NIEER
NATIONAL INSTITUTE FOR EARLY EDUCATION RESEARCH

EVALUATION OF THE PHILADELPHIA PREK PROGRAM

Year 2 Report

November 2018

Milagros Nores, PhD, W. Steven Barnett, PhD, Zijia Li, PhD, and Mariel Acevedo, 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 a project coordinator at the National Institute for Early Education Research (NIEER). She leads NIEER work on the PHLpreK Evaluation Study and conducts research at on topics related to P-3 evaluations. She has 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.

Suggested citation: Nores, M., Barnett, W.S., Li, Z. & M. Acevedo (2018). Evaluation of the Philadelphia PreK Program. Year 2 Report. New Brunswick, NJ: National Institute for Early Education Research.

Table of Contents

Study Background

This report is the second annual report and makes part of the evaluation study of Philadelphia's PreK Program (PHLpreK), conducted by the National Institute of Early Education Research (NIEER). The evaluation study examines key research questions on the effectiveness of PHLpreK on children's learning outcomes and the overall economy, as well as classroom quality and program design. This report summarizes the 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. The present report is one of the various components of this evaluation meant to be support a data-driven continuous improvement approach to support improvements in quality in the city's program.

Summary of the Literature on Classroom Quality

In the Year 1 report, we described the importance of high-quality preschool education, its potential to reduce the persistent achievement gaps observed in children from minority and low-income families as early as kindergarten and throughout primary (Nores, Francis & Barnett, 2017). We highlighted research that has shown that high-quality preschool education programs can produce lasting effects in school success and achievement and the potential for high-quality preschool programs to produce strong enough results to close half the achievement gap between children from low- and high-income families at kindergarten entry and even stronger results for minorities (Ceci & Papierno, 2005; Barnett, 2008; Duncan & Murnane, 2011; Nores & Barnett, 2015; Camilli et al., 2010, Friedman-Kraus, et al., 2016; Yoshikawa et al., 2013).

Creating, strengthening and supporting high quality programs requires understanding what high-quality preschool means. Small associations between structural features of preschool programs and children's learning gains have led researchers to concentrate more strongly on classroom process and in-service professional development to improve effectiveness (Pianta & Hamre, 2009; Hamre, et al, 2014). Observational measures have therefore become central to the field of early childhood, and are now part of continuous improvement cycles, quality rating systems, and program evaluations as well as key components of Head Start evaluations systems (Martinez-Beck, 2011; U.S. Department of Health and Human Services, Administration for Children and Families, 2010). The relation between child outcomes and the commonly used measures of quality are mostly moderate, yet this association has exhibited stronger at higher levels of process quality, and we take this into account (Burchinal, Kainz & Cai, 2011; Burchinal, et al., 2014; Hatfield, Burchinal, Pianta & Sideris, 2016; Hatfield, et al., 2015 Hestenes, et al., 2015; Weiland & Yoshikawa, 2013). Stronger findings between quality and children's early language and literacy skills are exhibited in language-rich classrooms and there is also evidence on associations with children's behavioral skills and executive functions (Hatfield, et al. 2015; Weiland, Ulvestad, Sachs & Yoshikawa, 2013).

The Philadelphia PreK Program

Philadelphia's preschool program has now finished its second year of operation. In the first year of operation, the program operated in 139 classrooms and home providers in the city. The

program has its roots in the vote of May 19, 2015, when Philadelphians overwhelmingly favored (80 percent) the creation of the Philadelphia Commission on Universal Pre-Kindergarten charged with proposing a universal pre-K program to provides quality, affordable, and accessible services to 3- and 4-year-olds throughout Philadelphia. The National Institute for Early Education Research (NIEER) functions as an external evaluator, conducting a multi-year, multi-site study that employs a combination of methods and designs to assess the program components, program quality, and impacts on children's learning and development. The current report summarizes classroom quality and children's learning outcomes in Year 2 of the PHLpreK program.

Results showed that PHLpreK classrooms are averaging high to moderate levels of quality in the dimensions of emotional support and classroom organization for children. In contrast, classrooms scores are quite low in instructional supports. Scores for all three dimensions of the CLASS measure dropped in relation to the previous year, and this drop was particularly important in emotional and instructional supports. Observations also revealed that classrooms continue to be generally balanced at implementing a variety of activity settings. However, about a fifth of the day is spent in transitions, where there are few opportunities for learning. Children engage in different content areas for some portion of the day, but there is evidence that no learning content occurs for almost a third of the day, albeit a small decrease is observed relative to the previous year. Classrooms exhibit a balance of didactic and scaffolded interactions, though children are rarely asked to explain or justify their thinking through metacognitive processes. Quality scores were examined separately for several subgroups of interest, including star level, number of PHLpreK classrooms, lead teacher credentials, hub, and curriculum. Minimal differences are found between subgroups. Higher quality classrooms exhibit less time in transition, more time with content areas, more scaffolded learning and integrate content more often.

This year we have also assessed children gains, how they differ among different children, and what aspects of centers and teaching and learning, contribute to gains. We find gains are smaller for Black and Hispanic children in some developmental areas relative to their White peers. We also find slight disadvantages for dual language children and IEP children. We do not find consistent and systematic effects across children's learning for the various teacher and center characteristics. We do find evidence of the importance of classroom quality, a smaller use of transitions and whole group, and the time spent on reading and vocabulary. Overall, teachers experience a variety of professional development and technical assistance and perceive it to be useful, yet a relatively small percentage of it is related to classroom quality. The report closes with recommendations of areas for which efforts to strengthen quality and built a system for continuous improvement are of concern and require a careful plan for change be designed and put into action.

Study Methods

The PHLpreK evaluation study is a multi-year, multi-site study that integrates several designs and components to provide a comprehensive assessment of the program's design, its quality, and its impact on children over time. The second year of the study included collection of child, teacher and classroom information to address the following questions:

1. What was the observed quality of children's classroom experiences? Did observed classroom quality improve relative to the prior year?
2. What where the learning gains of children in vocabulary, literacy, math, executive functions and social-emotional development through 2017–18? How did it relate to classroom quality and children's background characteristics?

The PHLpreK evaluation was designed to understand the development of the program over time and its association with children's learning and development. In Year 1, the research team measured classroom quality. In Year 2, the research team measured children's learning and development at the beginning and end of the school year, and repeated the measure of classroom quality. Measures and procedures are described below. Children were assessed in the Fall of 2017, and again at the end of the school year in the Spring of 2018. Classroom observations were conducted to assess teacher-child interactions and quantify the experiences of children through a typical learning day. Classroom observations were conducted between the months of February through May 2018. Quality was assessed using observation protocols widely established in the field, during two visits of approximately two and a half to three hours each.

1. Sample

In the 2017–18 school year, NIEER assessed 465 children in all 139 PHLpreK classrooms (15 which were home-based providers) at pre- and post-test. To recruit children from PHLpreK, children were distributed consent forms as part of the enrollment process across all classrooms and providers in the program. There were 530 children assessed at pre-test and 65 children were lost due to their attrition from the program. A total of 1,689 children had consent from their families to participate in the study. We randomly selected four children per classroom for assessment. The final sample of children was 67% African American, 13% Hispanic, 13% White and 7% Asian, mixed-race, or other (which is somewhat similar to the K-12 PHL school district demographics of 53% African American, 19% Latino, 14% White, and 13% other[1]).

For classroom observations, two instruments were used: CLASS Pre-K and EduSnap (described below). The CLASS was collected in 137 classrooms (both center-based and home-based) in two separate visits. The EduSnap was collected only in the 123 center-based classrooms.[2]

2. Measures

Measures on Classrooms

Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008)

The CLASS is an observational system that assesses classroom practices in preschool and kindergarten by measuring the interactions between students and adults. Observations consist of four to five 20-minute cycles followed by 10-minute coding periods.

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

[2] Two classrooms have not CLASS scores, one because it had not permanent teacher, the other one because of problems with the actual observation performed.

- Scores (codes) are assigned during various classroom activities, and then averaged across all cycles for an overall quality score. Interactions are measured through 10 dimensions, which are divided into three domains. The Emotional Support domain is measured by four dimensions: Positive Climate, Negative Climate, Teacher Sensitivity, and Regard for Student Perspectives. The Classroom Organization domain is measured by three dimensions: Productivity, Behavior Management, and Instructional Learning Formats. The Instructional Support domain is also measured by three dimensions: Concept Development, Quality of Feedback, and Language Modeling. Each dimension uses a 7-point Likert scale.

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

EduSnap is a tool that provides quantification of the experiences that children have throughout the school day. This measure provides an in-depth look at how students experience their day by documenting the actual time students spend in activity settings (e.g., whole group, free choice, transitions), content areas (e.g., reading, science, math), student learning approaches (e.g., collaboration, meta cognition), and teaching approaches (e.g., didactic, scaffolds). Data provide insight into curriculum balance, curriculum integration, and interactions between teachers and children. Observers coded classrooms on EduSnap for three hours in each classroom, starting first thing in the morning.

High-quality classrooms exhibit a balance across activity settings, content areas, and student learning and teaching approaches to best provide all children with a variety of experiences across the school day and within scheduled time blocks. The following information related to EduSnap is meant to help with the interpretation of observation results provided in this report.

- Access to a variety of activity settings 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.
- Frequent 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 would allow children access to multiple content areas and would strengthen learning across and within them.
- Providing children many opportunities to work together and to engage in metacognitive thinking supports both their social/emotional and academic development.
- 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.
- Scaffolded instruction involves teachers asking open-ended questions, engaging in feedback loops, and probing more deeply into children's thinking and understanding. This type of instruction enables the teachers to know specifically how much the children understood from a lesson, identify and remediate group or individual misunderstandings, and engage children in the learning process. Knowing this enables the teacher to respond by modifying the current and subsequent learning experiences and activities according to

the individual and group needs. Both didactic and scaffolded instruction are important teaching styles and should be incorporated in a balanced fashion throughout the course of each school day and within each lesson.

Measures on Children

The *Peabody Picture Vocabulary Test—Fourth Edition (PPVT-IV;* Dunn & Dunn, 2007) is an adaptive test composed of 228-items that measures 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. Frede, et al., 2009; Gormley, 2008; Jung et al., 2013; Ludwig & Phillips, 2008; Peisner-Feinberg, et al., 2014; Weiland & Yoshikawa, 2013).

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).

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 advance 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 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 8 one-tap and 8 two-tap trials in 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 were 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) 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 on 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.

