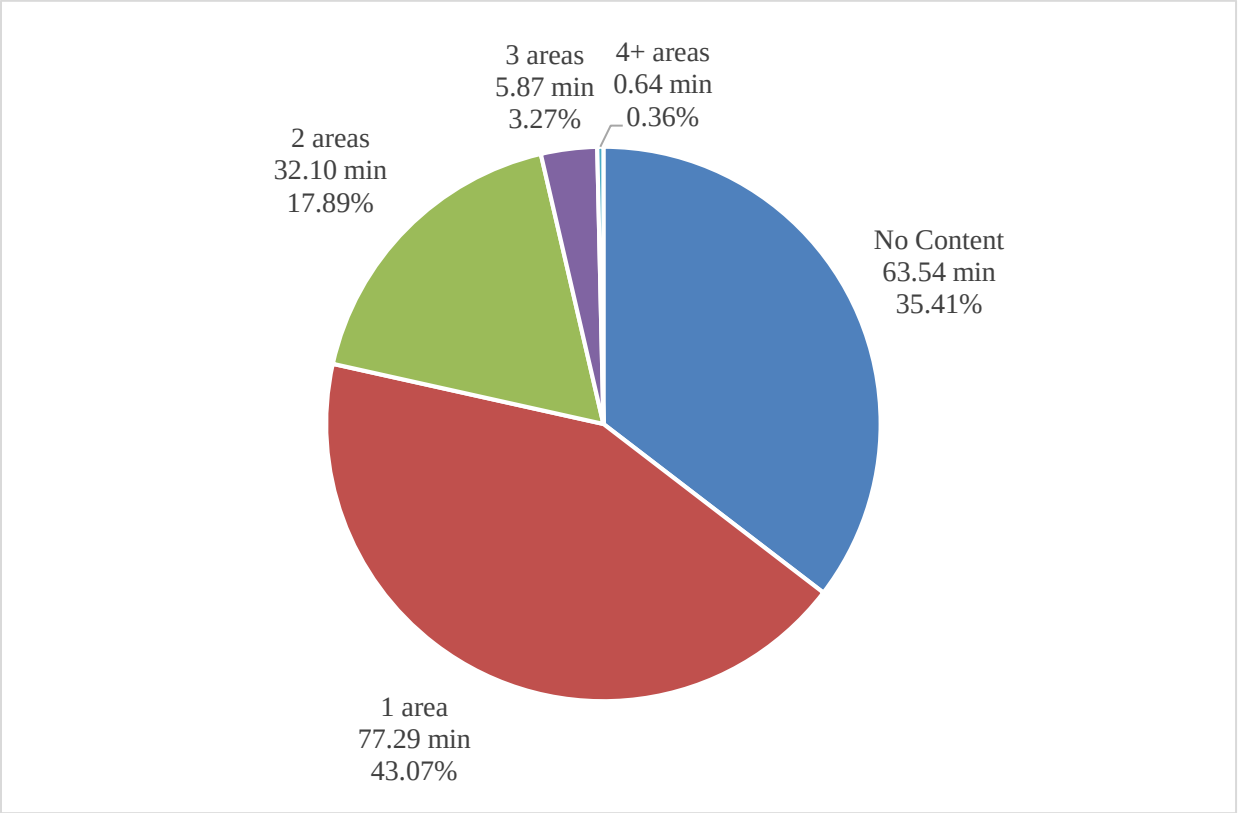


Table 7 reports the percentages of time of curriculum integration including no content, one content area, or more areas integrated across subgroup of interest. Subgroup patterns are aligned with overall results. About 24-39% of the day, there is no content. Another 38%-46% of the time, there no integration with one content area observed. Integration of two content areas varies between 17-27% of the time. Integration of three content areas is generally observed 3% of the time (about 6 minutes).

Table 7. Percentages of time spent in Curriculum Integration by Subgroup, n=129.

		% of Time				
		None	1	2	3	4 or more
STAR Ratings	1-2 (n=3)	23.51	38.08	28.71	8.97	0.00
	3 (n=69)	35.67	42.99	17.87	3.18	0.00
	4 (n=57)	35.73	43.44	17.35	3.09	0.00
Lead Teacher Credential	CDA/ECE (n=6)	33.95	43.79	17.82	3.43	0.00
	AA (n=51)	37.53	42.16	16.89	3.24	0.00
	BA/BS (n=45)	36.55	41.99	17.69	3.35	0.00
	MA/ME (n=18)	29.71	46.06	20.72	3.05	0.00
	Missing (n=9)	30.13	47.23	18.94	3.39	0.00
PHLpreK Partner Agency	Phila SD (n=15)	33.09	41.21	20.79	4.39	0.00
	PHMC (n=69)	34.85	44.05	17.66	3.14	0.00
	UAC (n=45)	37.05	42.19	17.28	3.10	0.00
Curriculum	Creative (n=89)	34.66	43.58	17.83	3.48	0.00
	Creative + (n=30)	36.39	41.75	18.59	3.06	0.00
	Mother Goose (n=9)	38.71	42.23	16.96	2.09	0.00
	Other (n=1)	42.78	44.44	11.11	1.67	0.00
Success By 6	Yes (n=5)	25.56	42.29	26.60	5.22	0.00
	No (n=124)	35.81	43.10	17.54	3.19	0.00
New Site	Yes (n=9)	35.69	41.68	18.59	3.86	0.00
	No (n=120)	35.39	43.18	17.84	3.23	0.00

EduSnap also assesses time spent on two student learning approaches: *collaboration* and *metacognition*. Average results are shown in Figure 10. The collaboration code accounts for children working together in activities where they are sharing ideas, completing assignments, or problem-solving. The metacognition code includes children being encouraged to provide evidence or reflect on their ideas or explaining their answers. Collaboration was observed about 14% of the time (and increased from 12% in 2018), and metacognition remains lacking.

Figure 10. Percentages of Time and Minutes Spent in Student Learning Approaches.

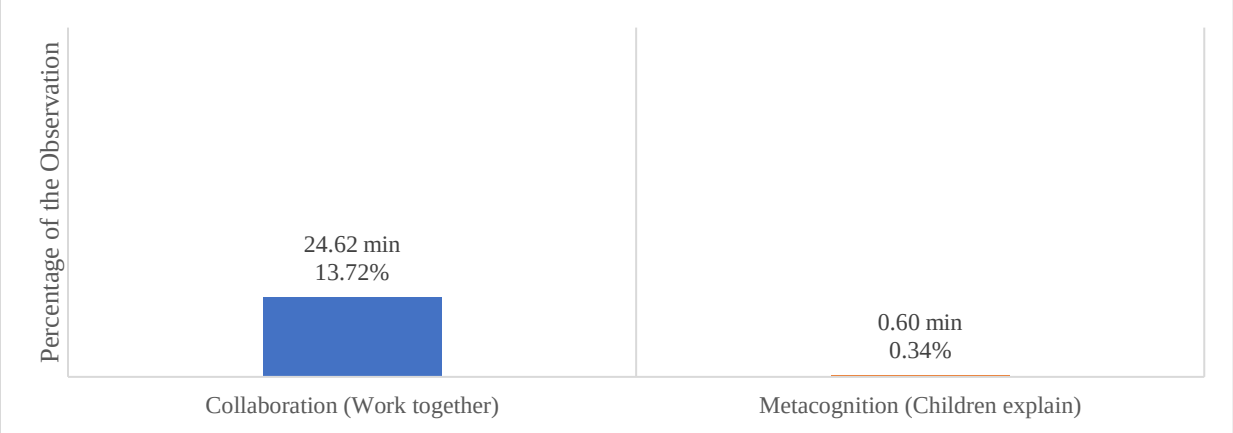


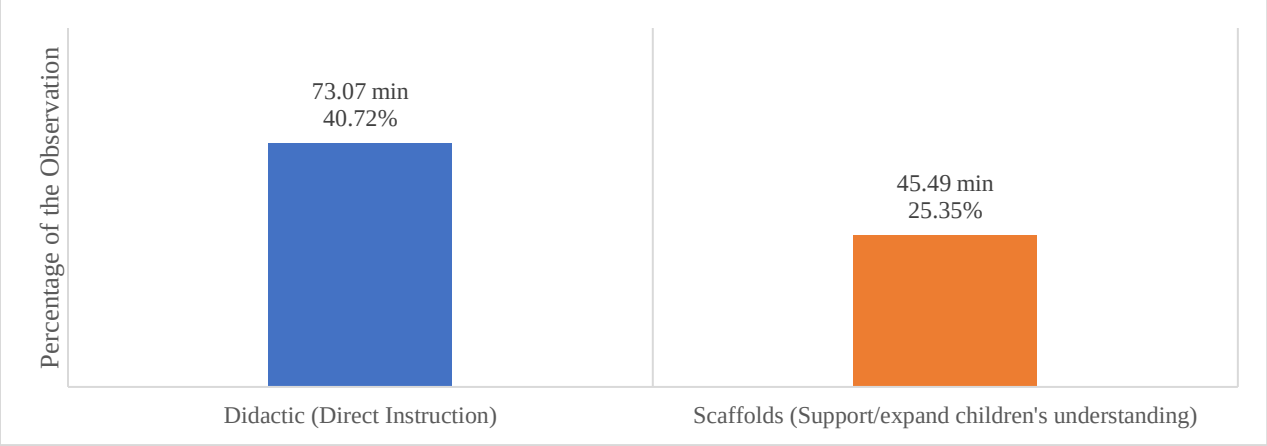
Table 8 reports average student learning approaches by type of setting. The lack of metacognitive approaches is consistent for all providers, star levels, teacher credentials, etc. Collaborative approaches vary between 11% and 17%. These are higher classrooms in classrooms with teachers with an MA/ME and 4-star-rated classrooms.

Table 8. Percentages of time spent in student learning approaches by subgroups, n=129.

		Student Learning Approaches	
		Collaboration	Metacognition
STAR Ratings	1-2 (n=3)	11.22	0.36
	3 (n=69)	13.25	0.44
	4 (n=57)	14.43	0.20
Lead Teacher Credential	CDA/ECE (n=6)	12.43	0.28
	AA (n=51)	12.56	0.26
	BA/BS (n=45)	12.88	0.34
	MA/ME (n=18)	17.30	0.71
	Missing (n=9)	18.25	0.00
PHLpreK Partner Agency	Phila SD (n=15)	12.96	0.37
	PHMC (n=69)	14.08	0.31
	UAC (n=45)	13.43	0.36
Curriculum	Creative (n=89)	14.82	0.31
	Creative + (n=30)	11.37	0.37
	Mother Goose (n=9)	10.87	0.31
	Other (n=1)	12.78	1.67
Success By 6	Yes (n=5)	22.60	0.44
	No (n=124)	13.37	0.33
New Site	Yes (n=9)	13.57	0.06
	No (n=120)	13.73	0.36

Finally, the EduSnap also provides an opportunity to assess the percentage of time children are exposed to two teaching and learning approaches: *didactic* and *scaffolded* teaching. These are illustrated in Figure 11. Classrooms in the sample continue to depend more on didactic approaches, and this has been the case for all three years assessed. The percentage of observed didactic approaches (42%) stayed relatively the same relative to 2018 (when it had shown an increase). The percentage of use of scaffolds, however, decreased back to the levels observed in 2017 (25%).

Figure 11. Percentages of Time and Minutes Spent in Teaching Approaches.



The percentage of time spent in didactic or scaffolded approaches are shown in Table 9 for the subgroup of interest. Classrooms with teachers with an AA appear to depend more on didactic approaches. PHMC supported classrooms evidence slightly higher scaffolds. Partners that have been with PHLpreK before this year evidenced higher scaffolds.

Table 9. Percentages of time spent in Teaching Approaches by Subgroup, n=129.

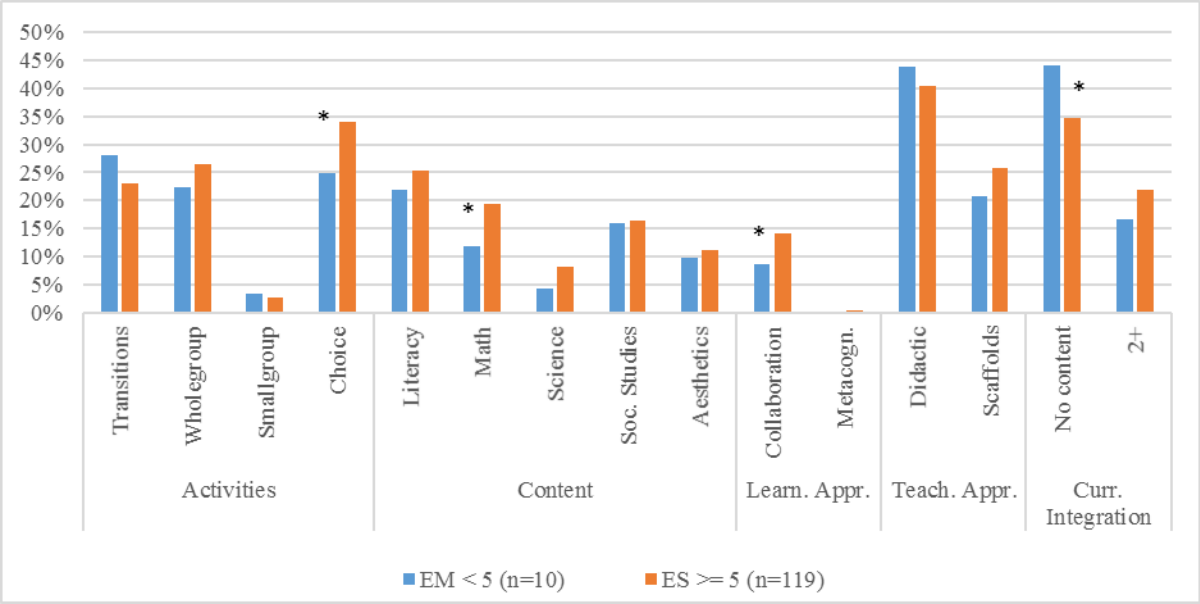
		Teaching and Learning approaches	
		Didactic	Scaffolds
STAR Ratings	1-2 (n=3)	39.21	44.21
	3 (n=69)	44.44	24.58
	4 (n=57)	36.30	25.29
Lead Teacher Credential	CDA/ECE (n=6)	39.45	27.18
	AA (n=51)	42.85	23.95
	BA/BS (n=45)	40.60	26.82
	MA/ME (n=18)	38.54	24.75
	Missing (n=9)	34.47	25.91
PHLpreK Partner Agency	Phila SD (n=15)	43.80	23.22
	PHMC (n=69)	39.26	25.70
	UAC (n=45)	41.93	25.52
Curriculum	Creative (n=89)	40.67	24.17
	Creative + (n=30)	42.66	26.48
	Mother Goose (n=9)	33.88	31.94
	Other (n=1)	48.33	37.22
Success By 6	Yes (n=5)	42.20	26.23
	No (n=124)	40.66	25.31
New Site	Yes (n=9)	47.05	19.01
	No (n=120)	40.25	25.82

2. What do high-quality classrooms have in common?

Given the information on the quality of classroom processes collected by the CLASS and the information of the time spent across settings, content, and teaching and learning approaches, it is possible to cross these two sources to assess patterns for the PHLpreK classrooms. This allows understanding whether the balance in activities, content, integration in curriculum, and teaching and learning in higher-quality classrooms differs relative to lower quality classrooms within the PHLpreK program.