3. Procedures

Trained and reliable observers conducted the observations of classroom quality. Training was provided in administering the observation protocol in two-day workshops. Training for EduSnap© was done by the developer and training for the CLASS by NIEER CLASS Teachstone© certified staff. Home-based classrooms were observed with the CLASS only. Both require the completion of kappa reliability with pre-coded videos online monitored by the developers of these instruments. In addition, collection teams were trained in procedures for conduct and required to complete background checks and human subjects' certification for the conduct of research. All observation score sheets were cleaned and entered at NIEER by trained staff.

Results

We first present the results by instruments, starting with the CLASS, then for EduSnap. A second section focuses on children's gains across a variety of child and center characteristics, and we associate children's gains to process quality. The final section provides our summary of the findings and recommendations.

1. Classroom Observations

CLASS Pre-K Results

Table 2, below, presents aggregate results across all PHLpreK classrooms for all CLASS dimensions and domains. The scoring patterns with instructional support scoring lower than other domains is consistent with the field and other preschool programs, as well as the previous year. Pre-K CLASS mean scores decreased from Spring 2017 from 5.85 to 5.64 for Emotional Support (ES), from 5.34 to 5.28 for Classroom Organization (CO), and from 2.41 to 2.05 for Instructional Support (IS). The fall is statistically significant for CLASS ES and CLASS IS. A discussion of each domain is subsequently presented.

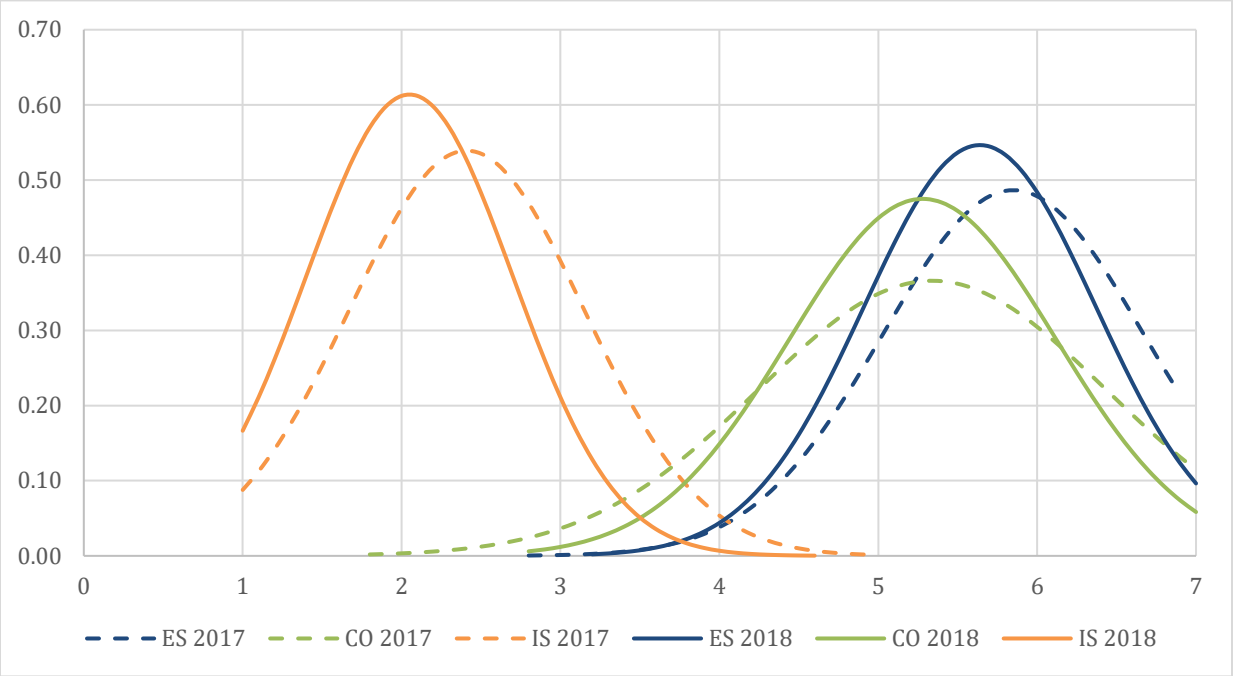
Table 2. PreK CLASS Dimension and Domain Means and Ranges.

CLASS Dimensions and Domains	2017 Mean (Range) N=139	2018 Mean (Range) N=137
Emotional Support Domain (ES)	5.85 (2.85-6.90)	5.64[a] (3.20-6.95)
1. Positive Climate	5.90 (1.60-7.00)	5.73 (3.20-7.00)
2. Negative Climate*	6.77 (5.00-7.00)	6.67 (4.00-7.00)
3. Teacher Sensitivity	5.69 (2.20-7.00)	5.52 (2.80-7.00)
4. Regard for Student Perspectives	5.03 (2.00-6.80)	4.65 (2.40-7.00)
Classroom Organization Domain (CO)	5.34 (1.87-6.93)	5.28 (2.80-6.93)
5. Behavior Management	5.49 (1.60-7.00)	5.48 (2.80-7.00)
6. Productivity	5.76 (1.80-7.00)	5.65 (2.80-7.00)
7. Instructional Learning Formats	4.77 (1.60-7.00)	4.72 (1.80-6.80)
Instructional Support Domain (IS)	2.41 (1.00-5.00)	2.05[a] (1.00-4.60)
8. Concept Development	2.09 (1.00-4.80)	1.84 (1.00-4.00)
9. Quality of Feedback	2.23 (1.00-5.00)	1.91 (1.00-4.40)
10. Language Modeling	2.91 (1.00-5.20)	2.41 (1.00-5.60)

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

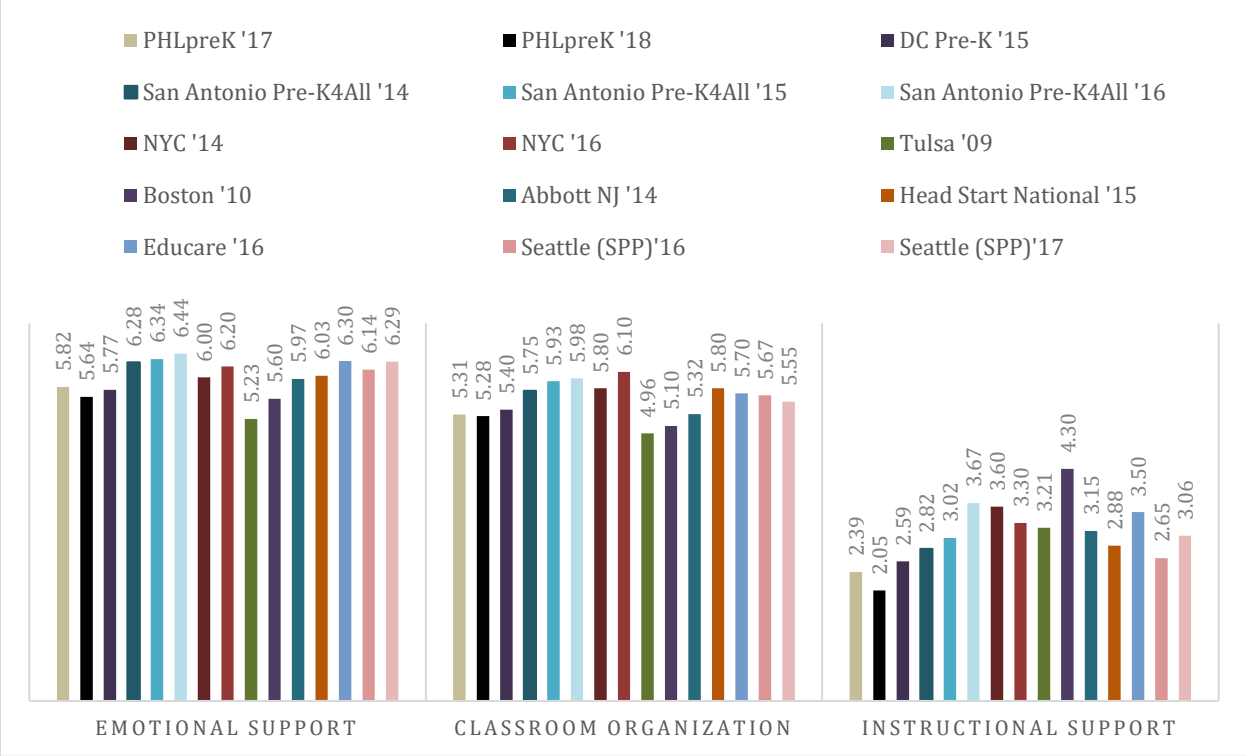
Figure 1, below, depicts the change in the distribution of scores between Year 1 and Year 2. There is a fall in CLASS ES scores with a higher concentration of scores along a lower mean for Year 2. For CLASS CO there seems to have been an improvement in some classrooms making the distribution have a shorter tail along the lower scores, however there is a concentration along a lower mean as well. This is the same case for CLASS IS. Some research appears to support (Burchinal et al. 2009; OPRE, 2010) thresholds for ES and CO above 5 and IS above 3 as necessary for a relation between quality and children's outcomes to be evidenced (other research defines these as slightly higher, at 5.5 and 3.5). In PHLpreK, 81.3% of the classrooms were above these thresholds in ES (down from 85.6%), 65.0% were above them in CO (down from 72.7%), and only 9.8% were above the threshold of 3 in IS (down from 19.4%).

Figure 1. Distribution of CLASS Scores by Domain 2017 & 2018.



The score patterns for the PHLpreK follow those of the National Overview of CLASS in pre-K classrooms of 2015 (OHS, 2015), which reported higher scores in ES (national mean of 6.03), followed by CO (national mean of 5.80) and with the lowest scores in the IS (national mean of 2.88). The PHLpreK CLASS from 2017 and 2018 scores together with those from various other programs in the U.S. are shown in Figure 2 for comparison purposes, including high-quality city programs and how they have evolved in their first few years in terms of quality. In contrast to the trends observed for PHLpreK, the SPP program in Seattle, the PreK4All program in San Antonio and the Pre-K for All program in NYC have all shown increases in CLASS scores in their first few years.

Figure 1. Comparison of PHLpreK CLASS scores with other programs.



Emotional Support Domain. The overall mean score for ES is 5.64 (SD 0.73), putting it in the high end of the mid range. The minimum score is 3.20, which indicates there were no classrooms in which there was a very low level of emotional support throughout all five cycles observed. This remains the same relative to the previous year. The highest scoring dimension is Negative Climate, with a mean of 6.67, indicating classrooms mostly exhibited very few negative interactions between teachers and children and also among children. The lowest scoring dimension remains Regard for Student Perspectives, with a mean of 4.65. This scoring reflects teachers sometimes showing flexibility and giving students responsibility, yet also sometimes restricting students' movement and choices throughout the day.

Classroom Organization Domain. The overall mean score for the Classroom Organization Domain is 5.28 (SD 0.73). The two higher scoring dimensions were Behavior Management and Productivity with means of 5.48 and 5.65, correspondingly. These high scores indicate teachers show effective methods to both prevent and redirect misbehavior, and most student behavior observed being consequently compliant and appropriate. In addition, teachers were observed managing their instructional time well, with little time wasted. This does not consider the quality of activities but rather that activities are available and planned.

Instructional Support Domain. IS assesses the interactions through which teachers deliver and facilitate high-order thinking skills, and develop language. This domain is the most difficult, yet critically central, when considering teacher practices that have impacts on children's learning. The mean score for this domain is 2.05 (SD 0.65) with averages ranging from 1 to 5 on a 7-point scale. The lowest scoring areas are Concept Development and Quality of Feedback, which focus on problem solving, prediction, experimentation, and classification, as well as children's creative processes and the planning that goes along with it, and the connection to concept, knowledge, and the real world. Quality of Feedback focuses on teachers' use of hints

and assistance to support children's understanding of a concept, as well as feedback loops (back and forth exchanges and follow-up questions), asking students to explain their thinking and expanding on their thought process. Scores for the Language Modeling subdomain were higher, but the mean was still below the "minimal" threshold. Having said this, this domain generally scores quite low across all programs, as shown in Figure 1.

Table 3 presents CLASS results for subgroups of interest. The patterns resemble those of the overall sample, with ES scoring higher, followed by CO scores, and IS scores scoring quite low regardless of the grouping. Patterns that emerge more clearly are that centers with a STAR rating of 2 or 4 score higher across all three domains[3], and CDA lead teachers scoring the lowest. Those with under various curriculum (other) evidenced the lowest scores. Success by 6 classrooms do evidence lower scores as well, which would be expected if the program is effectively targeting programs most in need.

Table 3. CLASS domains scores by subgroups, N = 137.

		CLASS Mean Scores		
		Emotional Support	Classroom Organization	Instructional Support
STAR Level	1-2 (n=28)	5.81	5.43	2.09
	3 (n=62)	5.49	5.15	1.96
	4 (n=47)	5.74	5.37	2.15
Number of PHL preK classrooms	1 (n=55)	5.63	5.24	1.93
	2 (n=40)	5.74	5.41	2.18
	3 (n=18)	5.66	5.36	2.34
	4 or more (n=24)	5.48	5.12	1.89
Lead Teacher Credential	CDA/ECE Course (n=5)	5.37	4.94	1.99
	AA (n=48)	5.69	5.29	2.00
	BA (n=44)	5.52	5.27	2.16
	MA (n=16)	5.75	5.36	2.09
	Missing (n=24)	5.75	5.31	1.93
PHLpreK Partner Agency	District 1199C (n=13)	5.48	5.10	1.97
	Phila SD (n=11)	5.97	5.68	1.81
	PHMC (n=54)	5.72	5.37	2.12
	UAC (n=59)	5.55	5.17	2.05
Curriculum	Creative (n=79)	5.65	5.25	1.99
	Creative + (n=19)	5.59	5.35	2.17
	Mother Goose (n=20)	6.09	5.71	2.55
	Other (n=19)	5.42	5.17	1.90
Success by 6	Yes (n=15)	5.28	5.03	1.88
	No (n=122)	5.69	5.31	2.07

[3] Across this and the following tables, 4 includes a district unrated center.

EduSnap results

EduSnap results are presented for the aggregate sample and also separately for the following subgroups: STAR rating level, number of classrooms per site, teacher credentials, PHLpreK partner agency, council district, and curriculum.

Activity Settings

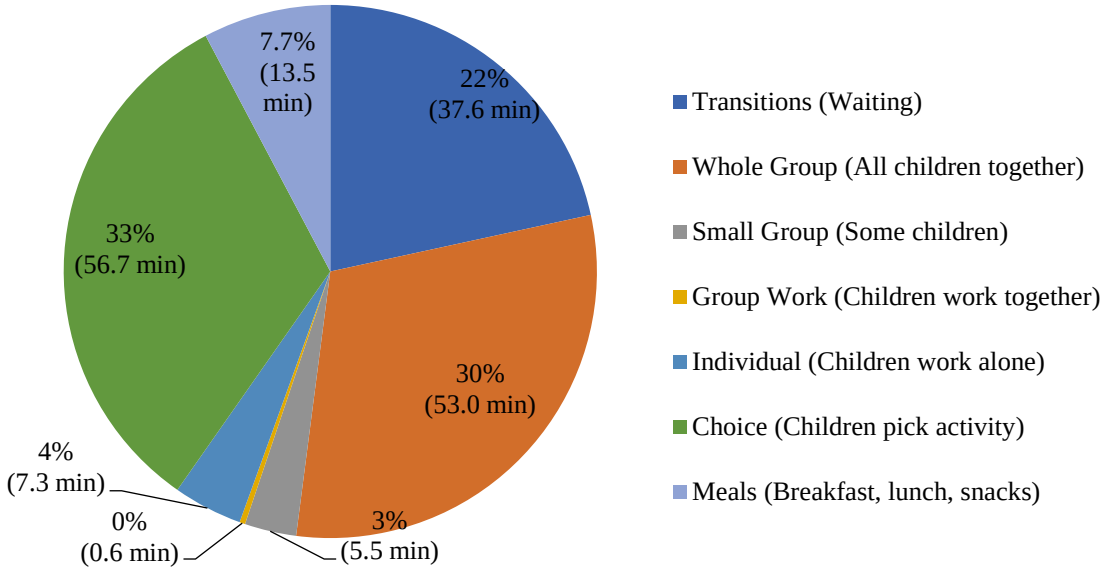
The first section of the EduSnap utilizes codes to account for 100% of the day. Coding in this portion creates descriptive information about the percentage of time during the observation that children are engaged in a particular activity setting. Activity settings are defined as follows:

- Transitions: Children are moving or waiting between locations or activity settings
- Whole Group: Children are engaged in teacher-led activities (50% or more of the children in the class)
- Small Group: Children are engaged in teacher led activities (Less than 50% of the children in the class)
- Group Work: Children are engaged in joint assignments that are not teacher-led
- Individual: Children work on individual assignments
- Choice: Children are engaged in activities they selected from a variety of unassigned options.
- Meals: Children are eating meals (breakfast or lunch) or snack.

A way of looking at children's experiences of the day is thinking how children experience activity settings as they occur on a minute-by-minute basis (Figure 3). The time spent in choice (33%) and whole group (30%) are on average well balanced, while the time spent on transitions (22%) is high and the time spent in small groups (3%) is quite low. Group work activities are geared more toward children in later grades so the percentages for this pre-K sample are of less importance. This is quite unchanged relative to the observations of 126 classrooms in Philadelphia in the spring of 2017 which found on average 27% of the time spent on whole group, 21% on transitions, and 33% on choice time. There was a slight decrease of 1-2% each on small group, group work and individual work time.

Converting time to minutes is helpful in interpreting how these percentages translate into teaching and learning throughout the day. One percent is on average equivalent to 1.7 minutes, five percent to 8.5 minutes, ten percent to 17 minutes, and so forth. Figure 3, below, includes minutes following this scale. The minutes are calculated based on the average time across observations which was 174 minutes, nearly three hours.

Figure 3. Percentages of Time and Minutes Spent in Activity Settings, n=123.



Patterns across subgroupings of programs by STAR ratings, number of contracted classrooms, curriculum, teacher qualifications or success by 6, are shown in Table 4. No strong patters emerge. Transitions are around 20% on average and up to about 25-29% in some cases without any clear patterns. Whole group activities vary between 20 and 37% while small group activities are generally low across the board (around 5%). MA/ME teachers appear to use small groups slightly more than other teachers. Choice is around 30-35% for most types of centers, and slightly more observed in classrooms with a teacher with a BA.

Table 4. Percentages of time spent in Activity Settings by Subgroups, n=123.

		% of Time						
		Meals	Transitions	Whole Group	Small Group	Individual	Choice	Group Work
STAR Ratings	1-2 (n=28)	8.22	22.98	27.58	1.92	34.49	0.06	8.22
	3 (n=56)	8.24	23.92	28.18	2.88	30.88	0.16	8.24
	4 (n=39)	6.75	17.39	35.68	4.32	33.16	0.79	6.75
Number of PHL preK classrooms	1 (n=39)	7.88	20.13	33.81	3.17	3.89	30.90	0.23
	2 (n=40)	5.79	21.03	26.57	2.41	5.10	38.97	0.14
	3 (n=20)	9.25	22.38	32.76	4.19	0.87	30.52	0.03
	4 or more (n=24)	9.63	24.47	29.38	3.34	6.49	25.58	1.10
Lead Teacher Credential	CDA/ECE (n=4)	6.42	18.82	40.79	0.70	7.64	25.63	0.00
	AA (n=43)	9.01	22.69	29.82	2.45	4.03	31.89	0.11
	BA/BS (n=40)	6.09	21.62	29.56	3.44	3.64	35.16	0.49
	MA/ME (n=13)	6.68	19.43	32.57	6.51	5.46	29.32	0.04
	Missing (n=23)	9.19	21.42	30.03	2.32	4.71	31.60	0.73
PHLpreK Partner Agency	Phila SD (n=11)	8.50	18.77	32.45	1.51	4.13	34.39	0.26
	PHMC (n=52)	6.81	21.95	34.81	4.26	2.50	29.06	0.62
	UAC (n=60)	8.46	21.89	26.25	2.42	5.89	34.98	0.10
Curriculum	Creative (n=78)	8.39	21.02	28.14	3.10	3.98	34.89	0.47
	Creative + (n=17)	5.07	19.47	36.46	3.80	4.85	30.13	0.23
	Mother Goose (n=11)	10.77	28.86	30.62	1.77	5.45	22.48	0.05
	Other (n=17)	5.67	21.94	34.70	3.38	4.45	29.83	0.03
Success By 6	Yes (n=15)	7.44	21.71	31.46	3.15	3.73	32.28	0.23
	No (n=108)	7.81	21.62	30.28	3.11	4.38	32.44	0.35

A second lens provided by the EduSnap is the percentage of time spent in different components of literacy (Figure 4). While different programs may be stronger in different aspects of literacy, or preferably balanced across these, what is significant relative to the 2017 observations is the large increase in the time spent on oral language (which increased from 12.8% in 2017). Although all components of literacy are very low, it is worth noting the particular emphasis on time spent on word identification (9.5% or about 16 minutes) and being read to (5.9% or about 10 minutes) and in contrast, the very little time spent on vocabulary (1.9%), reading (1.3%), or writing (1%, or averaging about one minute per day). Of these, the time spent on oral language, writing, vocabulary, word identification, and being read to increased relative to 2017. The other areas decreased slightly. Oral language includes children speaking with teachers on either knowledge, ideas, and/or feelings. It does not require high quality language to occur, and the CLASS instructional supports indicates low quality language modeling.