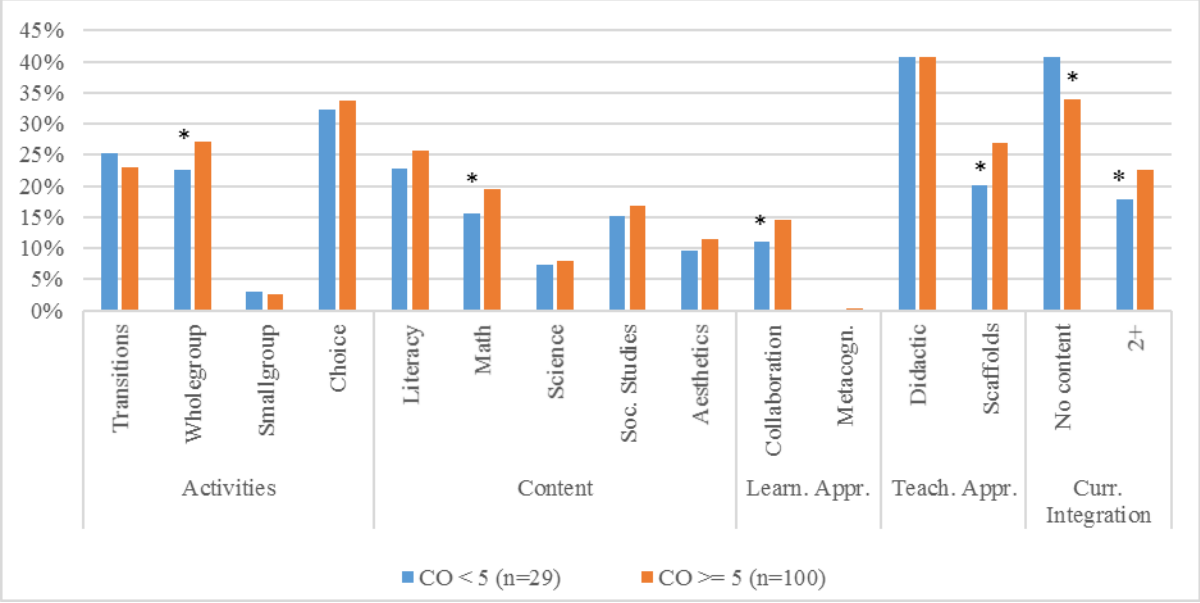
Average EduSnap percentages for lower versus higher scoring classrooms in Emotional Support are shown in Figure 12. We use a threshold of 5 to define lower (under 5) and higher (above or equal to 5) ES quality. Statistically significant differences include an asterisk. Higher ES quality classrooms were observed having higher percentages of choice (16 minutes), math (13 minutes) and collaboration (10 minutes) and lower percentages of time without content (17 minutes less).

Figure 12. EduSnap average percentages grouped by CLASS Emotional Support levels



For the CLASS Classroom Organization Domain (Figure 13), we also lower versus higher scoring classrooms at the threshold of 5. Statistically significant differences are highlighted with an asterisk. Higher CO quality classrooms are found to spend more time in whole group (8 minutes), more time in math (7 minutes), more time in collaborative activities (6 minutes), more time in scaffolded approaches (12 minutes), less time without any content (12 minutes), and more integration (9 minutes).

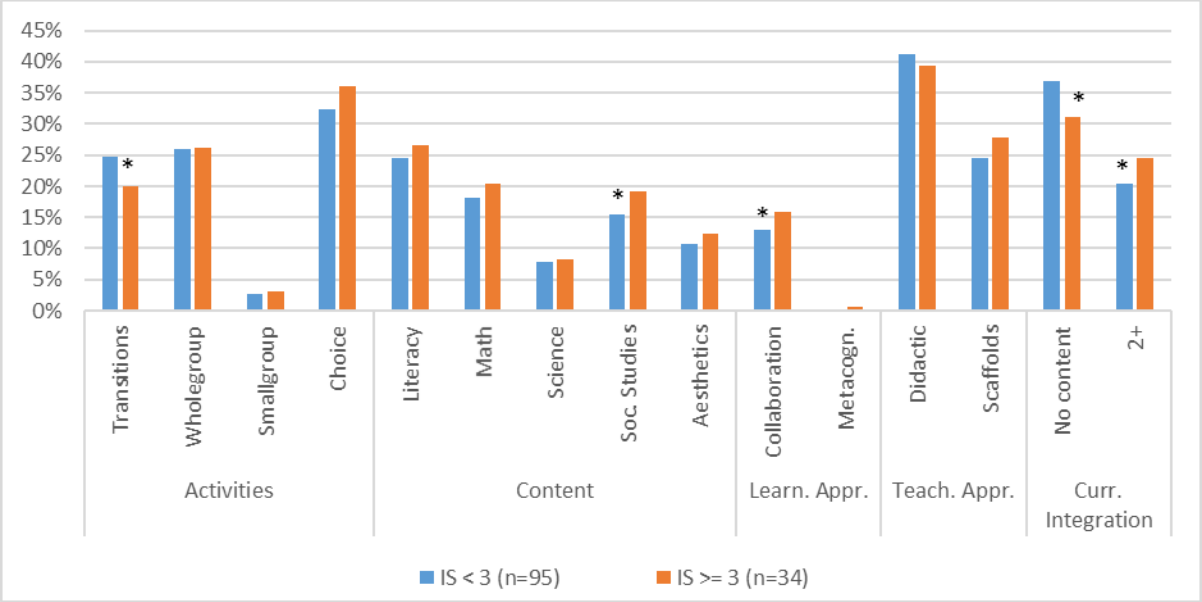
Figure 13. EduSnap average percentages grouped by CLASS Classroom Organization levels



Finally, we repeat this exercise for the Instructional Support domain, with the threshold defined at 3. Higher IS quality classrooms have on average lower transition times (9 minutes),

more time on social studies (6 minutes), more collaborative approaches (5 more minutes), less time spent without any content (10 minutes), and more integration (7 minutes on average).

Figure 14. EduSnap average percentages grouped by CLASS Instructional support levels



The findings above are a way to assess the associations between the different groupings and contents and activities captured in the EduSnap and the dimensions in CLASS. We also looked at the correlations between CLASS and EduSnap. These show statistically significant (albeit moderate) negative correlations between transitions and CLASS CO, positive ones between literacy and CLASS ES and CO, between time spent in math and all three dimensions, time spent in social studies and CLASS CO and IS, and between time spent in aesthetics and all three dimensions. This was also the case for time spent in collaborative activities and CLASS CO and ES. Time without any content was negatively correlated with all CLASS domains. Scaffolded approaches and integrated content were correlated with all CLASS domains as well.

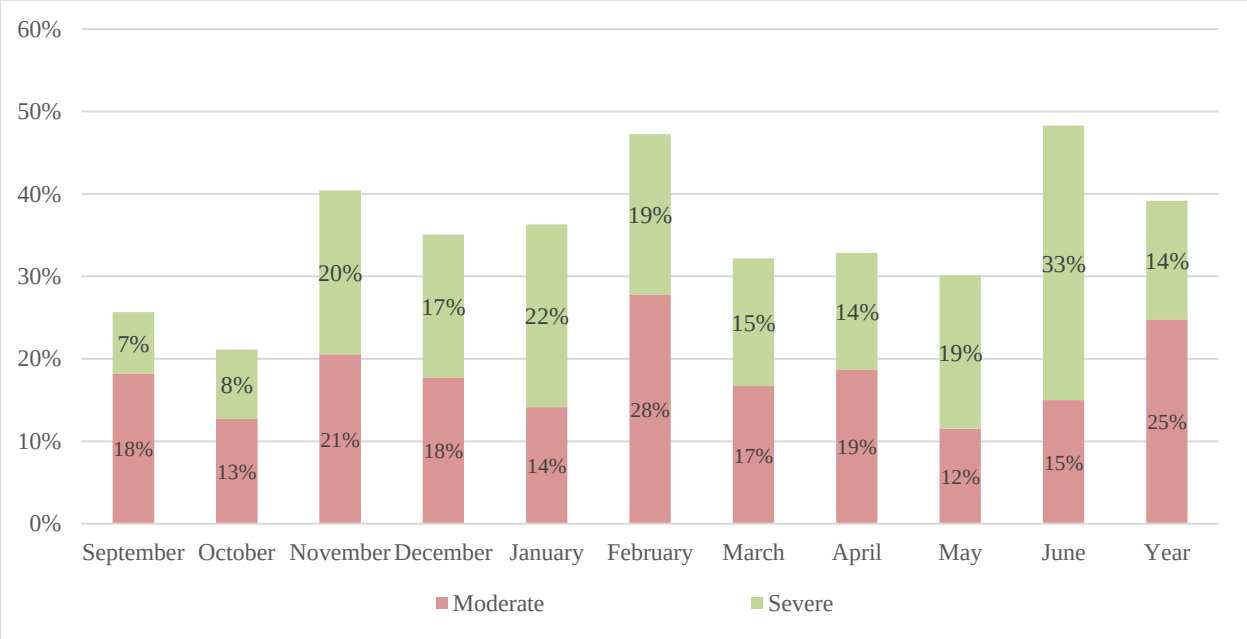
3. Children's absenteeism and teacher turnover in the PHLpreK program, 2018–2019

Absenteeism for the 2018-19 school year is reported in Figure 15 for children in our sample.[7] This differentiates between moderate (10-19.9% of the days) to severe (over 20% of the days) chronic absenteeism. We analyze absenteeism by month and for the whole academic year. Average daily attendance rate was 90% throughout the year (see Appendix Table B.1). However, chronic absenteeism varied by month, particularly increasing in the months of November, January and June. Chronic absenteeism is above 20% generally, most months it averaged above 30% and in critical months it increased up to 40% and more. Chronic absenteeism for the year was at 39%. In Newark (NJ) preschool chronic absenteeism in 2017-18 was 48% (53% for Black children and 44% for Latinx children; ACNJ, 2019). Chronic absenteeism in Chicago in

[7] Based on excused and unexcused absences versus expected days of attendance.

preschool were 45% of three-year-olds and 36% of four-year-olds in Chicago (23-57% for Black children and 30-35% for Latinx children: Ehrlich, Gwynne, & Allensworth, 2018). These rates are larger to what has been found for other large cities nationally (22% in Baltimore: Connolly & Olson, 2012; 27% in Washington, DC: Dubay & Holla, 2016).

Figure 15. Absenteeism by month and year, academic calendar.



Total teacher turnover as measured between September 2017 and February 2018. This turnover was reported upon observation of the research team in the center, and therefore likely underestimates turnover. That is, it does not include all possible turnover (between the time the assessors did not visit the classroom, and through the summer, or before October). Average turnover observed was of 18% (compared to 27% observed the previous year). Turnover was higher among teachers with lower or unknown qualifications, some agencies, and in sites that were not supported by Success by 6 (See appendix Table B.2). Turnover varied between 7% and 31% depending on the agency.

4. Children's gains in the PHLpreK program, 2018–2019

This evaluation measured gains child outcomes in receptive vocabulary (using the Peabody Picture Vocabulary Test), literacy (using the Woodcock-Johnson Tests of Achievement Letter-Word subtest), and math (using the Woodcock-Johnson Tests of Achievement Applied Problems subtest). Moreover, it evaluated executive functioning (EF) using two measures: the Dimensional Change Card Sort Game (DCCS) and the Peg Tapping task (PT). Socio-emotional development was measured with the ASEBA teacher reported form (C-TRF).

Child gains for the 2018–19 school year for the overall sample and for selected subgroups of interest are shown below and reported in detail in Appendix B. Included are only scores for children assessed in both fall and spring of the school year. Figures 16-18 report gains in standardized scores for the PPVT (vocabulary) and Woodcock-Johnson (literacy and math)

assessments which allow comparing results for children in the program in relation to growth due to age maturation (that is, in relation to growth due to children's natural average growth). These measures are standardized at the mean score of 100 and with a standard deviation of 15. Positive gains in standard scores point to gains that are larger than those of other children after adjusting for age.

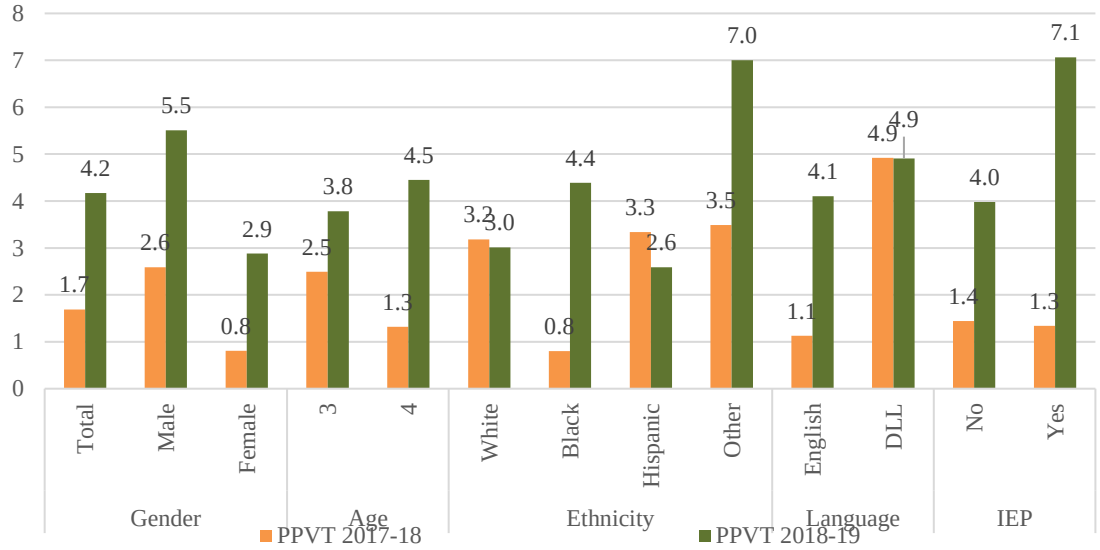
Children started on average at a lower level than the cohort assessed in 2017-18 (See appendix B) in the PPVT, and WJ LW and AP. However, they also grew more and as a consequence average spring scores do not differ much between the two cohorts. Children's standard scores increased on all three measures in relation to the norm and in relation to the 2017-18 cohort. Having said this, there was also more variance (children having more differences within the sample in terms of gains) this year for LW ID and AP measures. Other trends observed are: (a) larger receptive vocabulary gains between fall and spring for 4-year-olds, males, children identified as Black, dual language learners (DLLs) and children with an IEP; (b) larger literacy gains for females and children identified as White; (c) larger math gains for males, 4-year-olds, Hispanics and dual-language learners (DLLs). PPVT scores were lower this year for Hispanic children than in the 2017-18 cohort of children. For LW identification and applied problems, children regardless of gender, ethnicity, age, DLL status or IEP status, gained more as a group.

As comparison, it is useful to assess gains for lower income and minority children in other evaluations of preschool programs. For example, one-year gains for children in this year's sample was of 4.2 standard points on the PPVT, which are larger gains than those reported for 3 and 4-year-olds in the FACES study on Head Start (Table B.5a; Aikens, Klein, Tarullo, & West, 2013 and Table B.4a; Aikens, et al., 2017).[8] One-year gains in LW identification were of 0.56 standard points, quite smaller than one-year gains for in the FACES (although for the WJ-III). Lastly, gains in applied problems were on average 4.4 standard points, twice the one-year gains in FACES (for the WJ-III). Similar to PHLpreK children, Head Start children in the FACES study also scored well below average before and after a year in the program (Table B.4a; Aikens, et al., 2017).[9]

[8] FACES is The Head Start Family and Child Experiences Survey. This is an ongoing national longitudinal study of the cognitive, social, emotional, and physical development of Head Start children. The 2014-15 cohort of FACES evidence lower gains in these three measures than the 2009-10 cohort which. The 2014-15 gained 1.6 in the PPVT, 2.1 in letter word and 2.0 in applied problems. The 2014-15 gains for Head Start children were 3.4 in the PPVT, 5.8 standard points in LW ID and 2.0 in applied problems.

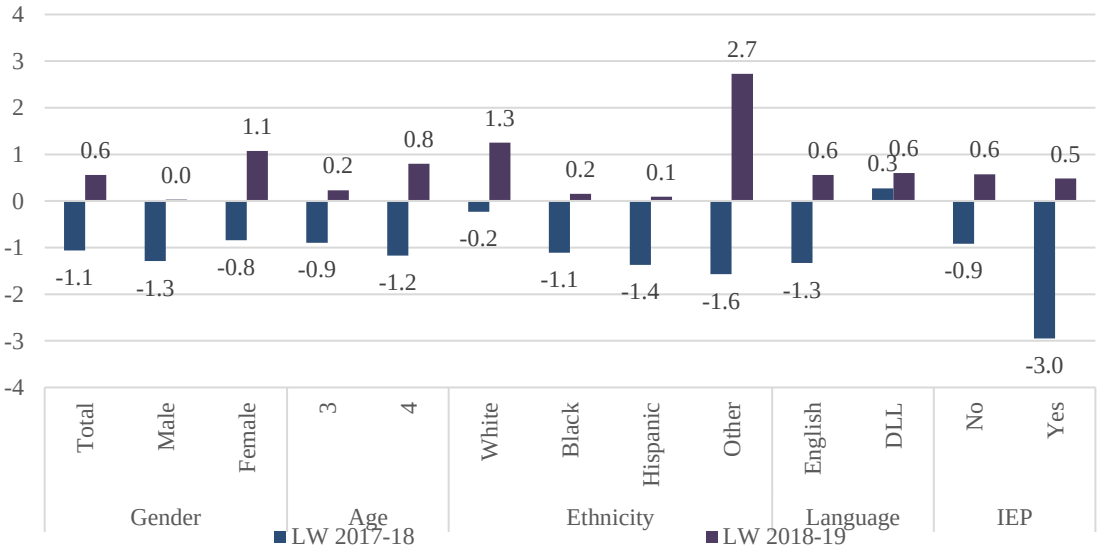
[9] Head Start children in the FACES study were 27% White, 23% African American and 42% Hispanic. A total of 40% spoke a language other than English at home. Almost a third of mothers of children in the 2014 study had less than a high school diploma (26%), another third reported a GED (33%) and another third some college or a vocational/technical degree (33%). Only 8% reported a Bachelor's or higher degree. Over 90% of the sample reported incomes below 200% of the federal poverty threshold. PPVT average fall scores for children in the FACES were 89.9 standard points. Slightly lower than the average in the PHLpreK sample.

Figure 16. Standard score gains for children in the PPVT, 2017-18 and 2018-19 cohorts



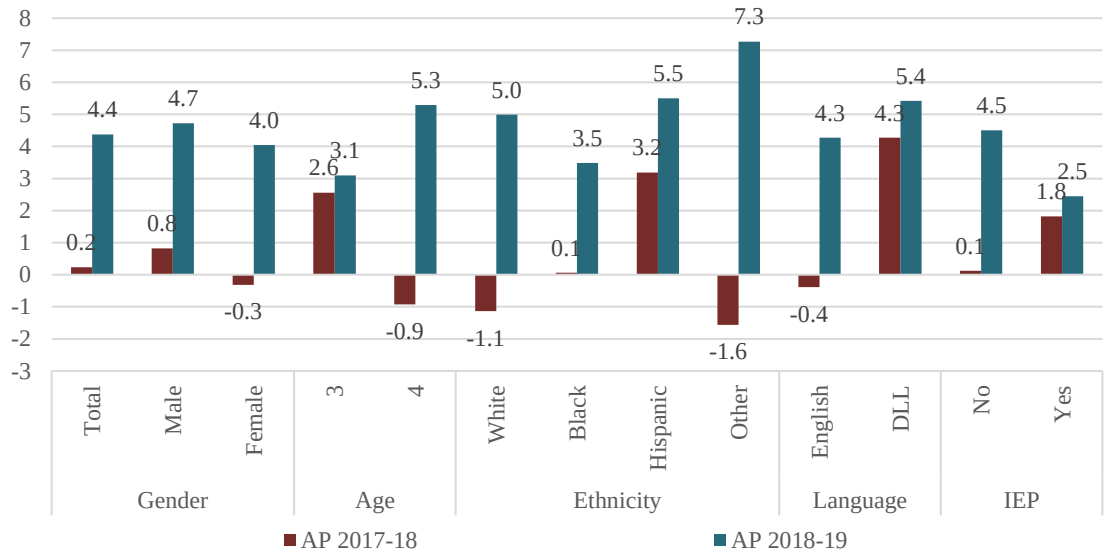
Note: For 2017-18 n= 465 for the PPVT; for 2018-19 n= 585.

Figure 17. Standard score gains for children in the WJ LW identification, 2017-18 and 2018-19 cohorts



Note: For 2017-18 n= 464 for the letter word identification subtest; for 2018-19 n= 585.

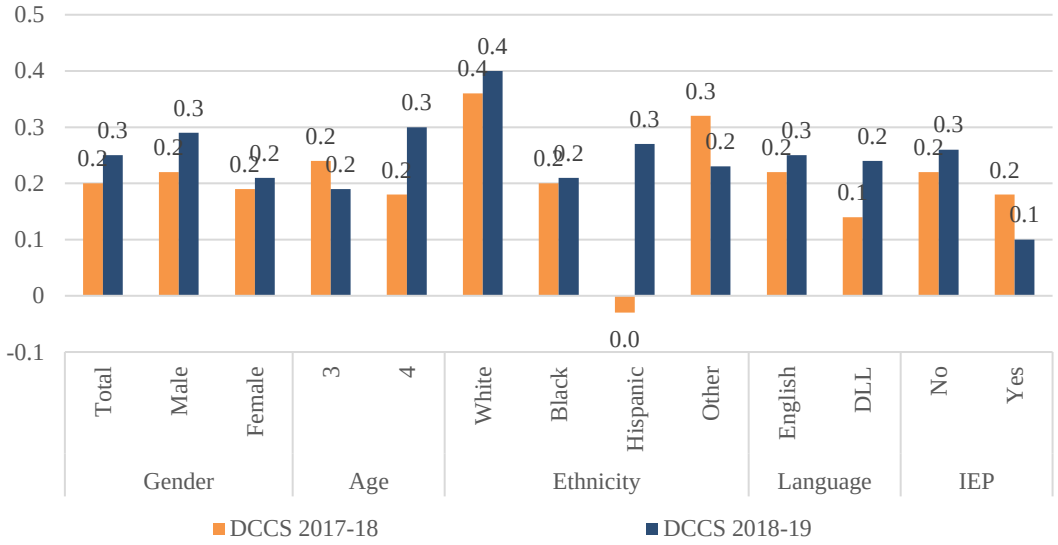
Figure 18. Standard score gains for children in the WJ applied problems, 2017-18 and 2018-19 cohorts



Note: For 2017-18 n= 462 for the applied problems subtest; for 2018-19 n= 585.

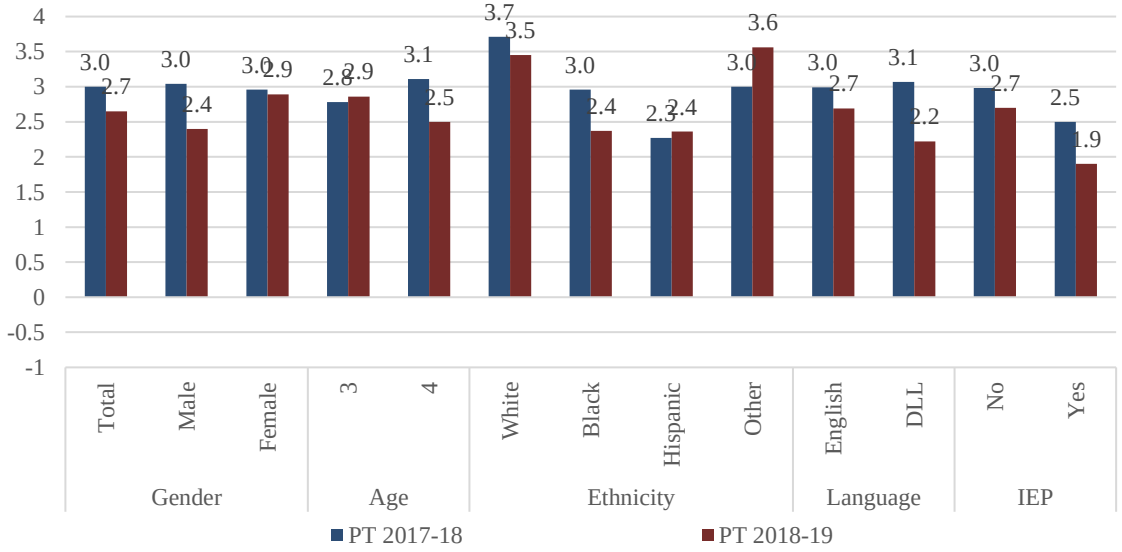
Figures 19-21 show gains in DCCS and Peg Tapping (executive function) and changes in the C-TRF (socio-emotional: which is reverse scored and therefore positive increases are not to be considered gains). As reference, the Learning-Related Cognitive Self-Regulation School Readiness Measures for Preschool Children Study (aka the Self-Regulation Measurement Study) (Meador, et al., 2013) reports average DCCS scores of 1.42 at 51–53 months of age and 1.62 at 57–59 months. This is an average difference of 0.20 between these two ages. Children gained in executive functions (DCCS and PT) at a higher level as children in the self-regulation study, with overall gains being 0.25. Stronger gains observed in children identified as White (0.40) smaller gains in children with an IEP (0.10). The self-regulation study also shows PT average scores of 6.02 at 51–53 months and 8.80 at 57–59 months, with a difference of 2.78. PHLpreK children gained at a similar level, with gains of 2.65 found for this preschool year. Other studies in Seattle and Boston have found similar results for children (Nores, et al., 2018; Weiland & Yoshikawa, 2013; Weiland, et al., 2013).

Figure 19. DCCS gains in children.



Note: n= 465 for 2017-18 and n=585 fir 2018-19.

Figure 20. Peg-Tapping gains in children.

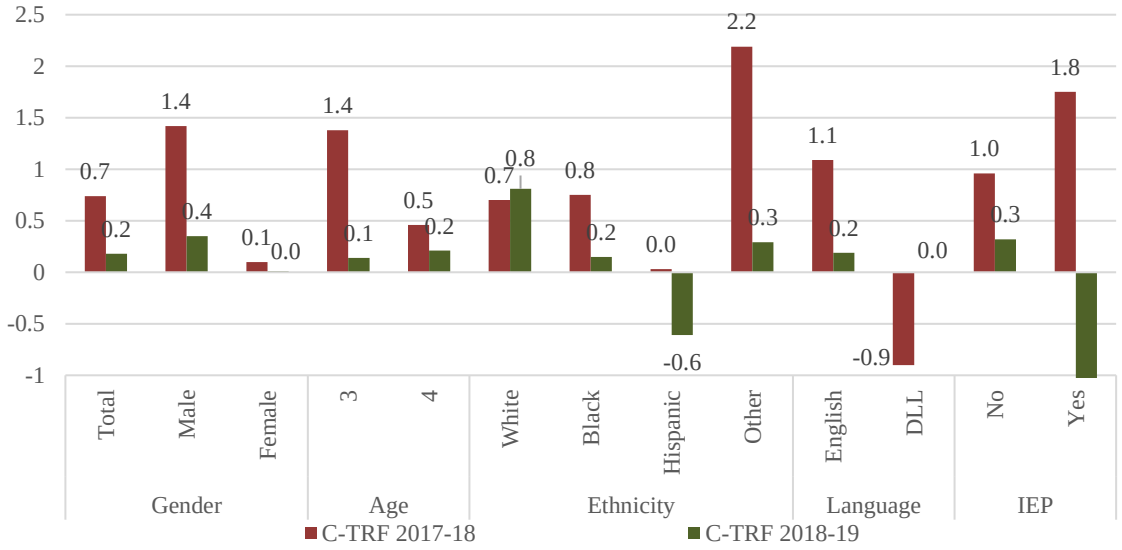


Note: n= 465 for 2017-18 and n=585 for 2018-19.

In relation to children's socio-emotional development, changes were quite small in relation to standards. T scores reported reflect how a child's score on each scale compare with the scores of the normative sample of peers. The incidence of socio-emotional problems increased between the fall and spring. However, these decreased for dual language learners and for children with an IEP. There were improvements (0.10) in average internalizing behaviors in relation to the norming sample (Appendix B). Improvements in relation to the norm were higher for children identified as African American, Hispanic and DLL. This was also the case for

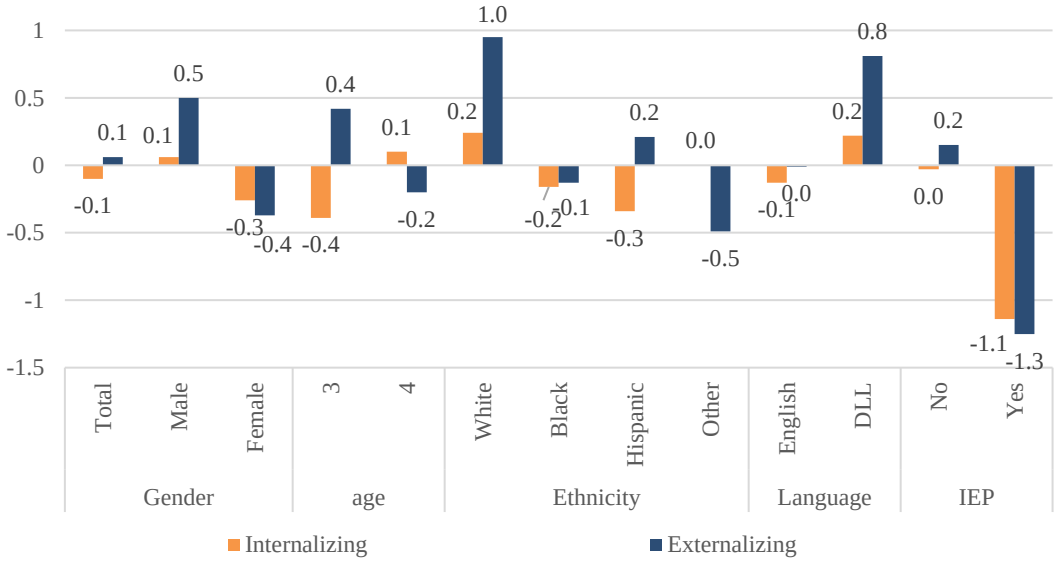
Females and children with an IEP. Similar improvements in externalizing behaviors were observed only for the latter two subgroups of children.

Figure 21. C-TRF changes in children (socio-emotional problems)



Note: n= 452 for 2017-18 and n=542 for 2018-19. For the C-TRF, negatives indicate gains.

Figure 22. C-TRF changes in internalizing and externalizing behavior.