Figure 4. Percentages of Time and Minutes Spent in Components of Literacy.

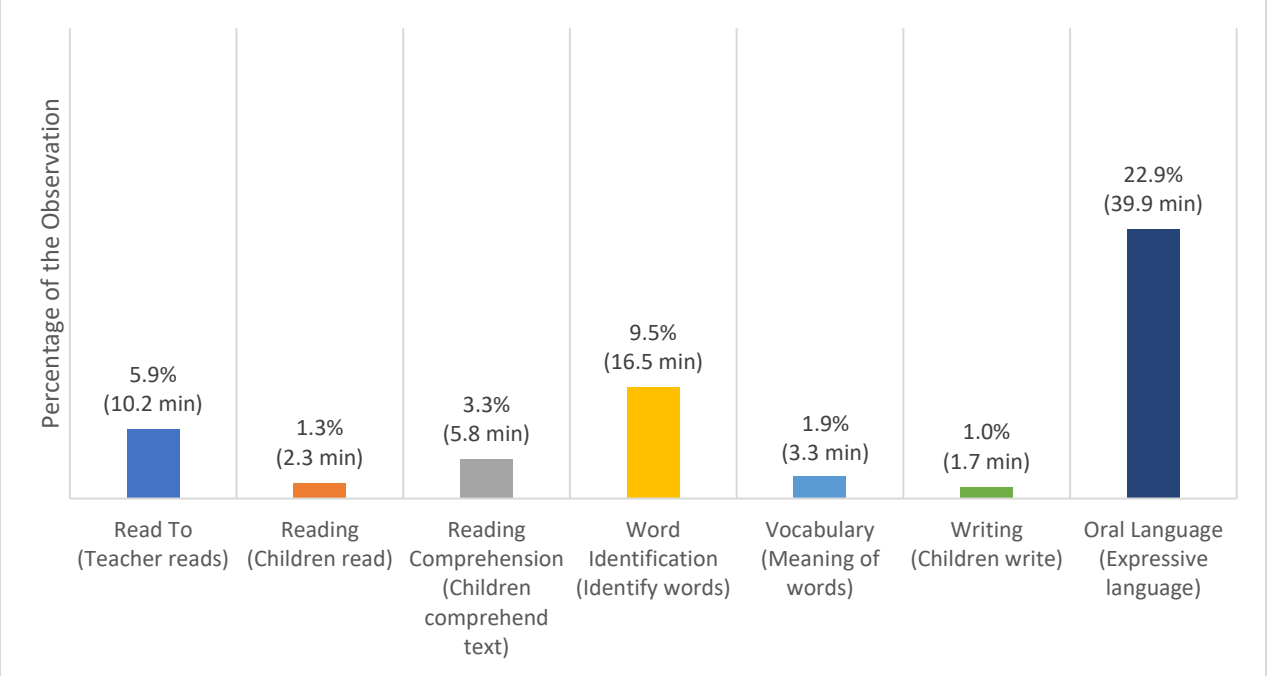


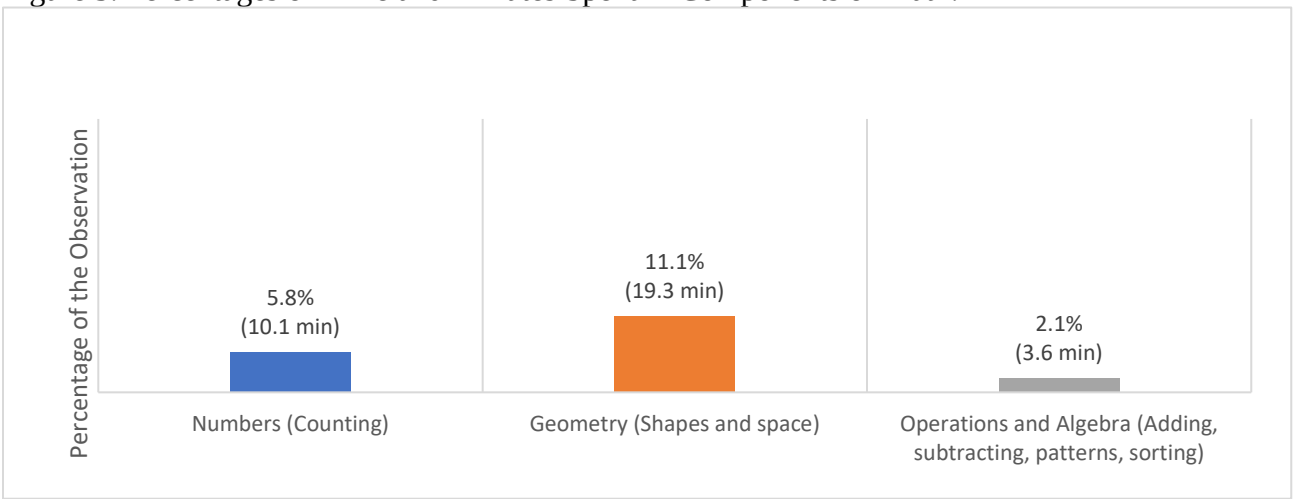
Table 5 reports these percentages for the subgroup of interests. It stands out the degree to which oral language is much higher in star 4 rated classrooms. Word identification is very low in the unrated/rated 1 and 2 classrooms, as are reading and writing (although this is generally low across the board).

Table 5. Percentages of time spent in Components of Literacy by Subgroup, n=123.

		Literacy Activities						
		Read-to	Reading	Compreh.	Word ID	Vocabulary	Writing	Oral Language
STAR Ratings	1-2 (n=28)	6.78	1.40	3.69	8.83	1.53	0.60	22.87
	3 (n=56)	5.26	1.23	3.01	9.08	1.54	1.06	17.68
	4 (n=39)	6.11	1.37	3.49	10.40	2.74	1.08	29.71
Number of PHL preK classrooms	1 (n=39)	6.07	1.49	3.66	8.65	11.57	0.25	27.16
	2 (n=40)	4.78	1.15	2.37	9.05	1.37	1.29	16.94
	3 (n=20)	7.36	1.45	5.11	9.22	1.37	1.30	28.80
	4 or more (n=24)	6.15	1.20	2.84	11.57	2.46	1.30	19.84
Lead Teacher Credential	CDA/ECE (n=4)	7.31	0.42	4.81	8.14	1.13	0.00	18.85
	AA (n=43)	5.54	1.29	3.32	9.79	2.92	0.57	26.88
	BA/BS (n=40)	5.52	1.22	2.69	8.76	1.44	0.83	21.67
	MA/ME (n=13)	6.21	1.36	4.20	11.92	1.99	2.61	14.14
	Missing (n=23)	6.69	1.67	3.66	8.79	1.00	1.17	22.07
PHLpreK Partner Agency	Phila SD (n=11)	6.85	2.06	4.66	6.99	2.16	0.25	32.63
	PHMC (n=52)	6.37	1.10	3.61	10.19	2.26	1.14	24.18
	UAC (n=60)	5.27	1.37	2.82	9.24	1.59	0.94	19.55
Curriculum	Creative (n=78)	5.39	1.27	3.05	8.67	2.16	0.98	23.93
	Creative + (n=17)	5.81	0.79	3.61	10.41	1.32	1.29	21.43
	Mother Goose (n=11)	4.53	1.37	2.07	8.28	2.27	0.31	23.25
	Other (n=17)	9.05	2.02	5.05	12.79	1.19	0.96	17.78
Success By 6	Yes (n=15)	5.41	0.99	2.16	8.80	1.76	0.50	24.90
	No (n=108)	5.94	1.36	3.48	9.53	1.94	1.03	22.37

Similarly, the EduSnap looks at the percentage of time spent in three specific components of math: numbers, geometry, and operations and algebra (Figure 5). It is notable that all components of math are extremely low. Relative to the 2017 report on Philadelphia, a slight increase is observed in the overall time spent in numbers and geometry, and a decrease in the time spent in operations and algebra.

Figure 5. Percentages of Time and Minutes Spent in Components of Math.



Percentages of time spent in components of math for each subgroup of interest are reported in Table 6. BA and MA teachers spend just a little more time on numbers, while teachers with a CDA spend a little more time on geometry and all teachers spend very little time on algebra. Creative plus classrooms seem to spend a slightly higher amount of time in numbers and algebra (and classrooms with the Mother Goose Time curriculum evidence very lower percentages of time across all three, particularly geometry and algebra). Overall though, differences are very small and there are no clearly distinguishable patterns.

Table 6. Percentages of time spent in components of math by subgroups, n=123.