Descriptive analyses of developmental gains for children's do not take into account the intersectionality of varied inter-relationships of social identities and interacting social processes that compound in the production of inequities (Becares & Priest, 2015). Estimations that account for varied socio-demographic identities allow understanding inequities between groups accounting for inter-group differences. Therefore, we next examine the association between children's learning end of year outcomes, their various demographic characteristics, and program features, as well as teacher qualifications using multi-level estimations. We include information

on children's start of year outcome, gender, race and ethnicity, home language, and IEP (we do not have information on children's family income levels). Program features for PHLpreK include star ratings, curriculum, teacher qualifications and classroom quality. The analyses also consider that scores of children who are in classrooms together cannot be considered to be independent of each other (that is, clustering of children within classrooms).

We present analyses including the CLASS and the EduSnap. Results are shown in the appendix C and summarized here. Table C.1. shows these for levels of CLASS, and Table C.2 includes the thresholds for quality of CLASS defined earlier. Tables C.3 and C.4 report results for raw scores. Multivariate analyses account for how children are grouped, their background and their preschool experience. That is, this allows understanding how children's gains differ among children, and what aspects of centers and teaching and learning, contribute to those gains.

Estimations show that female's and male's gains generally did not differ across outcomes measured. Children identified as African American evidenced lower spring scores in literacy, math and executive functions relative to their White peers, after controlling for their fall scores and other child and classroom aspects. That is, despite their larger overall gains graphed earlier, differences remain by the end of the year relative to White peers. Children identified as Hispanic or children with IEPs evidence a similar growth pattern, and by the end of the year initial differences relative to their White peers were generally not statistically significant. Child absenteeism rates, whether moderate or severe is also included. Child absenteeism, particularly severe rates of absenteeism, relate negatively to literacy and math gains although the opposite is true in relation to teacher's reported socio-emotional problems.

In relation to center and classroom characteristics, estimations show that children in 3- and 4- star rated programs performed higher in receptive vocabulary, literacy and math, than their peers in lower rated programs. This differed from last year's findings, where improvements in children did not align with star ratings. Improvements in quality for higher rated classrooms could be driving this result. This year, the number and percentage of centers rated 3 or 4 increased relative to the previous year and more classrooms were brought into these levels, and it is plausible that classrooms still under a 3-rating remain low quality and are statistically different. Curriculum this year also showed no systematic effect across the different child measures included in the evaluation.

CLASS CO evidences a positive association with receptive vocabulary (somewhat consistent with the findings for 2017-18), and even stronger so for CLASS CO above 5.5. Also, classrooms with emotional supports (ES) above 5.5 show positive associations with children's receptive vocabulary, literacy and math. CLASS IS levels above 3.5 were found to be positively related with receptive vocabulary and math outcomes. In relation to the EduSnap and within the boundaries of what was found in PHLpreK classrooms, estimations show that time spent in letter word identification and aesthetics, contribute positively to different areas of child development. On the flipside, it appears there are negative associations between small group activities and children's learning. The latter point towards the importance of intentionality in the use of activity settings.

The main patterns that emerge from the multivariate estimations are: (a) differences by race. African Americans remained at end of the school year however inequities appear to have improved relative to the previous year; (b) star levels this year show associations with children's gains; (c) time spent in small group negatively relate to some developmental outcomes, likely due to quality needed in these activity settings; (d) time spent in content areas matters for different children's domains; (e) higher levels of CLASS CO is related to stronger child

performance, (f) CLASS ES above 5.5 makes a differences for child outcomes; and (g) not much is observed for CLASS IS, unless levels are above 3.5 where it does relate to receptive unconstrained skills.

Overall, findings point that increases in quality among providers which has also moved them up in terms of star-levels appear to matter for children's gains. In addition, there are important findings that would indicate the importance of supporting classrooms to increase in quality across all CLASS domains, and even above the thresholds tested. Results would also point towards the importance of intentionality in integrating content, play and project-based experiences into small group settings.

5. Program supports: teacher and director perceptions

As part of the research, the last two years, we have included teacher and director surveys to support PHLpreK by assessing to what extent providers felt they were effectively served by existing program supports. Program supports could be those of PHLpreK or those provided by other government agencies. The surveys focused on capturing lead teacher's experience in early childhood, and their perceptions on professional development and technical assistance received in the last year. Responses were optional, although incentives were provided to teachers. As a consequence, the number of respondents may vary across questions. A total of 117 of 147 teachers and 73 of 82 directors completed the survey. If a director was also a lead teacher, it was indicated that they complete a director survey. Data was collected from the lead teacher. All lead teachers were surveyed, regardless of there full time or part time status, which should be taken into account when interpreting their responses.[10]

Teachers

On average, teachers reported being about 38 years of age (ranging between 22 and 69 years of age) with an average of 12 years of experience in early childhood. Teachers reported four years in the current program, and three years in the current classroom. These rates are aligned with those reported in the Spring of 2018. There seems to be alignment between the student and the teacher ethnic and racial composition. Of teachers that answered the question on race or ethnicity, 13% reported identifying as White, 67% as African American, and as 14% Hispanic. About 15% reported speaking Spanish. Of responding teachers, 61% reported a salary of under 20,000 to 40,000 a year. A large percentage chose not to share their salary (36%).

Table 10. Teacher's Annual Salary

Salary	Frequency	Percent
$20,000 or less	9	8.11
$20,001-$30,000	31	27.93
$30,001-$40,000	28	25.23
$40,001 or more	3	2.70
Do not wish to share	40	36.04
Total	**111**	**100.00**

[10] All PHLpreK providers are expected to meet the City living wage which in FY19 was $12.10.

The survey included questions on teachers' participation in professional development activities (PD). For this purpose, PD was defined as training and assistance for individual growth. The majority of teachers (79%) reported three or more in-service training days. Teachers mostly reported PD in study groups, direct instruction (outside consultant), outside training, follow-up supports for teachers trying out new skills, peer observation/feedback and coaching/consultation (Table 11). Teachers were also asked about the specific number of PD workshops attended, and 69% reported attending between one and six (Table 12).

Table 11. Responses to: In which of the following staff development and training activities have you participated since January 2018? (n=117)

Professional Development	Percent
Three or more in-service training days (training delivered at my program, by program leadership)	79.49
Workshops involving study groups or small-group problem solving	70.09
Direct instruction from an outside consultant on a specific topic	68.38
Peer observation and feedback	62.39
Follow-up support for a teacher trying out new skills and knowledge in the classroom	66.67
Visits to, or observations of, other schools	43.59
Release time for attending early childhood professional conferences	49.57
Enrollment in college or university courses	43.59
Workshops on computers and technology in the classroom	41.88
Training outside of my program, with participants from other programs	62.39
PD program that uses coaching/consultation	65.81
Other	7.69

Table 12. Responses to: How many professional development workshops do you recall attending since January 2018?

Workshops	Frequency	Percent
1-3	33	35.11
4-6	32	34.04
7-9	12	12.77
10 or more	17	18.09
Total	**94**	**100.00**

Teachers were asked about the degree to which the PD they attended has been beneficial to support their work with preschool children. There was a general agreement about how beneficial these workshops were. Teachers reported general child development, classroom quality and various curriculum trainings as well as arranging classroom space and learning materials, child assessments and health and safety as the most beneficial (Table 13). Family engagement and behavioral supports for children were also highly regarded.

Table 13. Responses to: Mark all the broad topic areas that were covered in professional development workshops that you attended since January 2018?

Professional Development Workshops	Not Beneficial or Not Too Beneficial	Adequate	Somewhat Beneficial or Highly Beneficial	N/A
General Child Development	1.85	10.19	79.63	8.33
Supporting English Language Learners (ELLs)	6.06	13.13	54.55	26.26
Creative Curriculum	4.72	13.21	70.75	11.32
Other Curriculum	9.38	15.63	75.00	0.00
Family Engagement/Partnership	2.91	9.71	71.84	15.53
Classroom Quality	3.81	6.67	77.14	12.38
Arranging Classroom Space and Learning Materials	2.94	6.86	77.45	12.75
Child Assessment	2.75	5.50	81.65	10.09
Nutrition	3.85	9.62	66.35	20.19
Information on the Philadelphia Nutrition Standards	5.94	11.88	56.44	25.74
Health and Safety	1.85	5.56	83.33	9.26
Early Childhood Mental Health/Social Emotional Development	5.71	7.62	72.38	14.29
Supporting Children with Challenging Behaviors (PBIS)	3.81	10.48	71.43	14.29
Other	1.92	7.69	40.38	50.00

Note: Response rates varied between 84% to 93% depending on the workshop.

The survey also inquired into teachers' participation in a PD program that uses coaching or consultation (e.g., My Teaching Partner, Practice-Based Coaching, etc.). Of responding teachers, 52% responded that they did participate in these types of programs (out of 113 respondents). About 40% reported that they participated in a PD program that uses coaching or consultation for less than a day (and reported this positively impacted their practice). Teachers were also probed on what type of PD is most needed in view of the challenges/needs they face every day as pre-K teachers. Teachers mostly expressed a need for supports in the following areas: behavior management, classroom quality and socio-emotional and mental health supports (Table 15). Teachers also mentioned other areas to a lesser degree as are: curriculum, working with children with disabilities or ELLs and child assessment.

Table 14. Responses to: How many hours have you participated in a professional development program since January 2018 that uses coaching/consultation (e.g., My Teaching Partner, Practice-Based Coaching, other)? (n=51)

Hours	Frequency	Percent
Less than a day	21	41.18
1-3 days	16	31.37
4-7 days	2	3.92
More than 7 days	6	11.76
Not Sure	6	11.76
Total	**51**	**100.00**

Table 15. Responses to: What type(s) of professional development do you feel are most needed given the challenges or needs you face as a pre-K teacher?

Professional Development	Frequency	Percent
Behavior Management	47	31.76
Child Assessment	6	4.05
Classroom Quality (Management)	15	10.14
Curriculum	10	6.76
English Language Learners (ELLs)	6	4.05
Family Engagement	8	5.41
Mental Health	15	10.14
Social Emotional Development	17	11.49
Special Needs	10	6.76
Other	14	9.46
Total	**148**	**100.00**

Note: Response rates varied as teachers reported multiple types of professional development.

The survey also inquired into Technical Assistance (TA). For the purposes of the survey, TA was defined as training and assistance for programmatic growth. Teachers reported requesting at least half a day to three days of TA, with 73% receiving some TA. The majority of teacher perceived all types of TA to be beneficial or highly beneficial (Table 16 with the exception of TA in relation to ELLs and PHL nutritional standards.

Table 16. Response to: Mark all the technical assistance you have received and the degree to which it has been beneficial for your everyday work supporting children and development of preschool children.

Technical Assistance	Not Beneficial or Not Too Beneficial	Adequate	Somewhat Beneficial or Highly Beneficial	N/A
Child Development	3.26	15.22	67.39	14.13
Supporting English Language Learners (ELLs)	7.14	19.39	42.86	30.61
Creative Curriculum	8.65	9.62	66.35	15.38
Family Engagement/Partnership	6.00	19.00	55.00	20.00
Classroom Quality	4.90	17.65	62.75	14.71
Classroom Space and Learning Materials	5.00	16.00	62.00	17.00
Child Assessment	2.02	18.18	64.65	15.15
Nutrition	6.06	19.19	52.53	22.22
Information on the PHL Nutrition Standards	5.26	16.84	48.42	29.47
Health and Safety	4.85	10.68	70.87	13.59
Early Childhood Mental Health/Social Emotional Development	4.08	17.35	64.29	14.29
Supporting Children with Challenging Behaviors (PBIS)	7.22	20.62	56.70	15.46
Other	3.77	9.43	26.42	60.38

Note: Response rates varied between 78% to 88% depending on the technical assistance topic.

Teachers were also asked about having received classroom materials from different sources in the last year. Teachers reported having received classroom materials mostly from PHLpreK (76%), merit grants (8%), Success by 6 (6%), or other sources (10%).

Directors

Directors were provided with a similar survey. It asked about their demographics, as well as PD and TA opportunities. Directors are on average 44 years old, 16% White, 67% African American, 9% Hispanic, and only eight of the 73 directors that responded the survey spoke Spanish. For those who reported their credentials (valid cases 71.7%), a large majority report an AA degree (36%) or a BA (33%) with only a small fraction reporting a master degree (16%) or some college (10%) or CDA/ECE (5%). A total of 11% reported having a CDA and 20% a teaching certification from PA. They report on average nine years of experience as a director, and seven in the current program. Most directors that shared their annual income reported between these to be over 40 thousand a year.

Similar to teachers, the survey asked directors about their participation in PD activities. Directors reporting a variety of modalities. Directors reported percentages over 80 for in-services training days and coaching/consultation (Table 17). On average they mostly reported (40%) 4-6 days of PD (Table 18).

Table 17. Responses to: In which of the following staff development and training activities have you participated since January 2018? (n=73)

Professional Development	Percent
Three or more in-service training days (training delivered at my program, by program leadership)	86.30
Workshops involving study groups or small-group problem solving	71.23
Direct instruction from an outside consultant on a specific topic	78.08
Peer observation and feedback	78.08
Follow-up support for a teacher trying out new skills and knowledge in the classroom	79.45
Visits to, or observations of, other schools	52.05
Release time for attending early childhood professional conferences	63.01
Enrollment in college or university courses	53.42
Workshops on computers and technology in the classroom	38.36
Training outside of my program, with participants from other programs	79.45
PD program that uses coaching/consultation	80.82
Other	10.96

Table 18. Responses to: How many professional development workshops do you recall attending since January 2018? (n=73)

Workshops	Frequency	Percent
1-3	14	19.18
4-6	30	41.10
7-9	13	17.81
10 or more	16	21.92
Total	**73**	**100.00**

Directors reported the following being the most beneficial areas of PD: child development, family engagement, child assessment and mental/socio-emotional health (Table 19). In terms of child assessment, 90% (out of 72 respondents) reported having received a copy of the Ages and Stages Screening Tool.

Table 19. Responses to: Mark all the broad topic areas that were covered in professional development workshops that you attended since January 2018.

Professional Development Workshops	Not Beneficial or Not Too Beneficial	Adequate	Somewhat Beneficial or Highly Beneficial	N/A
Child Development	4.35	18.84	75.36	1.45
Supporting English Language Learners (ELLs)	4.41	20.59	45.59	29.41
Creative Curriculum	4.17	8.33	72.22	15.28
Other Curriculum	17.39	13.04	56.52	13.04
Family Engagement/Partnership	8.45	11.27	76.06	4.23
Classroom Quality	5.88	14.71	69.12	10.29
Classroom Space and Learning Materials	3.08	23.08	61.54	12.31
Child Assessment	1.45	15.94	79.71	2.90
Nutrition	4.48	17.91	70.15	7.46
Information on the Philadelphia Nutrition Standards	5.97	22.39	53.73	17.91
Kindergarten Transition	7.14	21.43	70.00	1.43
Business Practice	7.35	10.29	66.18	16.18
Supervision	5.71	12.86	68.57	12.86
Health and Safety	4.29	17.14	72.86	5.71
Early Childhood Mental Health/Social Emotional Development	4.29	12.86	77.14	5.71
Supporting Children with Challenging Behaviors (PBIS)	5.80	11.59	68.12	14.49
Other	3.03	12.12	27.27	57.58

Note: Response rates varied between 88% to 98% depending on the workshop.

Directors were also asked most needed PD given the challenges they face as pre-K directors. Directors mostly expressed interest in behavior management strategies and classroom quality. Other interests include, family engagement and curriculum.

Table 20. What type(s) of professional development do you feel are most needed given the challenges or needs you face as a pre-K director?

Professional Development	Frequency	Percent
Behavior Management	21	27.63
Child Assessment	2	2.63
Classroom Quality (Management)	21	27.63
Curriculum	8	10.53
Family Engagement	8	10.53
Mental Health	2	2.63
Social Emotional Development	6	7.89
Special Needs	3	3.95
Other	5	6.58
Total	**76**	**100.00**

Directors were also asked about TA opportunities. A total of 40% (out of 63 respondents) reported having received no TA. Directors were asked about their perceptions of the benefits of the TA received on different content (Table 21). There was general agreement on the benefits, with the exception of TA on ELLs and nutrition.

Table 21. Response to: Mark all the technical assistance you have received and the degree to which it has been beneficial for your everyday work supporting children and development of preschool children.

Technical Assistance	Not Beneficial or Not Too Beneficial	Adequate	Somewhat Beneficial or Highly Beneficial	N/A
Child Development	1.72	10.34	58.62	29.31
Supporting English Language Learners (ELLs)	5.36	12.50	30.36	51.79
Creative Curriculum	-	6.67	60.00	33.33
Family Engagement/Partnership	1.67	3.33	63.33	31.67
Classroom Quality	1.64	9.84	59.02	29.51
Classroom Space and Learning Materials	1.67	16.67	50.00	31.67
Child Assessment	5.08	8.47	62.71	23.73
Nutrition	3.39	13.56	45.76	37.29
Information on the Philadelphia Nutrition Standards	6.90	13.79	43.10	36.21
Kindergarten Transition	1.69	10.17	66.10	22.03
Business Practice	5.00	11.67	51.67	31.67
Supervision	3.17	9.52	55.56	31.75
Health and Safety	1.52	9.09	60.61	28.79
Early Childhood Mental Health/Social Emotional Development	1.67	13.33	51.67	33.33
Supporting Children with Challenging Behaviors (PBIS)	3.33	13.33	46.67	36.67
Other	5.56	5.56	25.00	63.89

Note: Response rates varied between 71% to 90% depending on the technical assistance topic.

Directors also reported PD on Business Practices and Administration as beneficial (58%). Business and Financial Strength Technical Assistance was beneficial for 46% directors (n=68). Directors reported that TA was provided by UAC 22%, PHMC 36%, and other sources 25%.

Table 22. Response to: Have you attended any of the following since January 2018? How beneficial were these?

Professional Development/ Technical Assistance	Not Beneficial or Not Too Beneficial	Adequate	Somewhat Beneficial or Highly Beneficial	N/A
Professional development on Business Practices and Administration	4.35	8.70	57.97	28.99
Have you used Business and Financial Strength Technical Assistance	2.94	7.35	45.59	44.12
Other Business training or technical assistance	4.92	3.28	40.98	50.82

Note: Response rates varied between 83% to 94% depending on the training/technical assistance.

Directors were also asked about having received classroom materials from different sources. They reported having received classroom materials mostly from PHLpreK (57%), Merit grants (23%), and others reported having received materials from Success by 6 (3%) or other sources (7%).

Discussion of Findings

This report summarizes the findings for the 2018-19 school year for Philadelphia's preschool program. The program has concluded its third year of operations and continues to grow since its inception through solidifying partnerships with providers across the city. The purpose of this component of the evaluation 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 maturation.

Pre-K classrooms in these programs are averaging high to moderate levels of quality as measured by the CLASS Emotional Support and Classroom Organization domains. The Instructional Support domain is still scoring low but has shown an important improvement in the last year. In summary, classrooms on average are nurturing and safe environments for children and are adequately structured and organized. The increase in Instructional Support is a positive signal and is a trend that could be built on. 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, opportunities to engage in problem-solving activities, and planning and production processes that incorporate and build upon children and their initiatives. The EduSnap provides some insight into plausible areas to improve quality as it relates to time use in the classroom.

The EduSnap observations captured classrooms being generally effective at implementing a variety of activity settings. Whole group and choice both allow for more content or scaffolded learning to occur and are both important and related to different CLASS domains. However, the EduSnap showed that a fourth of the day continues to be spent in transitions and a third without any content, and these two were found to be highly related, and in turn, they were

found to be related to instructional support. Children were found experiencing a stronger percentage of didactic interactions than scaffolded ones. Metacognitive processes continue to be entirely absent in classrooms. Analyses of children's gains also point towards issues of quality and a need to strengthen intentionality in the use of small group settings and choice.

We also assessed how children's gains differ among different subgroups of children and what aspects of centers and teaching and learning, contribute to reducing any differences among children. Findings show no differences by gender. Children in the 2018-19 cohort started the school year at a lower level on average than children in the 2017-18 cohort. However, they also gained more across most measures throughout the school year. In receptive vocabulary, children identified as African American or Hispanic evidence higher overall gains, which implied a reduction in their differences relative to their white peers, although part of these differences remained for African American children by the end of the school year. Larger gains relative to the previous year are also observed for dual language children (math) and children with an IEP (receptive vocabulary, literacy, math, and socio-emotional development) to a degree that by the spring, no statistically significant differences remain relative to their English speaking peers and children without an IEP, respectively.

Star levels this year did show associations with children's gains in receptive vocabulary, literacy and math. No systematic differences were observed for children's gains in relation to either curriculum or teacher qualifications. CLASS CO levels are related to children's receptive vocabulary and math, similarly to 2017-18 findings. ES, CO and IS levels above the thresholds tested in this report were found to significantly matter for children's cognitive outcomes.

Given the low percentage of time in which integration of content occurs, content appears to compete across areas. Time spent in word identification, numbers and aesthetics support different areas of cognitive development. This indicates that the time spent on content is likely essential for children's performance in related content areas. Integration would strengthen this by increasing content altogether.

Teachers seem to perceive a need for additional supports mostly around socio-emotional development, children's mental health, and behavior management. Directors also mirrored this request and acknowledged a need for further supports on classroom quality.

The 2018-19 increase in quality scores is important for the program to progress towards the quality needed for long-term impacts on children. Supports for teachers on classroom quality would ensure this trend persists in future years. Continuing to increase classroom quality in the program require particular focus around strengthening instructional supports (concept development, quality of feedback, language modeling, metacognition) across choice, small group, and group work activity settings. In addition, increasing the use of content (time spend on it and quality of content) would require efforts to create awareness on why a third of the time is spent without content, and increasing content with reductions on such time, as well as increasing overall content through stronger integration across areas.

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

Achenbach, T. M. (2009). The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.

Advocates for Children of New Jersey (ACNJ). (2019). Newark kids count 2019: A city profile of child well-being. Retrieved from https://acnj.org/downloads/2019_03_28_Newark_Kids_Count_2019.pdf

Aikens, N., A. Kopack Klein, E. Knas, J. Hartog, M. Manley, L. Malone, L. Tarullo, and S. Lukashanets. (2017) Child and Family Outcomes During the Head Start Year: FACES 2014–2015 Data Tables and Study Design. OPRE Report 2017-100. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

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., & 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).

Bécares, L., & Priest, N. (2015). Understanding the influence of race/ethnicity, gender, and class on inequalities in academic and non-academic outcomes among eighth-grade students: Findings from an intersectionality approach. *PLOS One*, 10(10), e0141363.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child development*, 78(2), 647-663.

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

Camilli, G., Vargas, S., Ryan, S., & Barnett, W.S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. Teachers College Record, 112, 579-620.

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.

Connolly, F., & Olsen, L. S. (2012). Early elementary performance and attendance in Baltimore City Schools' pre-kindergarten and kindergarten. Baltimore, MD: Baltimore Education Research Consortium.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do". *Developmental psychobiology*, 29(4), 315-334.

Dubay, L., & Holla, N. (2016). Does attendance in early education predict attendance in elementary school? An analysis of DCPS's Early Education Program. Washington, DC: Urban Institute.

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). New York: Russell Sage Foundation and Spencer Foundation.

Dunn, L. M., & Dunn, D. M. (2007). PPVT-4: Peabody picture vocabulary test. Pearson Assessments.

Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., Cai, K., Clifford, R.M., Ebanks, C., Griffin, J.A. & Henry, G. T. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child development*, 78(2), 558-580.

Ehrlich, S. B., Gwynne, J. A., & Allensworth, E. M. (2018). Quarterly pre-kindergarten attendance matters: Early chronic absence patterns and relationships to learning outcomes. *Early Childhood Research Quarterly*, 44, 136–151.

Frede, E. C., Jung, K., Barnett, W. S., Lamy, C. E., & Figueras, A. (2009). The APPLES blossom: Abbott Preschool Program Longitudinal Effects Study (APPLES): Preliminary effects through second grade. National Institute for Early Education Research: New Brunswick, NJ.

Friedman-Krauss, A., Barnett, S., & Nores, M. (2016). How much can high-quality universal pre-K reduce achievement gaps? Washington, DC: Center for American Progress and National Institute for Early Education Research. Retrieved from https://cdn.americanprogress.org/wp-content/uploads/2016/04/01115656/NIEER-AchievementGaps-report.pdf

Gormley, W. T. (2008). The Effects of Oklahoma's Pre-K Program on Hispanic Children*. Social Science Quarterly, 89(4), 916-936.

Graham, G. (2013, March 11). Tulsa's preschool programs seen as national model. Tulsa World: Boulder, Tulsa. Retrieved from: http://www.tulsaworld.com/news/local/tulsa-s-preschool-programs-seen-as-national-model/article_a932f79f-ec5c-5619-a89a-0ef88d271830.html

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

Jung, K., Barnett, W. S., Hustedt, J. T., & Francis, J. (2013). Longitudinal effects of the Arkansas Better Chance Program: Findings from first grade through fourth grade. National Institute for Early Education Research: New Brunswick, NJ

Ludwig, J., & Phillips, D. A. (2008). Long-term effects of Head Start on low-income children. Annals of the New York Academy of Sciences, 1136(1), 257-268.

Meador, D. N., Turner, K. A., Lipsey, M. W., & Farran, D. C. (2013). Administering Measures from the PRI Learning-Related Cognitive Self- Regulation Study. Nashville, TN: Peabody Research Institute. Available at https://my.vanderbilt.edu/cogselfregulation/files/2012/11/SR-Measure-Training-Manual-final.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.

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., 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.

Office of Head Start, U.S. Department of Health and Human Services, Administration for Children and Families. (2015). A National Overview of Grantee CLASS Scores in 2015. Washington, DC.

Peisner-Feinberg, E. S., LaForett, D. R., Schaaf, J. M., Hildebrandt, L. M., Sideris, J., & Pan, Y. (2014). Children's outcomes and program quality in the North Carolina Pre-kindergarten Program: 2012–2013 Statewide evaluation. Frank Porter Graham Child Development Institute: Chapel Hill, NC.

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.

Qi, C. H., Kaiser, A. P., Milan, S., & Hancock, T. (2006). Language performance of low-income African American and European American preschool children on the PPVT–III. *Language, Speech, and Hearing Services in Schools*.

Ritchie, S., Weiser, B., Mason, E., & Holland, A. (2015). The EduSnap Professional Learning System. Durham, NC: Snapshot. Retrieved from: https://getEduSnap .com.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). Woodcock-Johnson IV tests of achievement. Riverside Publishing.

Weiland, C. & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, 84(6), 2112-2130.

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly*, 28(2), 199-209.

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.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature protocols*, 1(1), 297.

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.

EduSnap Classroom Observation Measure (Ritchie, Weiser, Mason, & Holland, 2015)

The EduSnap Classroom Observation Measure is a tool that quantifies the experiences that children have throughout the day. The measure provides an in-depth look at how students experience their day by documenting the actual time students spend in various activities, such as activity settings (e.g., whole group, free choice, transitions), content areas (e.g., reading, science, math), student learning approaches (e.g., collaboration, metacognition), and teaching approaches (e.g., didactic, scaffolds). The information collected provides insight into curriculum balance, curriculum integration, and interactions between teachers and children. While there is no consensus on the exact number of minutes any one child should spend in a particular activity, it is understood that high-quality experiences for children are represented by a balance across activity settings, content areas, and student learning & teaching approaches so that all children are provided with a variety of experiences throughout the school day.

Table A.2. EduSnap Section Descriptions.

Section	Description
Activity Settings	It is important as children optimize their learning in different ways - some have greater or less tolerance for large and small groupings, noise level, auditory, visual, tactile and kinesthetic input.
	Activity settings include: Whole group, Transitions, Small group, Group work, Individual, Choice, Meals
Content Areas	Frequency of exposure to learning opportunities increases children's academic achievement. Developing literacy and math skills and processes is essential for success at all levels of learning and should be emphasized according to children's needs and developmental levels. However, these should be balanced with the importance of learning foundational knowledge in all subject areas, including science, social studies, art, and music. A well integrated curriculum allows access to multiple content areas to children and strengthens learning across and within them.
	Content areas include: Literacy, Math, and Other (Science, Gross Motor, Social Studies, Aesthetics)
Student Learning Approaches	Providing children many opportunities to work together and to engage in metacognitive thinking supports both their social/emotional and academic development.
	Student learning approaches include: Collaboration and Metacognition
Teaching Approaches	Scaffolded instruction involves teachers' use of open-ended questions, feedback loops, and probing to more deeply engage children's thinking and understanding. This type of instruction enables teachers to gauge how much the children understood from a lesson, identify and remediate group or individual misunderstandings, and engage children in their learning process. It therefore enables the teacher to respond by modifying the current and subsequent learning experiences and activities according to individual and group needs. Both didactic and scaffolded instruction are important teaching styles to be incorporated in a balanced fashion throughout the course of each school day and within each lesson. Didactic instruction provides children with needed practice and repetition that helps them build their skill base across the curriculum, giving them models, demonstrations, information, and guidance.
	Teaching approaches include: Scaffolds and Didactic

Child Measures

The *Peabody Picture Vocabulary Test—Fourth Edition (PPVT-IV;* Dunn & Dunn, 2007) is an adaptive test comprised of 228-items measuring receptive vocabulary in standard English. The PPVT is predictive of general cognitive abilities and is a direct measure of vocabulary size. That is adaptive means that a portion of the test is used with rules for establishing a floor, below which the child is assumed to know all the answers and a ceiling above which the child is assumed to know none of the answers. It is designed for use with population ages 2.5 and above. The PPVT has shown concurrent validity (e.g., Qi, Kaiser, Milan, & Hancock, 2006) and the results of these tests are found to be strongly correlated with school success (Blair & Razza, 2007; Early, et al., 2007). This instrument has been used in various preschool studies (e.g., Barnett, et al., 2018; Frede, et al., 2009; Gormley, 2008; Jung et al., 2013; Ludwig & Phillips, 2008; Peisner-Feinberg, et al., 2014; Weiland & Yoshikawa, 2013) and capture large gains for low income, dual-language and non-white children. In the Faces study (Aikens, et al., 2017) Cronbach's alpha reliability for the PPVT-4 was 0.97.

The *Woodcock-Johnson Psycho-Educational Battery—Fourth Edition (WJ- IV;* Woodcock, McGrew, Mather, & Schrank, 2001) includes multiple subtests. Only the *Applied Problems* and *Letter-Word Identification* subtests were used. WJ- IV is normed on a stratified random sample of 6,359 English-speaking subjects in the United States. The WJ is also an adaptive test, used with populations above age 3. Correlations of the WJ with other tests of cognitive ability and achievement are reported to range from 0.60 to 0.70. This measure has been used in numerous large-scale preschool studies (e.g., Early, et al., 2007; Gormley, 2008; Graham, 2013; Peisner-Feinberg, et al., 2014; Weiland & Yoshikawa, 2013; Wong, Cook, Barnett, & Jung, 2008). In the Faces study (Aikens, et al., 2017) Cronbach's alpha reliability for the WJ-LW III was 0.90 and for the WJ-AP III was 0.88.