		% of Time		
		Numbers	Geometry	Algebra
STAR Ratings	1-2 (n=28)	4.93	12.47	2.17
	3 (n=56)	5.33	10.38	2.09
	4 (n=39)	7.14	11.08	1.96
Number of PHL preK classrooms	1 (n=39)	5.93	11.25	2.80
	2 (n=40)	4.92	12.35	2.46
	3 (n=20)	6.12	10.17	1.46
	4 or more (n=24)	6.86	9.44	1.84
Lead Teacher Credential	CDA/ECE (n=4)	3.80	11.38	1.44
	AA (n=43)	5.72	11.15	1.79
	BA/BS (n=40)	6.22	11.63	2.64
	MA/ME (n=13)	6.58	10.09	2.30
	Missing (n=23)	5.21	10.49	1.57
PHLpreK Partner Agency	Phila SD (n=11)	5.52	9.91	1.26
	PHMC (n=52)	7.08	10.64	2.53
	UAC (n=60)	4.77	11.68	1.82
Curriculum	Creative (n=78)	5.65	11.32	1.89
	Creative + (n=17)	6.59	11.72	3.60
	Mother Goose (n=11)	5.08	7.22	0.83
	Other (n=17)	6.28	11.82	2.16
Success By 6	Yes (n=15)	5.09	13.00	1.97
	No (n=108)	5.90	10.81	2.08

The EduSnap also considers the importance of content area balance, and therefore captures the percentage of time that children spent in each content area (Figure 6). Similar to the 2017 report on Philadelphia, the percentage of time spent on literacy (34.8%) dominates over any other content areas, e.g., doubling the time spent in math. This translates into about an average of 61 minutes of the 2 hours and 54 minutes observed on average across classrooms. Science and gross motor remain as the two content areas with the lowest percentages of time observed. The percentage of time spent in all content areas increased, particularly for literacy (25.1% to 34.8%), social studies (16.0% to 19.8%), and aesthetics (14.5% to 17.2%). The language dominance is driven by Oral Language. In this aspect, the CLASS instructional support scores clarify that this quantity does not include necessarily quality in language modelling

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

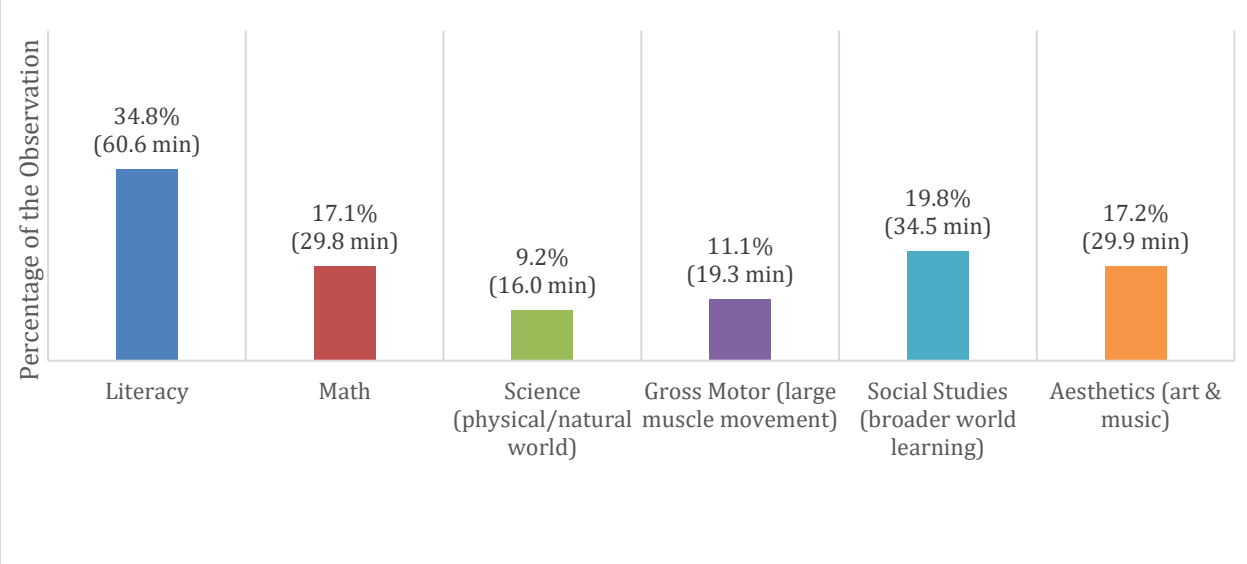


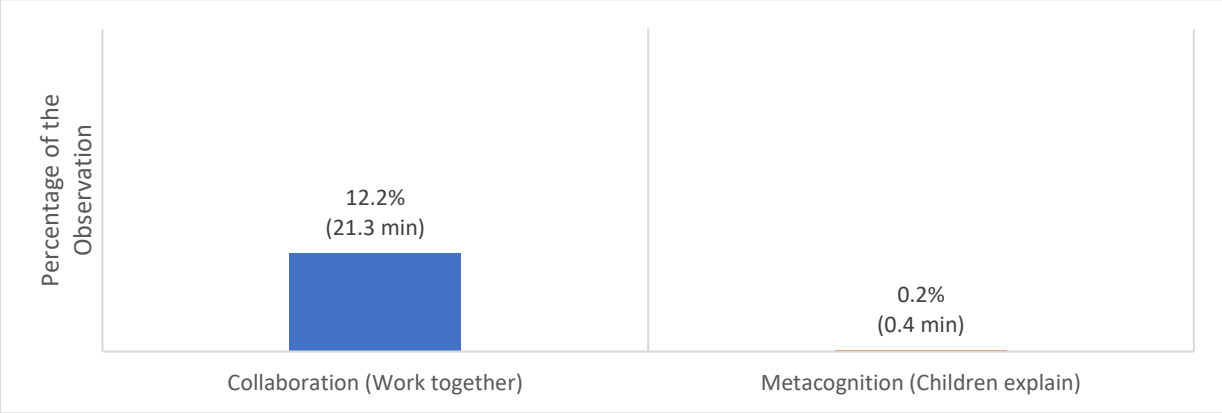
Table 7 below reports the percentage of time spent in the different content areas for each subgroups of interest. Mother Goose Time classrooms spend the lowest percentages of time in math and science content. STAR 4 rated classrooms spend more time on literacy and aesthetics than other classrooms but this is not the case for other content areas.

Table 7. Percentages of time spent in various content areas by subgroups, n=123.

		% of Time					
		Literacy	Math	Science	Gross Motor	Social Studies	Aesthetics
STAR Ratings	1-2 (n=28)	34.91	17.21	10.06	10.36	21.81	15.31
	3 (n=56)	29.80	16.19	8.60	10.95	18.95	15.87
	4 (n=39)	41.29	18.29	9.26	11.62	19.57	20.33
Number of PHL preK classrooms	1 (n=39)	37.57	17.33	10.32	11.26	19.88	17.51
	2 (n=40)	29.14	17.89	8.68	12.34	19.37	16.14
	3 (n=20)	40.01	16.18	10.22	12.17	21.74	15.37
	4 or more (n=24)	34.40	16.11	7.09	7.49	18.76	19.76
Lead Teacher Credential	CDA/ECE (n=4)	29.97	15.92	11.98	8.29	11.18	24.78
	AA (n=43)	37.70	16.61	9.47	11.00	20.44	18.05
	BA/BS (n=40)	33.04	18.44	9.45	12.83	21.24	16.18
	MA/ME (n=13)	29.64	17.54	7.87	9.48	17.72	17.88
	Missing (n=23)	35.18	15.56	8.21	9.28	18.77	15.46
PHLpreK Partner Agency	Phila SD (n=11)	40.41	15.38	11.06	12.61	16.30	13.03
	PHMC (n=52)	36.58	18.01	9.39	12.10	18.72	19.38
	UAC (n=60)	31.84	16.59	8.57	9.80	21.38	15.98
Curriculum	Creative (n=78)	34.50	17.08	9.77	10.42	19.47	16.77
	Creative + (n=17)	34.59	19.79	9.06	12.94	21.57	20.19
	Mother Goose (n=11)	34.24	12.14	6.63	10.84	21.13	16.45
	Other (n=17)	35.38	17.58	7.93	12.00	18.70	16.34
Success By 6	Yes (n=15)	35.48	17.89	6.27	11.68	23.04	16.43
	No (n=108)	34.49	16.97	9.54	10.93	19.35	17.26

EduSnap also provides a lens to the student learning approaches *collaboration* and *metacognition* (Figure 7). The collaboration code considers situations when children are working together in activities where they are sharing ideas, completing an assignment, or solving a problem. The metacognition code includes instances where children are encouraged to provide evidence for their ideas or explain their answers. Collaboration was scarcely observed (and it decreased from 13% in the previous year to 12%) and metacognition was again absent.

Figure 7. Percentages of Time and Minutes Spent in Student Learning Approaches.



Student learning approaches by type of setting are summarized in Table 8. Significantly, the absence of metacognition is true for all providers, star levels, teacher credentials agencies and curriculum approaches. Collaborative approaches vary between 9% and 17%, and are higher classrooms with teachers with a BA/BS and a 4-star rating.

Table 8. Percentages of time spent in student learning approaches by subgroups, n=123.

		Student Learning Approaches	
		Collaboration	Metacognition
STAR Ratings	1-2 (n=28)	10.97	0.10
	3 (n=56)	9.30	0.22
	4 (n=39)	17.13	0.28
Number of PHL preK classrooms	1 (n=39)	12.88	0.31
	2 (n=40)	13.73	0.09
	3 (n=20)	10.75	0.17
	4 or more (n=24)	9.56	0.29
Lead Teacher Credential	CDA/ECE (n=4)	11.02	0.28
	AA (n=43)	11.42	0.28
	BA/BS (n=40)	13.22	0.04
	MA/ME (n=13)	9.99	0.51
	Missing (n=23)	13.12	0.20
PHLpreK Partner Agency	Phila SD (n=11)	13.63	0.21
	PHMC (n=52)	12.83	0.27
	UAC (n=60)	11.32	0.16
Curriculum	Creative (n=78)	14.01	0.21
	Creative + (n=17)	10.18	0.24
	Mother Goose (n=11)	7.50	0.36
	Other (n=17)	8.66	0.10
Success By 6	Yes (n=15)	10.50	0.23
	No (n=108)	12.39	0.21

The EduSnap also captures the teaching and learning approaches utilized by teachers to engage children. Specifically, these include *didactic* and *scaffolds*. The percentages of time for these categories are presented in Figure 8. Classrooms in the sample tend to weigh toward didactic approaches over scaffolded approaches. This was the case in 2017 as well, although this year the use of didactic approaches and scaffolded approaches have increased relative to the 2017 report on Philadelphia. Didactic approaches increased from 35% to 40% and scaffolded approaches from 25% to 32%.

Figure 8. Percentages of Time and Minutes Spent in Teaching Approaches.