The *Dimensional Change Card Sort Task* (DCCS; Zelazo, 2006) is an executive function task requires children to sort a set of cards based on different sorting criteria given by the examiner. The test assesses attention-shifting and short-term memory combined. Scores on the DCCS reflect a pass/fail system on each of three levels of increasing difficulty. Raw scores range between 0 and 3, where a score of 0 means a child did not pass the first level, which includes a color sorting task. In addition, full scores reflect the level of total passes. In the first level, children are tasked with sorting two objects by a color rule, in a second level by a shape rule, and in the advanced level, children are asked to ignore color or shape by adding a border to cards to indicate which attribute to sort by. There are no standard score equivalents. However, in a study of test-retest reliability, means by age for children age 48 months or younger were 1.14 for 48–50 months they were 1.33, for 51–53 months they were 1.42, and for 54–56 months they were 1.58 (Meador et al., 2013).

The *Peg Tapping Test* (PT; Diamond & Taylor, 1996) requires children to follow directions to tap a peg twice when the experimenter taps once and vice versa. It requires children to inhibit a natural tendency to mimic the experimenter while remembering the rule for the correct response, tapping into inhibitory control, attention, and short-term memory. Sixteen trials are conducted with eight one-tap and eight two-tap trials in a random sequence. The final score for Peg Tapping is a sum of all the 16 items that comprise the test. While there are no standard score equivalents, in a study of test-retest reliability, means by age for children age 48 months or younger was 4.05, for 48–50 months they were 4.57, for 51–53 months they were 6.02, and for 54–56 months they were 7.87 (Meador et al., 2013).

The *Caregiver-Teacher Report Form* (C-TRF: Achenbach, 2009 & Achenbach, et al., 2013) ages 1½–5 is a short questionnaire for obtaining teachers' reports of their child's competencies and problems. It is normed based on 1,192 children. It has also been tested in 14 societies with 9,389 children. Teachers were instructed to rate the child's behavior early in the fall and again late in the spring. It consists of a 99-item list of behaviors to which the teacher gives a response of 0, 1, or 2 (not true, somewhat true, or very true). Scores included in this report are for total behavior problems.

Appendix B. Outcomes.

Table B.1. Child absenteeism AY 2018–19.

	Attendance		Moderate		Severe		Valid N
	Count	Valid %	Count	Valid %	Count	Valid %	
August	81	84.4%	9	9.4%	6	6.3%	96
September	310	74.3%	76	18.2%	31	7.4%	417
October	347	78.9%	56	12.7%	37	8.4%	440
November	261	59.6%	90	20.5%	87	19.9%	438
December	278	65.0%	76	17.8%	74	17.3%	428
January	270	63.7%	60	14.2%	94	22.2%	424
February	220	52.8%	116	27.8%	81	19.4%	417
March	280	67.8%	69	16.7%	64	15.5%	413
April	276	67.2%	77	18.7%	58	14.1%	411
May	285	69.9%	47	11.5%	76	18.6%	408
June	138	51.7%	40	15.0%	89	33.3%	267
Year	275	60.8%	112	24.8%	65	14.4%	452

Note. Absence rate information as collected by PHMC from September 2018 through June of 2019.

Table B.2. Frequency and percentage of lead teacher replaced in AY 2018-19.

Subgroups		**Turnover**
STAR Level	1-2 (n=3)	0.00%
	3 (n=72)	20.83%
	4 (n=72)	15.28%
Lead Teacher Credential	CDA/ECE Course (n=8)	0.00%
	AA (n=53)	20.75%
	BA (n=48)	12.50%
	MA, Med, MSc (n=23)	13.04%
	Other, Studying, Missing (n=15)	40.00%
PHLpreK Partner Agency	UAC (n=45)	31.11%
	PHMC (n=73)	12.50%
	1199c (n=14)	6.67%
	SDP (n=15)	13.33%
Curriculum	Creative (n=94)	15.96%
	Creative + (n=42)	23.81%
	Mother Goose (n=10)	10.00%
	Other (n=1)	0.00%
Success by 6	Yes (n=5)	0.00%
	No (n=142)	18.31%
New Site	Yes (n=11)	9.09%
	No (n=136)	18.38%
Total		**17.69%**

Note. Changes of lead teachers are based on the information collected as of September 2018 and June 2019 and likely do not account for full turnover.

Table B.3. PPVT raw score means and gains by child characteristics

		PPVT Raw F18			PPVT Raw S19		PPVT Raw Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		585	56.55	23.77	71.79	24.02	15.14	15.14
Gender	Male	287	55.48	22.43	71.52	24.46	16.62	15.97
	Female	298	57.57	24.99	72.06	23.64	13.71	14.17
Age	3	251	46.30	19.90	62.05	21.92	16.01	15.96
	4	334	64.24	23.56	78.89	23.00	14.50	14.50
Ethnicity	White	84	66.67	26.83	81.09	26.87	14.14	12.57
	Black	367	56.09	22.91	71.30	23.08	15.40	16.72
	Hispanic	82	49.59	20.93	64.33	23.20	13.41	10.89
	Other	51	54.12	24.30	70.31	21.95	17.67	13.61
Language	English	535	57.87	23.42	73.09	23.55	15.18	15.54
	DLL	50	42.36	23.01	58.58	25.06	14.73	10.33
IEP	No	552	57.17	23.74	72.32	23.61	14.98	15.04
	Yes	33	46.06	22.06	63.81	28.85	17.48	16.65

Table B.4. PPVT standard score means and gains by child characteristics

		PPVT SS F18			PPVT SS S19		PPVT SS Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		585	92.60	17.48	96.70	16.53	4.17	12.93
Gender	Male	287	91.36	16.55	96.26	17.09	5.51	13.99
	Female	298	93.80	18.27	97.13	15.99	2.88	11.70
Age	3	251	93.03	16.39	96.58	16.86	3.78	13.96
	4	334	92.28	18.27	96.80	16.31	4.45	12.14
Ethnicity	White	84	101.33	19.63	104.41	17.82	3.01	10.53
	Black	367	92.13	16.44	96.21	16.04	4.39	14.38
	Hispanic	82	87.49	16.23	90.81	15.21	2.59	8.91
	Other	51	89.86	18.50	95.60	15.03	7.00	11.36
Language	English	535	93.65	16.94	97.61	16.13	4.10	13.21
	DLL	50	81.38	19.35	87.47	17.85	4.91	9.69
IEP	No	552	93.20	17.27	97.18	16.06	3.98	12.88
	Yes	33	82.58	18.17	89.42	21.51	7.06	13.45

Table B.5. WJ-LW Raw score means and gains by child characteristics

		LWIDNT Raw F18			LWIDNT Raw S19		LWIDNT Raw Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		585	6.17	5.66	9.61	7.34	3.29	3.87
Gender	Male	287	6.49	6.48	9.84	8.43	3.24	4.33
	Female	298	5.87	4.74	9.39	6.11	3.33	3.38
Age	3	251	4.44	4.56	7.19	5.93	2.64	3.14
	4	334	7.47	6.06	11.37	7.76	3.76	4.27
Ethnicity	White	84	7.73	6.46	11.79	9.38	4.05	6.20
	Black	367	6.11	5.60	9.42	6.90	3.12	3.26
	Hispanic	82	5.06	4.15	7.97	5.34	2.86	2.96
	Other	51	5.86	6.47	9.56	8.06	3.69	3.48
Language	English	535	6.30	5.80	9.82	7.57	3.34	3.93
	DLL	50	4.78	3.65	7.49	3.79	2.71	3.17
IEP	No	552	6.12	5.54	9.50	7.01	3.22	3.47
	Yes	33	7.09	7.47	11.32	11.23	4.23	7.81

Table B.6. WJ-LW standard score means and gains by child characteristics

		LWIDNT SS F18			LWIDNT SS S19		LWIDNT SS Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		585	92.91	14.33	93.69	14.30	0.56	10.24
Gender	Male	287	93.49	14.69	93.51	15.08	0.03	10.55
	Female	298	92.36	13.98	93.88	13.53	1.07	9.93
Age	3	251	95.49	14.11	96.05	14.81	0.23	11.17
	4	334	90.98	14.22	91.98	13.70	0.80	9.53
Ethnicity	White	84	98.60	14.42	99.64	13.70	1.25	11.29
	Black	367	92.58	14.24	93.10	14.52	0.15	10.27
	Hispanic	82	90.48	13.23	90.27	12.43	0.09	8.73
	Other	51	90.10	14.55	92.58	14.07	2.73	10.27
Language	English	535	93.29	14.24	94.12	14.36	0.56	10.21
	DLL	50	88.92	14.89	89.40	13.07	0.60	10.69
IEP	No	552	92.90	14.16	93.74	13.91	0.57	10.08
	Yes	33	93.15	17.25	93.06	19.58	0.48	12.73

Table B.7. WJ-AP raw score means and gains by child characteristics

		APPROB Raw F18			APPROB Raw S19		APPROB Raw Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		585	6.63	4.05	10.04	4.42	3.39	3.21
Gender	Male	287	6.57	3.91	9.98	4.60	3.52	3.35
	Female	298	6.68	4.19	10.09	4.24	3.27	3.06
Age	3	251	4.75	3.42	7.99	4.00	3.33	3.22
	4	334	8.04	3.92	11.53	4.11	3.44	3.21
Ethnicity	White	84	8.70	4.12	12.49	4.61	3.74	3.20
	Black	367	6.33	4.07	9.49	4.14	3.18	3.25
	Hispanic	82	5.90	3.38	9.43	4.48	3.60	2.91
	Other	51	6.59	3.88	10.33	4.43	3.87	3.32
Language	English	535	6.76	4.04	10.17	4.36	3.39	3.23
	DLL	50	5.20	3.90	8.64	4.82	3.44	3.05
IEP	No	552	6.68	4.02	10.12	4.28	3.42	3.18
	Yes	33	5.79	4.53	8.81	6.14	2.97	3.67

Table B.8. WJ-AP standard score means and gains by child characteristics

		APPROB SS F18			APPROB SS S19		APPROB SS Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		585	83.82	17.00	88.03	17.48	4.37	15.08
Gender	Male	287	83.33	16.58	87.32	18.56	4.72	15.54
	Female	298	84.29	17.40	88.72	16.38	4.04	14.63
Age	3	251	84.86	16.38	87.47	19.10	3.10	16.22
	4	334	83.04	17.43	88.44	16.21	5.29	14.14
Ethnicity	White	84	94.25	14.78	99.25	16.01	4.99	13.38
	Black	367	82.27	17.06	85.39	17.26	3.49	15.72
	Hispanic	82	81.07	15.78	85.97	16.72	5.50	14.31
	Other	51	82.51	16.13	89.10	15.05	7.27	14.57
Language	English	535	84.51	16.59	88.64	17.24	4.27	14.97
	DLL	50	76.40	19.54	81.84	18.83	5.42	16.24
IEP	No	552	84.19	16.67	88.58	16.65	4.50	14.93
	Yes	33	77.61	21.05	79.74	26.09	2.45	17.31

Table B.9. DCCS Final score means and gains by child characteristics

		DCCS Final F18			DCCS Final S19		DCCS Final Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		585	1.19	0.56	1.44	0.62	0.25	0.66
Gender	Male	287	1.14	0.55	1.42	0.63	0.29	0.67
	Female	298	1.24	0.56	1.46	0.61	0.21	0.65
Age	3	251	1.08	0.49	1.25	0.57	0.19	0.60
	4	334	1.27	0.60	1.58	0.61	0.30	0.70
Ethnicity	White	84	1.36	0.63	1.76	0.64	0.40	0.74
	Black	367	1.14	0.55	1.35	0.59	0.21	0.63
	Hispanic	82	1.21	0.51	1.49	0.58	0.27	0.64
	Other	51	1.24	0.51	1.46	0.62	0.23	0.69
Language	English	535	1.19	0.56	1.45	0.62	0.25	0.65
	DLL	50	1.16	0.58	1.40	0.62	0.24	0.71
IEP	No	552	1.19	0.56	1.45	0.62	0.26	0.67
	Yes	33	1.21	0.48	1.32	0.60	0.10	0.47

Table B.10. Peg Tapping score means and gains by child characteristics

		PT Final F18			PT Final S19		PT Final Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		585	4.29	5.28	7.02	5.96	2.65	4.92
Gender	Male	287	4.33	5.15	6.66	5.95	2.40	4.74
	Female	298	4.26	5.41	7.36	5.97	2.89	5.09
Age	3	251	2.15	3.93	4.98	5.48	2.86	4.73
	4	334	5.90	5.59	8.51	5.87	2.50	5.05
Ethnicity	White	84	5.81	5.79	9.35	5.94	3.45	4.87
	Black	367	3.95	5.18	6.46	5.87	2.37	4.71
	Hispanic	82	4.29	5.00	6.34	5.83	2.36	4.86
	Other	51	4.22	5.35	7.67	5.94	3.56	6.14
Language	English	535	4.36	5.33	7.15	5.97	2.69	5.00
	DLL	50	3.56	4.73	5.71	5.84	2.22	4.05
IEP	No	552	4.38	5.34	7.16	5.97	2.70	4.96
	Yes	33	2.82	3.88	4.87	5.56	1.90	4.18

Table B.11. C-TRF Total Problems raw score means and gains by child characteristics

		C-TRF Raw F18			C-TRF Raw S19		C-TRF Raw Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		542	16.35	22.00	17.54	22.51	1.07	18.78
Gender	Male	267	18.40	22.44	20.40	24.70	1.29	20.35
	Female	275	14.35	21.41	14.78	19.85	0.85	17.16
Age	3	231	20.22	26.27	20.98	25.36	1.35	19.88
	4	311	13.47	17.70	15.01	19.83	0.86	17.98
Ethnicity	White	82	13.46	16.94	16.90	21.88	2.84	19.29
	Black	335	17.40	23.88	18.32	22.24	0.57	19.57
	Hispanic	77	17.26	21.73	17.63	27.19	1.25	16.73
	Other	48	12.46	14.85	13.57	18.18	0.84	15.68
Language	English	497	16.34	22.00	17.40	21.81	0.83	19.01
	DLL	45	16.36	22.19	18.93	29.03	3.54	16.13
IEP	No	510	15.54	21.54	16.84	21.82	1.27	18.50
	Yes	32	29.19	25.43	28.03	29.62	-1.93	22.56

Table B.12. C-TRF Total Problems T score means and gains by child characteristics

		C-TRF T F18			C-TRF T S19		C-TRF T Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		542	46.15	11.58	46.34	12.30	0.18	10.39
Gender	Male	267	46.40	12.03	46.92	13.13	0.35	10.61
	Female	275	45.90	11.15	45.79	11.45	0.01	10.20
Age	3	231	48.32	12.51	48.07	13.27	0.14	10.70
	4	311	44.53	10.58	45.08	11.40	0.21	10.19
Ethnicity	White	82	44.70	10.44	45.71	12.75	0.81	11.53
	Black	335	46.64	11.94	47.00	12.09	0.15	10.26
	Hispanic	77	46.78	11.81	45.54	13.69	-0.61	9.77
	Other	48	44.19	10.40	44.32	11.05	0.29	10.23
Language	English	497	46.12	11.57	46.38	12.11	0.19	10.51
	DLL	45	46.40	11.84	45.97	14.47	0.00	9.10
IEP	No	510	45.67	11.45	45.95	12.18	0.32	10.43
	Yes	32	53.78	11.26	52.41	12.84	-1.89	9.79

Table B.13. C-TRF Internalizing Problems raw score means and gains by child characteristics