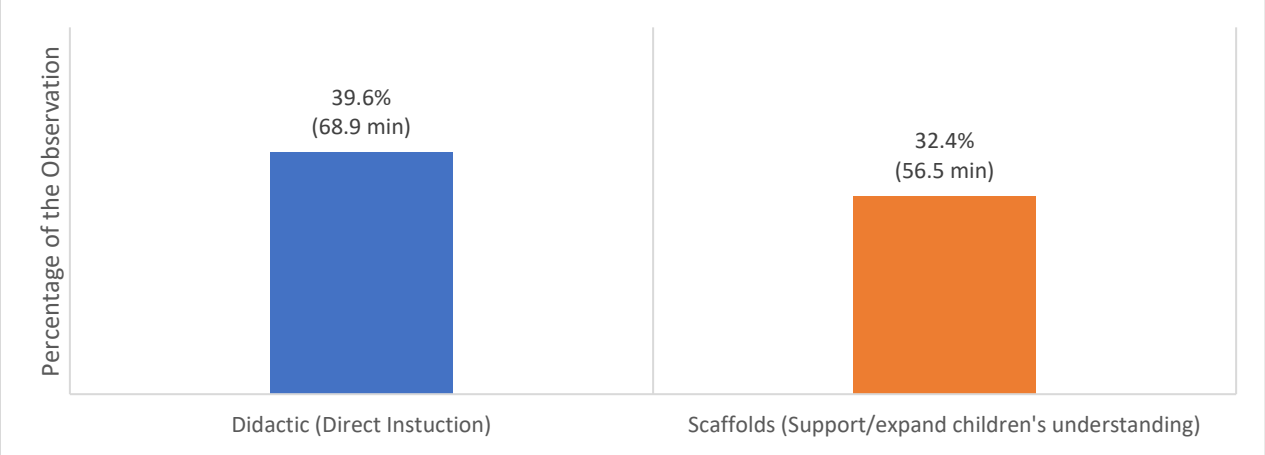


Table 9 describes the percentage of time spent in didactic or scaffolded approaches for each subgroup of interest. Higher percentage of scaffolding or didactic approaches are not observed in higher star level classrooms

Table 9. Percentages of time spent in Teaching Approaches by Subgroup, n=123.

		Teaching and Learning approaches	
		Didactic	Scaffolds
STAR Ratings	1-2 (n=28)	39.24	31.57
	3 (n=56)	43.98	27.51
	4 (n=39)	34.12	39.44
Number of PHL preK classrooms	1 (n=39)	37.28	36.72
	2 (n=40)	40.63	25.99
	3 (n=20)	37.22	32.20
	4 or more (n=24)	44.55	30.30
Lead Teacher Credential	CDA/ECE (n=4)	41.53	30.94
	AA (n=43)	37.31	36.82
	BA/BS (n=40)	40.21	30.66
	MA/ME (n=13)	46.16	24.85
	Missing (N=23)	39.72	30.70
PHLpreK Partner Agency	Phila SD (n=11)	32.89	39.30
	PHMC (n=52)	41.35	34.34
	UAC (n=60)	39.67	29.08
Curriculum	Creative (n=78)	32.83	37.07
	Creative + (n=17)	32.71	44.10
	Mother Goose (n=11)	31.99	47.73
	Other (n=17)	29.07	42.73
Success By 6	Yes (n=15)	34.61	32.49
	No (n=108)	40.49	32.18

In terms of content, the EduSnap also considers the importance of curriculum integration across areas (Figure 9). In the sample, classrooms evidenced no content 28% of the time (which translates into about 50 minutes), no integration 43% of the time (one content area occurring about 75 minutes), and integration of two or more content areas about 28% of the time. Relative to the 2017 report on Philadelphia, an increase in integration of two or more content areas is apparent (in 4% points), as well as a decrease in time at which no content is being carried out (from 33%).

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

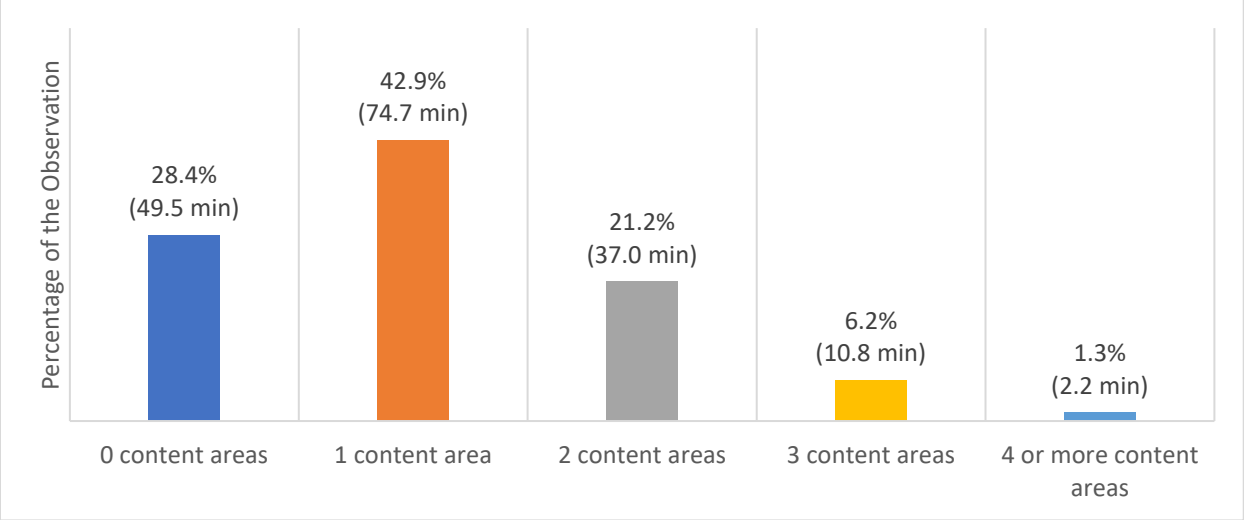


Table 10 reports the percentages of time of curriculum integration for none, one, or more areas being integrated across each subgroup of interest. The results align with the overall results. About 20-30% of the time there is no integration in most classrooms and about 40% of the time there is one content area. Integration in two content areas varies between 18-23% of the time. Integration in three content areas is quite low and is less than 7% of the time (about 12 minutes). Four-star classrooms evidence slightly higher percentages of integration of two or three areas.

Table 10. Percentages of time spent in Curriculum Integration by Subgroup, n=123.

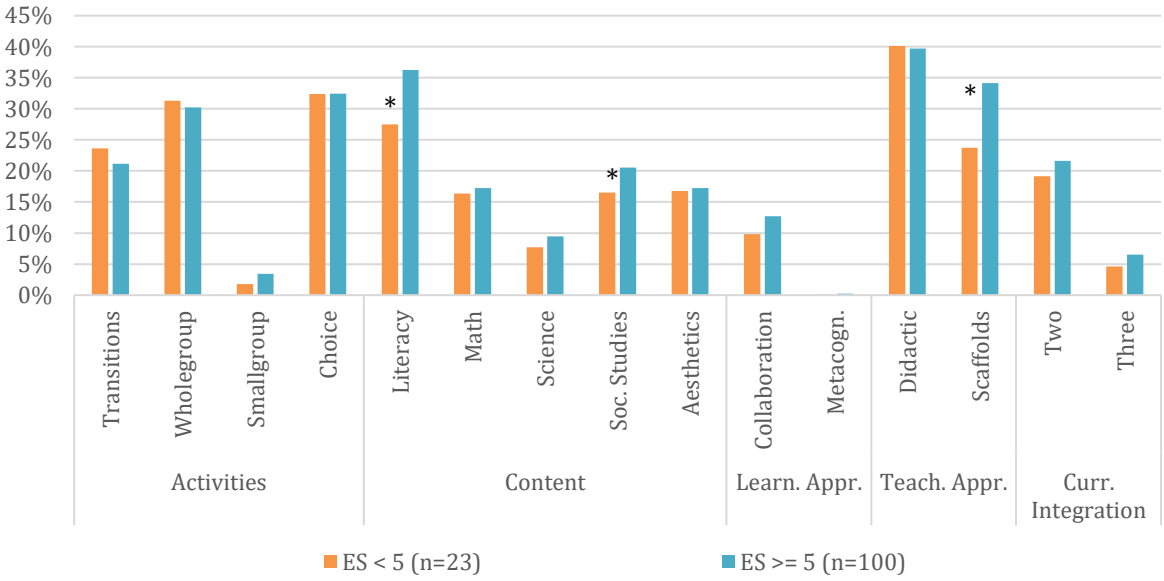
		% of Time				
		None	1	2	3	4 or more
STAR Ratings	1-2 (n=28)	28.40	43.37	20.04	6.76	1.43
	3 (n=56)	32.55	41.71	19.64	5.13	0.96
	4 (n=39)	22.99	44.12	24.19	7.17	1.53
Number of PHL preK classrooms	1 (n=39)	28.37	40.12	22.84	6.90	1.77
	2 (n=40)	27.90	46.82	19.69	5.00	0.58
	3 (n=20)	26.43	42.03	22.59	7.42	1.53
	4 or more (n=24)	31.80	41.37	19.77	5.77	1.29
Lead Teacher Credential	CDA/ECE (n=4)	31.29	41.21	22.02	5.07	0.42
	AA (n=43)	28.11	41.07	22.13	7.05	1.64
	BA/BS (n=40)	26.56	44.30	21.79	6.22	1.12
	MA/ME (n=13)	31.12	44.74	18.50	4.48	1.16
	Missing (N=23)	31.02	42.88	19.68	5.48	0.93
PHLpreK Partner Agency	Phila SD (n=11)	30.81	40.32	20.84	5.74	2.29
	PHMC (n=52)	26.06	43.19	22.73	6.74	1.30
	UAC (n=60)	30.35	43.03	19.89	5.72	1.02
Curriculum	Creative (n=78)	29.60	42.09	20.58	6.35	1.38
	Creative + (n=17)	21.76	46.97	23.69	6.63	0.95
	Mother Goose (n=11)	34.50	38.01	20.11	6.51	0.86
	Other (n=17)	26.85	45.39	22.05	4.51	1.20
Success By 6	Yes (n=15)	28.09	42.86	20.87	6.74	1.44
	No (n=108)	28.64	42.85	21.22	6.07	1.22

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

Lastly, we assessed children's experiences through the day for higher scoring classrooms in the three domains of CLASS versus lower scoring classrooms. This is meant to understand whether the balance in activities, content, integration in curriculum, and teaching and learning in higher quality classrooms differs relative to lower quality classrooms.

Figure 10 illustrates selected average EduSnap percentages for lower versus higher scoring classrooms in Emotional Support. We defined lower versus higher scoring classrooms in Emotional Support as those below and above 5, respectively. Statistically significant differences are marked with an asterisk. Worth highlighting is that: higher ES quality classrooms also evidenced more time in literacy activities (about 16 minutes), more time in social studies (about 7 minutes), and more in scaffolded activities (18 minutes).

Figure 10. EduSnap average percentages grouped by CLASS Emotional Support levels



We repeated this exercise with the CLASS Classroom Organization Domain (Figure 11). We defined lower versus higher scoring classrooms again as those below and above 5, respectively. Statistically significant differences are also marked with an asterisk. Higher CO quality classrooms also evidenced less time in transitions (about 6 minutes), more time in literacy activities (about 23 minutes), more time in math (about 8 minutes), more time in collaborative activities (6 minutes), less didactic approaches (14 minutes), more scaffolded approaches (24 minutes), and more integration (15 minutes).

Figure 11. EduSnap average percentages grouped by CLASS Classroom Organization levels

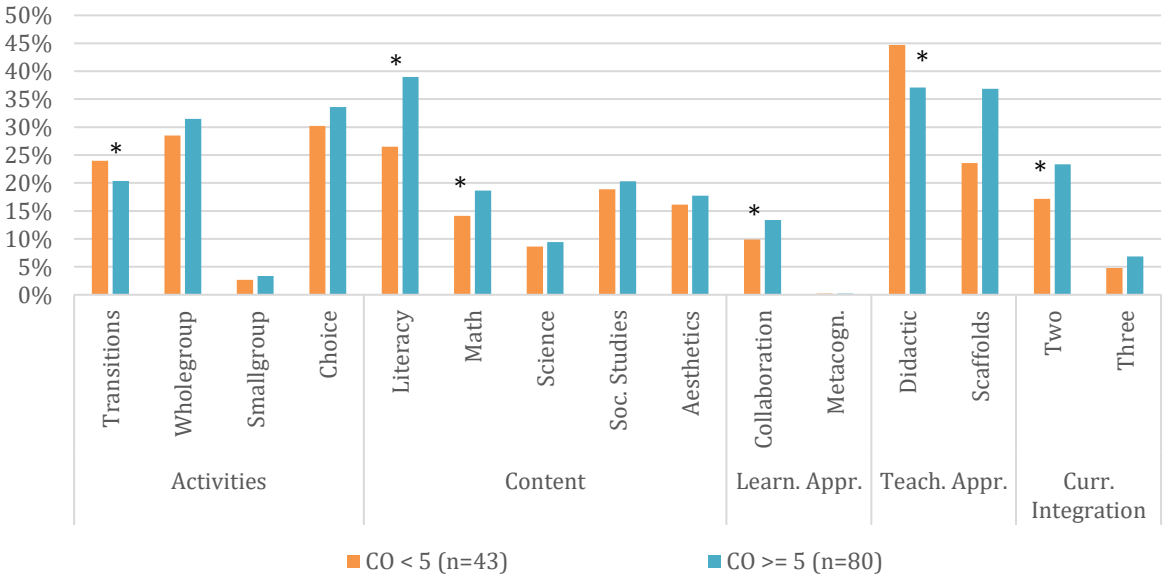
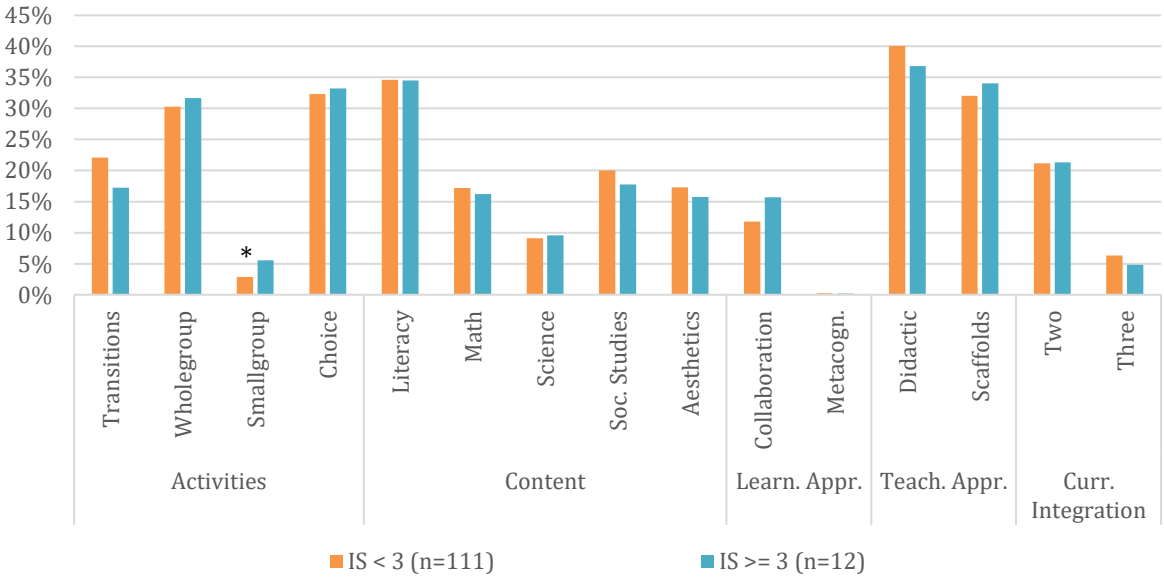


Figure 12 compares average EduSnap percentages for classrooms with instructional support below and above the threshold of 3. Higher IS quality classrooms spend on average slightly less time on transitions (about 8 minutes), use small group more (only about 5 more minutes on average), use more collaborative approaches (7 more minutes on average), and slightly more scaffolding. While not illustrated here, we also found differences in content and teaching approaches for classrooms that scored above the median of 1.9 (which decreased from 2.3 in 2017), for the use of free choice and less use of a whole group setting.

Figure 12. EduSnap average percentages grouped by CLASS Instructional support levels



The findings above are also estimating correlations between the different groupings and contents and activities captured in the EduSnap and the CLASS dimensions. These show

statistically significant (albeit moderate) negative correlations between transitions and the three CLASS dimensions (negative), didactic approaches and CLASS ES and CO, and no content and all three dimensions and positive between choice grouping and IS, literacy content and CLASS ES and CO, math content and CLASS CO, Collaborative and scaffolded approaches and CLASS ES and CO, and one content area and CLASS IS and two content areas and CLASS ES and CO.

We also examined the associations between CLASS and EduSnap and various classroom/center characteristics as well as teacher qualifications through simple regression analyses. We found positive associations between higher star ratings and less use of transitions, with programs with a star rating of 4 evidencing more collaborative approaches and more scaffolded instruction. Positive associations were also observed for Creative curriculum combined and time on social studies, and Creative alone and collaborative approaches (as well as less didactic approaches). We found a small association between CLASS dimensions and the Mother Goose Time curriculum, half of which are home providers.

3. Children's absenteeism and teacher turnover in the PHLpreK program, 2017-2018

We report children absenteeism in Appendix A, Table A.1, overall and for selected subgroups. Absenteeism between the months of March and June is included and was on average 11%. However, absenteeism was higher for Hispanic children (17%).

In addition, table A.2. reports teacher turnover as measured between October 2017 and February 2018. This turnover was reported upon observation of the research team in the center, and therefore likely underestimates turnover, as it does not account for 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 27% with turnover being higher among center types of centers, for teachers for which there is no information on qualifications, for some agencies, and within the curriculum types grouped under "other." Turnover varied between 8% and 36% depending on the agency.

4. Children's gains in the PHLpreK program, 2017-2018

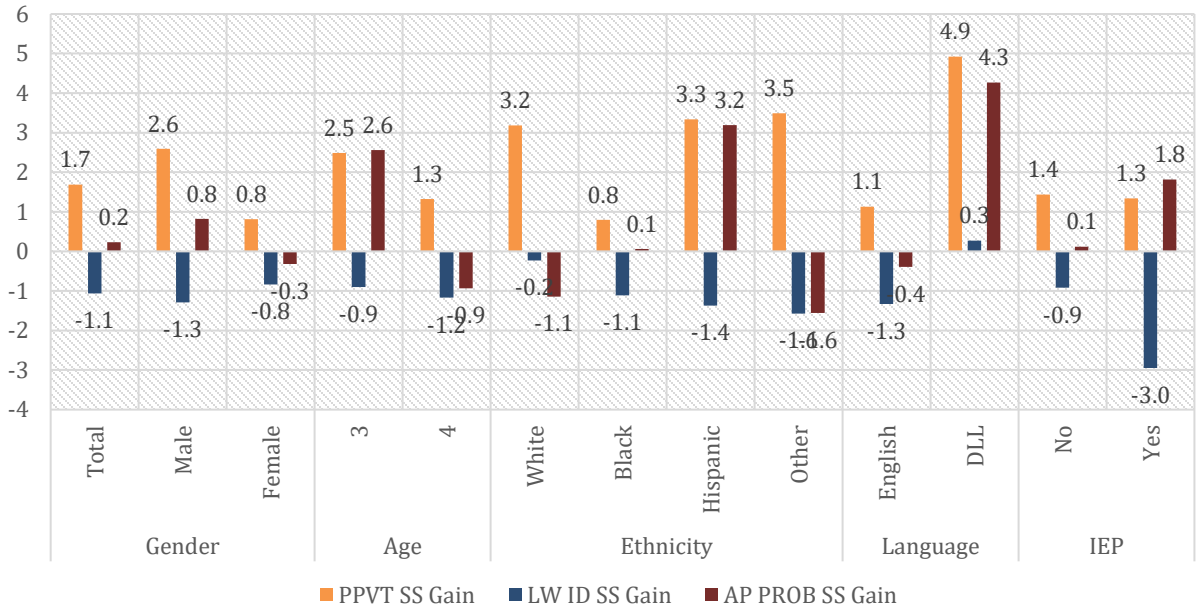
As described earlier, this evaluation assessed child outcomes in receptive vocabulary (using the Peabody Picture Vocabulary Test), literacy (using the Woodcock-Johnson Tests of Achievement Letter-Word subtest), and math (using the Woodcock-Johnson Tests of Achievement Applied Problems subtest). Moreover, it evaluated executive functioning (EF) using two measures: the Dimensional Change Card Sort Game (DCCS) and the Peg Tapping task (PT), as well as socio-emotional development (C-TRF).

Child gains for the 2017–18 school year for children in the PHLpreK sample and for selected subgroups of interest are reported depicted below and reported in detail in Appendix A. Only valid scores for children assessed in both fall and spring of the school year are included. Figure 13 provides standardized scores for the PPVT (vocabulary) and Woodcock-Johnson (literacy and math) assessments which allow to compare to expected gains after controlling for age (that is, to growth on maturation alone). Standard scores for these measures are standardized at the 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 from a similar background adjusting for age. Overall, children's standard scores increased on the PPVT and for most children in applied

problems. Other trends observed are: (a) larger fall to spring gains for children identified as White, Hispanic, Other, and DLL (dual language learner) in PPVT, (b) negative standard score gains for all children in letter-word identification, which were larger for IEP children, and (c) larger standard score gains in applied problems for three-year-old children, children identified as Hispanic, and DLL children.