		C-TRF IP Raw F18			C-TRF IP Raw S19		C-TRF IP Raw Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		542	4.57	6.64	4.73	6.45	0.20	6.10
Gender	Male	267	4.89	6.43	5.36	7.05	0.24	6.21
	Female	275	4.27	6.84	4.12	5.77	0.16	6.00
Age	3	231	5.69	8.08	5.55	7.21	0.08	6.35
	4	311	3.74	5.19	4.12	5.77	0.29	5.92
Ethnicity	White	82	4.06	4.83	4.66	6.15	0.44	6.03
	Black	335	4.81	7.39	4.91	6.37	0.08	6.34
	Hispanic	77	4.90	6.00	4.73	7.70	0.31	5.44
	Other	48	3.27	4.25	3.70	5.67	0.40	5.69
Language	English	497	4.57	6.70	4.69	6.20	0.13	6.14
	DLL	45	4.60	5.97	5.07	8.67	1.00	5.68
IEP	No	510	4.35	6.54	4.51	6.16	0.24	5.94
	Yes	32	8.16	7.34	8.00	9.37	-0.41	8.13

Table B.14. C-TRF Internalizing Problems T score means and gains by child characteristics

		C-TRF IP T F18			C-TRF IP T S19		C-TRF IP T Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		542	45.70	10.52	45.72	10.88	-0.10	10.39
Gender	Male	267	46.13	10.50	46.63	11.33	0.06	10.33
	Female	275	45.27	10.54	44.85	10.38	-0.26	10.47
Age	3	231	47.38	11.69	47.03	11.68	-0.39	10.93
	4	311	44.44	9.39	44.77	10.17	0.10	10.00
Ethnicity	White	82	45.38	9.17	45.74	10.87	0.24	10.51
	Black	335	45.90	11.07	46.11	10.74	-0.16	10.47
	Hispanic	77	46.30	10.44	45.32	12.30	-0.34	10.48
	Other	48	43.88	8.83	43.81	9.99	0.00	9.96
Language	English	497	45.68	10.55	45.75	10.62	-0.13	10.42
	DLL	45	45.82	10.31	45.41	13.59	0.22	10.21
IEP	No	510	45.30	10.42	45.35	10.64	-0.03	10.32
	Yes	32	52.06	10.22	51.41	12.97	-1.14	11.52

Table B.13. C-TRF Externalizing Problems raw score means and gains by child characteristics

		C-TRF EP Raw F18			C-TRF EP Raw S19		C-TRF EP Raw Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		7.80	11.03	8.46	11.86	0.44	9.51	7.80
Gender	Male	9.12	11.58	10.10	12.83	0.67	10.16	9.12
	Female	6.52	10.33	6.87	10.63	0.21	8.84	6.52
Age	3	9.44	12.61	10.31	13.22	0.95	10.16	9.44
	4	6.58	9.53	7.09	10.57	0.06	9.02	6.58
Ethnicity	White	5.87	9.04	7.70	11.54	1.59	9.67	5.87
	Black	8.37	11.55	8.97	11.95	0.23	10.09	8.37
	Hispanic	8.12	11.61	8.46	13.20	0.54	7.06	8.12
	Other	6.63	9.06	6.45	9.84	-0.40	8.49	6.63
Language	English	7.80	10.98	8.38	11.70	0.32	9.73	7.80
	DLL	7.73	11.66	9.23	13.52	1.67	6.76	7.73
IEP	No	7.45	10.86	8.21	11.79	0.56	9.59	7.45
	Yes	13.37	12.34	12.20	12.48	-1.38	8.24	13.37

Table B.14. C-TRF Externalizing Problems T score means and gains by child characteristics

		C-TRF EP T F18			C-TRF EP T S19		C-TRF EP T Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Total		542	48.34	10.17	48.44	10.89	0.06	8.74
Gender	Male	267	48.49	10.42	49.09	11.32	0.50	8.74
	Female	275	48.20	9.95	47.81	10.46	-0.37	8.73
Age	3	231	50.08	10.98	50.14	11.93	0.42	8.99
	4	311	47.05	9.34	47.19	9.91	-0.20	8.55
Ethnicity	White	82	46.27	8.97	47.39	10.83	0.95	9.31
	Black	335	48.86	10.48	48.91	10.87	-0.13	8.94
	Hispanic	77	48.79	10.58	48.58	12.07	0.21	7.83
	Other	48	47.54	8.99	47.00	9.63	-0.49	7.72
Language	English	497	48.32	10.16	48.38	10.74	-0.01	8.85
	DLL	45	48.64	10.42	49.10	12.64	0.81	7.38
IEP	No	510	47.99	10.12	48.14	10.84	0.15	8.82
	Yes	32	53.97	9.55	53.07	10.92	-1.25	7.34

Appendix C. Child Estimations.

Table C.1. Multivariate analyses of children's 2018-19 posttest (spring) standard score in relation to child and site or classroom characteristics including the CLASS (scale) and EduSnap.

	Receptive Vocabulary	Literacy	Math	DCCS Final	Peg Tapping	Socio-Emotional (inverted measure)
Female	-1.822~	0.791	-0.936	0.003	0.442	-0.343
	(1.00)	(0.87)	(1.22)	(0.05)	(0.42)	(0.74)
Af.Am.	-1.808	-2.775~	-5.988**	-0.360***	-1.923**	-0.356
	(1.70)	(1.42)	(2.03)	(0.08)	(0.69)	(1.39)
Hisp.	-1.943	-2.237	-2.142	-0.144	-2.299*	-2.537
	(2.40)	(2.03)	(2.86)	(0.12)	(0.98)	(2.03)
Other Race/Ethn.	0.637	-0.489	-1.329	-0.203~	-1.207	-1.332
	(2.13)	(1.81)	(2.56)	(0.10)	(0.87)	(1.59)
DLL	-2.298	-1.478	-3.381	-0.115	-0.357	2.943
	(2.31)	(1.97)	(2.78)	(0.11)	(0.95)	(1.87)
IEP	0.867	1.891	-3.102	-0.165	-0.974	-2.398
	(2.20)	(1.89)	(2.65)	(0.11)	(0.91)	(1.69)
Moderate Attendance Risk	-0.938	-0.607	0.310	0.023	-1.237*	-0.821
	(1.35)	(1.17)	(1.64)	(0.07)	(0.56)	(1.05)
Severe Attendance Risk	-3.127	-3.250~	-3.635	-0.009	-0.629	-2.938~
	(1.96)	(1.69)	(2.36)	(0.10)	(0.81)	(1.53)
No Attendance Information	0.524	-1.312	-1.377	0.002	-0.744	-1.686
	(1.51)	(1.26)	(1.77)	(0.08)	(0.60)	(1.80)
Star 3	10.782*	9.320*	15.984**	-0.013	2.750	-4.341
	(4.48)	(3.73)	(5.22)	(0.23)	(1.80)	(5.21)
Star 4	10.030*	8.456*	15.297**	-0.001	3.675*	-4.843
	(4.45)	(3.70)	(5.18)	(0.23)	(1.78)	(5.15)
Creative+	1.630	2.011	0.606	-0.096	0.952	-1.146
	(1.60)	(1.32)	(1.85)	(0.08)	(0.63)	(1.91)
Mother Goose	1.378	-0.636	-1.269	0.087	1.843*	0.584
	(2.31)	(1.91)	(2.69)	(0.12)	(0.92)	(3.73)
LT Turnover	-1.065	1.119	-1.344	0.046	-0.723	-0.604
	(1.56)	(1.29)	(1.81)	(0.08)	(0.63)	(2.08)
LT Associate	-3.716	-1.251	2.602	0.181	0.755	2.272
	(2.73)	(2.26)	(3.17)	(0.14)	(1.08)	(3.35)
LT Bachelor	-3.764	-0.824	1.242	0.201	1.506	-0.360
	(2.76)	(2.28)	(3.21)	(0.14)	(1.10)	(3.31)
LT Master	0.346	-1.430	4.921	0.062	1.505	3.297
	(3.12)	(2.58)	(3.62)	(0.16)	(1.24)	(3.80)
LT Other	-1.109	-2.188	3.682	-0.026	0.927	3.900
	(3.55)	(2.94)	(4.11)	(0.18)	(1.41)	(4.23)
LT Af.Am.	0.428	0.311	-1.714	0.058	-0.592	1.545
	(2.14)	(1.77)	(2.48)	(0.11)	(0.85)	(2.46)
LT Hisp.	-0.596	-0.122	-5.047~	-0.088	-0.489	4.224
	(2.56)	(2.12)	(2.96)	(0.13)	(1.02)	(3.42)
LT Other	-2.685	-1.064	-6.066	-0.105	-0.043	0.998
	(3.44)	(2.86)	(4.01)	(0.17)	(1.37)	(4.11)
LT Race Blank	-3.815	0.225	-4.603	-0.020	-0.658	-5.425~

	(2.54)	(2.11)	(2.95)	(0.13)	(1.01)	(3.02)
CLASS ES	1.847	1.210	1.830	0.026	-0.010	-0.596
	(1.55)	(1.28)	(1.79)	(0.08)	(0.61)	(2.10)
CLASS CO	1.906~	0.331	1.845	-0.005	0.446	1.793
	(1.12)	(0.93)	(1.30)	(0.06)	(0.45)	(1.42)
CLASS IS	-2.073*	-0.441	-2.026*	0.007	-0.042	-0.164
	(0.85)	(0.70)	(0.98)	(0.04)	(0.34)	(1.17)
Transitions	-0.037	-0.119~	0.012	0.001	0.034	-0.020
	(0.08)	(0.06)	(0.09)	(0.00)	(0.03)	(0.10)
Small Group	-0.308*	-0.062	-0.314~	-0.011	0.036	-0.238
	(0.15)	(0.12)	(0.17)	(0.01)	(0.06)	(0.18)
Choice	-0.025	-0.034	-0.025	0.000	0.009	0.081
	(0.05)	(0.04)	(0.06)	(0.00)	(0.02)	(0.06)
Oral	-0.063	-0.016	0.072	-0.004	-0.026	-0.105
	(0.09)	(0.07)	(0.10)	(0.00)	(0.03)	(0.11)
Read To	-0.183	-0.213~	-0.039	-0.002	0.023	-0.228
	(0.15)	(0.13)	(0.18)	(0.01)	(0.06)	(0.19)
Word Ident	0.056	0.226**	0.206~	-0.001	0.066	0.044
	(0.10)	(0.08)	(0.12)	(0.01)	(0.04)	(0.13)
Numbers	0.159	0.107	0.189	0.002	0.048	0.071
	(0.11)	(0.09)	(0.12)	(0.01)	(0.04)	(0.13)
Geometry	-0.016	-0.122~	-0.017	-0.002	0.023	0.007
	(0.09)	(0.07)	(0.10)	(0.00)	(0.04)	(0.11)
Science	-0.123	-0.102	-0.152~	-0.002	-0.009	0.083
	(0.08)	(0.07)	(0.09)	(0.00)	(0.03)	(0.10)
Gross Motor	-0.111	-0.039	-0.052	0.000	0.037	0.104
	(0.09)	(0.07)	(0.10)	(0.00)	(0.04)	(0.11)
Aesthetics	0.277**	0.099	0.298*	0.006	-0.001	0.221~
	(0.10)	(0.08)	(0.12)	(0.01)	(0.04)	(0.12)
Didactic	-0.015	-0.065	-0.066	-0.006*	-0.002	0.088
	(0.06)	(0.05)	(0.07)	(0.00)	(0.02)	(0.07)
N	457	457	457	457	456	397

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are 3-year-olds, Males, White, English, Non-IEP, Creative Curriculum, Star Level 1, Lead Teacher some college and below, Lead Teacher White. Other controls are pre-test, age in months, days between tests, age cohort (3 versus 4 year olds), class size, missing absence data, and hub. Standard scores are used for PPVT, and WJ or WM, T scores are used for C-TRF. Errors are clustered by site. Estimations including self reported teacher race and education were consistent with findings reported here. Estimations including CLASS dimension as thresholds (above cutoff points), also were consistent with findings reported here.

Table C.2. Multivariate analyses of children's 2018-19 posttest (spring) standard score in relation to child and site or classroom characteristics including the CLASS (scale and cutoff) and EduSnap.

	Receptive Vocabulary	Literacy	Math	DCCS Final	Peg Tapping	Socio- Emotional (inverted measure)
Female	-1.742~	0.766	-0.864	0.005	0.466	-0.360
	(1.00)	(0.86)	(1.21)	(0.05)	(0.42)	(0.74)
Af.Am.	-2.098	-2.683~	-6.305**	-0.368***	-2.077**	-0.444
	(1.68)	(1.41)	(2.02)	(0.08)	(0.69)	(1.39)
Hisp.	-2.211	-2.296	-2.875	-0.154	-2.639**	-2.913
	(2.40)	(2.05)	(2.89)	(0.12)	(0.99)	(2.04)
Other Race/Ethn.	0.925	-0.458	-1.133	-0.195~	-1.165	-1.300
	(2.11)	(1.79)	(2.54)	(0.10)	(0.87)	(1.59)
DLL	-2.451	-1.265	-3.346	-0.113	-0.354	3.062
	(2.29)	(1.96)	(2.76)	(0.11)	(0.94)	(1.87)
IEP	0.849	1.801	-3.445	-0.168	-1.117	-2.584
	(2.20)	(1.88)	(2.64)	(0.11)	(0.91)	(1.69)
Moderate Attendance Risk	-1.363	-0.985	-0.153	0.010	-1.323*	-1.052
	(1.35)	(1.16)	(1.64)	(0.07)	(0.56)	(1.05)
Severe Attendance Risk	-3.333~	-2.969~	-3.949~	-0.017	-0.810	-2.938~
	(1.95)	(1.69)	(2.36)	(0.10)	(0.81)	(1.53)
No Attendance Information	-0.363	-2.309~	-2.266	-0.027	-0.878	-2.481
	(1.50)	(1.28)	(1.81)	(0.08)	(0.62)	(1.79)
Star 3	8.831*	7.571*	15.053**	-0.067	2.908	-5.699
	(4.38)	(3.76)	(5.28)	(0.23)	(1.82)	(5.22)
Star 4	7.587~	6.585~	13.743**	-0.074	3.615*	-6.496
	(4.34)	(3.72)	(5.23)	(0.23)	(1.80)	(5.14)
Creative+	1.212	1.729	0.165	-0.109	0.845	-1.539
	(1.54)	(1.31)	(1.85)	(0.08)	(0.63)	(1.86)
Mother Goose	2.360	-0.750	-0.855	0.113	1.883*	-0.862
	(2.30)	(1.96)	(2.78)	(0.12)	(0.95)	(3.76)
LT Turnover	0.366	2.093	-0.287	0.086	-0.615	0.515
	(1.56)	(1.33)	(1.87)	(0.08)	(0.65)	(2.14)
LT Associate	-5.399*	-2.129	1.509	0.133	0.625	1.517
	(2.66)	(2.28)	(3.21)	(0.14)	(1.10)	(3.31)
LT Bachelor	-5.549*	-1.632	0.024	0.149	1.321	-0.812
	(2.70)	(2.31)	(3.25)	(0.14)	(1.12)	(3.28)
LT Master	-0.912	-2.072	4.367	0.028	1.535	3.273
	(3.02)	(2.59)	(3.64)	(0.16)	(1.25)	(3.77)
LT Other	-3.379	-3.358	1.814	-0.096	0.585	2.595
	(3.47)	(2.97)	(4.17)	(0.18)	(1.43)	(4.24)
LT Af.Am.	0.606	0.266	-1.913	0.061	-0.716	1.303
	(2.06)	(1.76)	(2.47)	(0.11)	(0.85)	(2.40)
LT Hisp.	-0.983	-0.252	-5.925*	-0.105	-0.855	2.657
	(2.48)	(2.12)	(2.98)	(0.13)	(1.02)	(3.43)
LT Other	-2.812	-1.093	-6.247	-0.109	-0.109	0.465
	(3.30)	(2.83)	(3.98)	(0.17)	(1.37)	(4.00)
LT Race Not Reported	-4.889*	-0.310	-6.088*	-0.056	-1.123	-6.283*
	(2.49)	(2.13)	(2.99)	(0.13)	(1.03)	(3.00)

CLASS ES > 5.5	5.119*	6.082**	4.712~	0.156	0.506	7.247*
	(2.24)	(1.92)	(2.70)	(0.12)	(0.93)	(2.91)
CLASS CO > 5.5	4.444*	0.949	2.813	0.127	0.465	0.434
	(1.96)	(1.69)	(2.37)	(0.10)	(0.81)	(2.52)
CLASS IS > 3.5	4.410~	0.594	6.230*	0.138	2.386*	4.223
	(2.55)	(2.19)	(3.08)	(0.13)	(1.06)	(3.20)
CLASS ES	-0.103	-1.482	0.079	-0.031	-0.136	-3.635
	(1.80)	(1.54)	(2.17)	(0.09)	(0.74)	(2.43)
CLASS CO	0.649	0.410	1.430	-0.040	0.477	2.420
	(1.42)	(1.22)	(1.71)	(0.07)	(0.59)	(1.81)
CLASS IS	-3.965**	-0.828	-4.543**	-0.052	-0.957~	-2.014
	(1.27)	(1.08)	(1.52)	(0.07)	(0.52)	(1.71)
Transitions	-0.058	-0.155*	0.012	0.000	0.045	-0.038
	(0.08)	(0.07)	(0.09)	(0.00)	(0.03)	(0.10)
Small Group	-0.329*	-0.102	-0.338*	-0.012	0.033	-0.261
	(0.14)	(0.12)	(0.17)	(0.01)	(0.06)	(0.18)
Choice	-0.036	-0.050	-0.036	-0.000	0.008	0.058
	(0.05)	(0.04)	(0.06)	(0.00)	(0.02)	(0.06)
Oral	-0.086	-0.025	0.062	-0.005	-0.026	-0.092
	(0.08)	(0.07)	(0.10)	(0.00)	(0.03)	(0.11)
Read To	-0.190	-0.225~	-0.052	-0.002	0.018	-0.295
	(0.15)	(0.13)	(0.18)	(0.01)	(0.06)	(0.19)
Word Ident	0.059	0.198*	0.200~	-0.001	0.067	-0.007
	(0.10)	(0.09)	(0.12)	(0.01)	(0.04)	(0.13)
Numbers	0.091	0.073	0.146	0.000	0.043	0.058
	(0.11)	(0.09)	(0.13)	(0.01)	(0.04)	(0.13)
Geometry	0.028	-0.120	0.014	-0.001	0.032	0.007
	(0.09)	(0.07)	(0.11)	(0.00)	(0.04)	(0.11)
Science	-0.111	-0.124~	-0.143	-0.002	-0.003	0.060
	(0.08)	(0.07)	(0.09)	(0.00)	(0.03)	(0.10)
Gross Motor	-0.075	-0.024	-0.050	0.001	0.027	0.054
	(0.09)	(0.08)	(0.11)	(0.00)	(0.04)	(0.11)
Aesthetics	0.243*	0.059	0.288*	0.005	0.007	0.177
	(0.10)	(0.09)	(0.12)	(0.01)	(0.04)	(0.13)
Didactic	-0.000	-0.058	-0.049	-0.006~	0.004	0.106
	(0.06)	(0.05)	(0.07)	(0.00)	(0.02)	(0.07)
N	457	457	457	457	456	397

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are 3-year-olds, Males, White, English, Non-IEP, Creative Curriculum, Star Level 1, Lead Teacher some college and below, Lead Teacher White. Other controls are pre-test, age in months, days between tests, age cohort (3 versus 4 year olds), class size, missing absence data, and hub. Standard scores are used for PPVT, and WJ or WM, T scores are used for C-TRF. Errors are clustered by site. Estimations including self reported teacher race and education were consistent with findings reported here. Estimations including CLASS dimension as thresholds (above cutoff points), also were consistent with findings reported here.

Table C.3. Multivariate analyses of children's 2018-19 posttest (spring) raw scores in relation to child and site or classroom characteristics including the CLASS (scale) and EduSnap.

	Receptive Vocabulary	Literacy	Math	Socio- Emotional (inverted measure)
Female	-2.590*	0.188	-0.396	-2.017
	(1.24)	(0.37)	(0.27)	(1.36)
Af.Am.	-1.641	-1.121~	-1.468**	-1.045
	(2.10)	(0.59)	(0.45)	(2.52)
Hisp.	-1.172	-0.712	-0.694	-3.344
	(2.97)	(0.85)	(0.64)	(3.69)
Other Race/Ethn.	1.056	-0.289	-0.470	-1.961
	(2.63)	(0.76)	(0.57)	(2.91)
DLL	-3.228	-0.597	-0.579	5.417
	(2.85)	(0.83)	(0.61)	(3.40)
IEP	1.378	1.249	-0.611	-4.484
	(2.72)	(0.79)	(0.59)	(3.02)
Moderate Attendance Risk	-1.753	0.022	-0.059	-0.966
	(1.67)	(0.49)	(0.36)	(1.90)
Severe Attendance Risk	-2.842	-1.051	-0.458	-5.718*
	(2.41)	(0.71)	(0.52)	(2.79)
No Attendance Information	-1.049	-1.023~	-0.463	-1.160
	(1.86)	(0.53)	(0.39)	(3.12)
Star 3	13.562*	2.604~	3.698**	-8.281
	(5.51)	(1.56)	(1.15)	(9.01)
Star 4	12.246*	2.562~	3.552**	-10.558
	(5.46)	(1.55)	(1.14)	(8.91)
Creative+	2.029	1.192*	-0.007	-4.327
	(1.96)	(0.56)	(0.41)	(3.31)
Mother Goose	0.682	-0.380	-0.388	1.896
	(2.84)	(0.80)	(0.59)	(6.44)
LT Turnover	-1.389	0.088	-0.301	1.368
	(1.91)	(0.54)	(0.40)	(3.60)
LT Associate	-5.259	-0.757	0.804	4.071
	(3.36)	(0.95)	(0.70)	(5.75)
LT Bachelor	-4.804	-0.582	0.506	-0.976
	(3.39)	(0.96)	(0.71)	(5.70)
LT Master	0.072	-1.157	1.102	5.827
	(3.83)	(1.08)	(0.80)	(6.54)
LT Other	-2.558	-1.337	0.801	7.130
	(4.36)	(1.23)	(0.91)	(7.29)
LT Af.Am.	0.924	0.426	-0.153	0.496
	(2.62)	(0.74)	(0.55)	(4.25)
LT Hisp.	-0.868	-0.017	-1.430*	5.227
	(3.14)	(0.89)	(0.65)	(5.90)
LT Other	-3.674	-0.512	-1.518~	1.261
	(4.22)	(1.20)	(0.88)	(7.11)
LT Race Not Reported	-4.922	0.291	-0.842	-10.783*
	(3.12)	(0.88)	(0.65)	(5.21)
CLASS ES	2.514	0.798	0.663~	-0.519

	(1.90)	(0.54)	(0.40)	(3.62)
CLASS CO	2.666~	-0.083	0.314	2.437
	(1.38)	(0.39)	(0.29)	(2.43)
CLASS IS	-2.745**	-0.308	-0.535*	0.608
	(1.04)	(0.29)	(0.22)	(2.01)
Transitions	-0.070	-0.038	0.008	0.027
	(0.09)	(0.03)	(0.02)	(0.17)
Small Group	-0.415*	0.024	-0.045	-0.646*
	(0.18)	(0.05)	(0.04)	(0.31)
Choice	-0.026	-0.013	-0.012	0.156
	(0.06)	(0.02)	(0.01)	(0.11)
Oral	-0.115	-0.030	0.003	-0.116
	(0.11)	(0.03)	(0.02)	(0.19)
Read To	-0.241	-0.107*	-0.017	-0.051
	(0.19)	(0.05)	(0.04)	(0.34)
Word Ident	0.059	0.059~	0.036	0.086
	(0.13)	(0.04)	(0.03)	(0.22)
Numbers	0.227~	0.011	0.031	-0.012
	(0.13)	(0.04)	(0.03)	(0.22)
Geometry	-0.044	-0.049	0.007	-0.010
	(0.11)	(0.03)	(0.02)	(0.19)
Science	-0.176~	-0.017	-0.032	0.070
	(0.10)	(0.03)	(0.02)	(0.17)
Gross Motor	-0.141	-0.014	-0.008	0.160
	(0.11)	(0.03)	(0.02)	(0.19)
Aesthetics	0.367**	0.029	0.083**	0.419~
	(0.12)	(0.04)	(0.03)	(0.22)
Didactic	-0.034	-0.021	-0.029~	0.072
	(0.07)	(0.02)	(0.01)	(0.13)
N	457	457	457	403

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are 3-year-olds, Males, White, English, Non-IEP, Creative Curriculum, Star Level 1, Lead Teacher some college and below, Lead Teacher White. Other controls are pre-test, age in months, days between tests, age cohort (3 versus 4 year olds), class size, missing absence data, and hub. Standard scores are used for PPVT, and WJ or WM, T scores are used for C-TRF. Errors are clustered by site. Estimations including self reported teacher race and education were consistent with findings reported here. Estimations including CLASS dimension as thresholds (above cutoff points), also were consistent with findings reported here.

Table C.4. Multivariate analyses of children's 2018-19 posttest (spring) raw scores in relation to child and site or classroom characteristics including the CLASS (scale and cutoff) and EduSnap.

	Receptive Vocabulary	Literacy	Math	Socio- Emotional (inverted measure)
Female	-2.507*	0.177	-0.383	-2.020
	(1.24)	(0.36)	(0.27)	(1.36)
Af.Am.	-1.915	-1.093~	-1.535***	-1.203
	(2.08)	(0.59)	(0.45)	(2.53)
Hisp.	-1.455	-0.783	-0.902	-3.948
	(2.98)	(0.86)	(0.64)	(3.72)
Other Race/Ethn.	1.309	-0.343	-0.470	-1.909
	(2.62)	(0.75)	(0.56)	(2.91)
DLL	-3.373	-0.479	-0.525	5.569
	(2.83)	(0.82)	(0.61)	(3.41)
IEP	1.302	1.191	-0.713	-4.728
	(2.72)	(0.79)	(0.58)	(3.03)
Moderate Attendance Risk	-2.231	-0.070	-0.150	-1.281
	(1.67)	(0.49)	(0.36)	(1.91)
Severe Attendance Risk	-3.011	-0.893	-0.478	-5.790*
	(2.41)	(0.71)	(0.52)	(2.80)
No Attendance Information	-2.058	-1.305*	-0.667~	-2.095
	(1.86)	(0.54)	(0.40)	(3.13)
Star 3	11.494*	2.277	3.609**	-9.645
	(5.43)	(1.58)	(1.17)	(9.19)
Star 4	9.669~	2.237	3.345**	-12.357
	(5.38)	(1.57)	(1.16)	(9.05)
Creative+	1.554	1.125*	-0.095	-4.787
	(1.91)	(0.55)	(0.41)	(3.28)
Mother Goose	1.597	-0.656	-0.478	0.176
	(2.85)	(0.83)	(0.61)	(6.60)
LT Turnover	0.152	0.246	-0.153	2.669
	(1.93)	(0.56)	(0.41)	(3.77)
LT Associate	-6.967*	-0.819	0.688	3.137
	(3.30)	(0.96)	(0.71)	(5.80)
LT Bachelor	-6.614*	-0.591	0.385	-1.554
	(3.35)	(0.98)	(0.72)	(5.75)
LT Master	-1.180	-1.172	1.093	5.800
	(3.75)	(1.09)	(0.80)	(6.59)
LT Other	-4.878	-1.407	0.563	5.422
	(4.30)	(1.25)	(0.92)	(7.44)
LT Af.Am.	1.070	0.366	-0.238	0.182
	(2.56)	(0.74)	(0.55)	(4.22)
LT Hisp.	-1.290	-0.044	-1.630*	3.223
	(3.08)	(0.90)	(0.66)	(6.01)
LT Other	-3.793	-0.487	-1.537~	0.560
	(4.09)	(1.19)	(0.88)	(7.04)
LT Race Not Reported	-6.073*	0.228	-1.121~	-12.070*
	(3.09)	(0.90)	(0.66)	(5.27)
CLASS ES > 5.5	5.923*	1.816*	1.133~	7.956
	(2.78)	(0.81)	(0.60)	(5.11)

CLASS CO > 5.5	4.350~	-0.513	0.063	0.884
	(2.43)	(0.71)	(0.52)	(4.40)
CLASS IS > 3.5	4.547	-0.166	1.206~	6.195
	(3.17)	(0.92)	(0.68)	(5.61)
CLASS ES	0.200	-0.074	0.188	-3.756
	(2.23)	(0.65)	(0.48)	(4.26)
CLASS CO	1.520	0.283	0.477	3.048
	(1.76)	(0.51)	(0.38)	(3.12)
CLASS IS	-4.706**	-0.271	-1.012**	-2.049
	(1.57)	(0.46)	(0.34)	(3.02)
Transitions	-0.094	-0.047~	0.009	0.014
	(0.09)	(0.03)	(0.02)	(0.17)
Small Group	-0.442*	0.010	-0.053	-0.669*
	(0.18)	(0.05)	(0.04)	(0.31)
Choice	-0.040	-0.019	-0.015	0.130
	(0.06)	(0.02)	(0.01)	(0.11)
Oral	-0.137	-0.028	0.003	-0.104
	(0.10)	(0.03)	(0.02)	(0.19)
Read To	-0.249	-0.113*	-0.022	-0.131
	(0.18)	(0.05)	(0.04)	(0.33)
Word Ident	0.055	0.044	0.029	0.028
	(0.12)	(0.04)	(0.03)	(0.23)
Numbers	0.156	0.009	0.027	-0.028
	(0.13)	(0.04)	(0.03)	(0.23)
Geometry	0.000	-0.056~	0.008	-0.001
	(0.11)	(0.03)	(0.02)	(0.19)
Science	-0.167~	-0.028	-0.034	0.052
	(0.10)	(0.03)	(0.02)	(0.17)
Gross Motor	-0.105	-0.019	-0.015	0.102
	(0.11)	(0.03)	(0.02)	(0.20)
Aesthetics	0.330**	0.020	0.083**	0.379~
	(0.12)	(0.04)	(0.03)	(0.22)
Didactic	-0.018	-0.020	-0.025~	0.093
	(0.07)	(0.02)	(0.01)	(0.13)
N	457	457	457	403

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are 3-year-olds, Males, White, English, Non-IEP, Creative Curriculum, Star Level 1, Lead Teacher some college and below, Lead Teacher White. Other controls are pre-test, age in months, days between tests, age cohort (3 versus 4 year olds), class size, missing absence data, and hub. Standard scores are used for PPVT, and WJ or WM, T scores are used for C-TRF. Errors are clustered by site. Estimations including self reported teacher race and education were consistent with findings reported here. Estimations including CLASS dimension as thresholds (above cutoff points), also were consistent with findings reported here.