As comparison, one-year gains for children in the sample are of 1.7 standard points on the PPVT, about a third of the 4.5 point one-year gains reported for 4-year-olds in the FACES study. Similarly, one-year gains in LW identification were of -1.1 standard points, quite a contrast with one-year gains for 4-year-olds in the FACES study of 5.0 standard points. Lastly, gains in applied problems were on average 0.2 standard point, about a tenth of one-year gains for 4-year-olds in FACES of 2.2 standard points. 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.5a; Aikens, Klein, Tarullo, & West, 2013).

Figure 13. Standard score gains for children in the PPVT, the WJ LW identification and the WJ applied problems

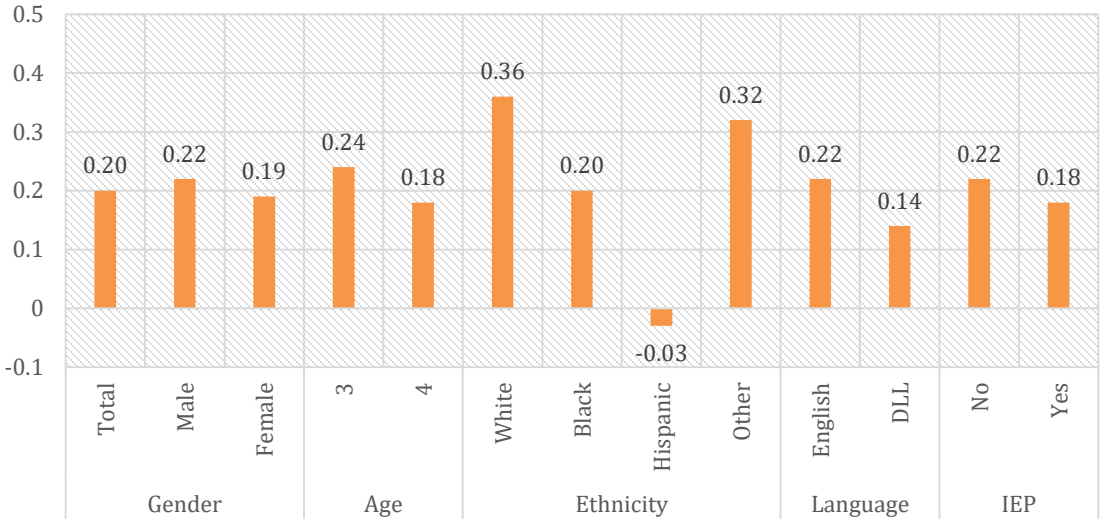


Note: n= 465 for the PPVT, n= 464 for the letter word identification and n= 462 for the applied problems.

Figures 14 and 15 below depict gains in DCCS and Peg tapping as well as changes in the C-TRF (which is reverse scored and therefore positive increases are not to be considered gains). Children gains in executive functions (DCCS and PT) across the board. In particular, there are stronger gains observed in children identified as White and smaller gains in children identified as Hispanic. 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, which is equivalent to the growth observed in this this study. This study also reports for the 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 advanced above this level throughout the preschool year.

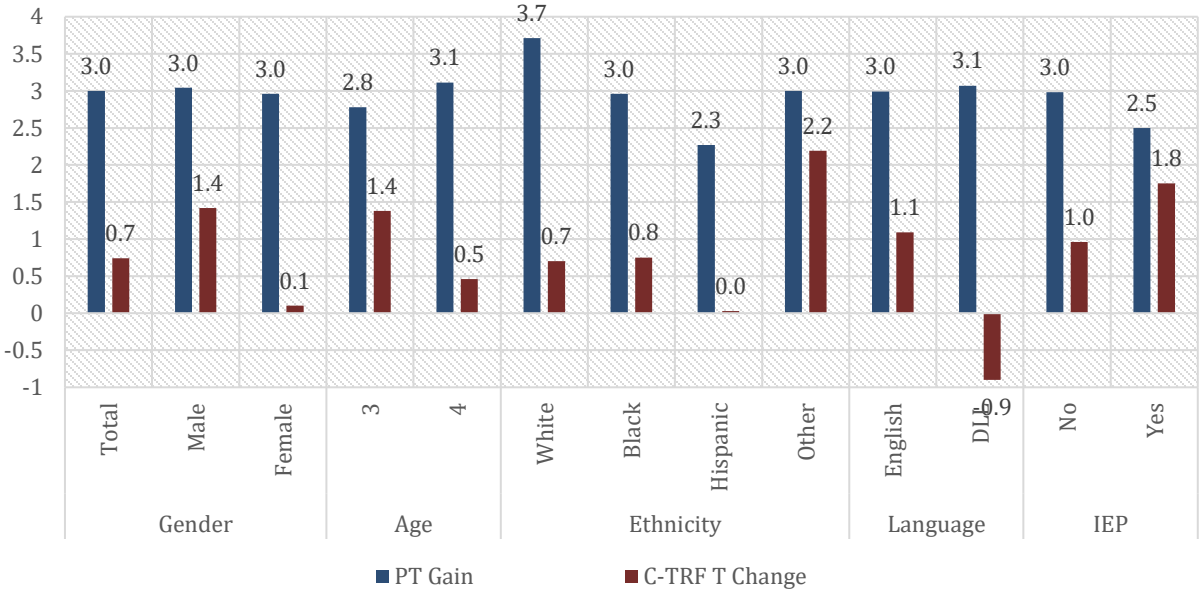
Figure 14. DCCS gains in children.



Note: n= 465 for the DCCS

As it pertains children's socio-emotional development, changes were for the most part negative, with a few exceptions: these were on average zero for females, and children identified as Hispanics and positive for DLL children.

Figure 15. Peg Tapping gains and C-TRF changes in children.



Note: n= 465 for the PT, n= 452 for the C-TRF.

The section that follows assesses the development of children and relates these to various aspects. In particular, we examine the association between children's learning gains and program features, as well as teacher qualifications using multi-level estimations. We control for children's characteristics, which also allows us to capture whether the program is allowing different groups of children to progress equally and/or reducing gaps. We include information on children's gender, race and ethnicity, home language, and IEP (we do not have information on children's income levels). Program features for PHLpreK include star rating, whether they are a Success by 6 program, curriculum, teacher qualifications and classroom quality. The analyses also take into account 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 conduct analyses including the two measures of quality, the CLASS and the EduSnap. Results are shown in the appendix B and summarized here. Table B.1. shows these for levels of CLASS, and Table B.2 at the cutoffs of quality of CLASS. Tables B.3 and B.4 report results for raw scores for those measures that where standardized in the first two tables. In essence, multivariate analyses account for how children are grouped, where they come from, and their background. With this in mind, these assess how children's gains differ among children, and what aspects of centers and teaching and learning, contribute to those gains.

Results show that female's and male's gains did not differ for most outcomes, with the exception of the socio-emotional dimension (the only one measured through teacher report), for which girls show stronger results by the spring. Children identified as African American and Hispanic evidence lower literacy and executive function gains through the school year relative to their White peers. Children identified as African American also evidence lower gains in receptive vocabulary. Smaller gains are also observed for dual language children in executive functions relative to their English only peers however, they evidence stronger gains in socio-emotional development. IEP children evidence lower gains in receptive vocabulary and on one of the executive functions measures, relative to non-IEP children. Absenteeism rate (measured for the March to June period) shows a strong negative association with literacy and a slight negative association with executive functions and socio-emotional development.

In terms of center-based factors, children in 3- and 4- star rated programs have lower gains in receptive vocabulary relative to 2-star rated programs. This is also the case in literacy for all 2-star and higher centers, after accounting for other child and center characteristics. This provides some indication that star ratings may not accurately be capturing aspects of quality that relate to children performance. Some differences by hub were detected that likely capture neighborhood effects, as income is not available for children and therefore not included as a control (not shown in the appendix). Teacher turnover (measured as a change in teacher at some point between pre-test and post-test) surprisingly shows a positive association with receptive vocabulary (was the change positive because turnover resulted in the recruitment of better teachers?) but a negative association with children's gain in one measure of executive functions.

Curriculum appears to have no systematic effect across the different child measures included (with a few positive and negative effects observed). Children exposed to teachers with different degrees (whether an Associate, BA or MA) evidence similar learning progress by the spring, although some benefits where observed for children with a teacher with a BA on literacy. CLASS CO evidences a positive association with receptive vocabulary, and a slight positive association with literacy, math and socio-emotional development. On the other hand, CLASS IS showed a negative association with socio-emotional development (this measure is inverted

therefore increases are to be interpreted negatively). Results for CLASS ES are surprising, as it shows negative associations with vocabulary, math and socio-emotional development.

As for the quantification of the experiences (the EduSnap), a higher use of transitions shows a slight and large negative association with literacy spring scores. The use of whole group also evidences a strong negative association with literacy scores. Negative associations also for group work and choice are indicative of a potential need to work on the quality of how these groupings are used, which is reflected in the low CLASS IS scores generally. Small group does however show a positive association with math.

The time teacher spends reading shows large positive associations with literacy and math scores, and children's time reading also shows slightly significant associations with math scores. Interestingly, reading comprehension also evidences a significant negative association with math and literacy, which is surprising although also aligned with the low CLASS IS scores and suggests further need for professional development on this area. There is a very large and statistically significant association for the time spent on children's vocabulary and receptive vocabulary, as is expected, on the time spent on writing and children's literacy, and on the time spent on geometry and math scores.

In sum, the main patterns that emerge from the multivariate analyses are: (a) there are some differences by ethnicity and race, as well as by language background and IEP that may require stronger efforts on the program to address equity; (b) absenteeism has important effects on children's performance; (c) star levels does not appear to be a strong indicator of quality as it relates to children's gains; (d) time spent in content areas strongly relates to children's outcomes, (e) higher levels of CLASS CO is related to stronger child performance, and (f) not much is observed for CLASS IS likely due to the low levels of CLASS IS quality (low variation).

5. Program supports: teacher and director perceptions

The evaluation also included surveys to directors and teachers to understand to what extent they were being served by any existing program supports. The surveys included understanding the personnel's experience in the field of early childhood, and in education more generally, as well as understanding their perceptions on professional development and technical assistance received since the launch of PHLpreK. Responses were optional, therefore n varies question to question. A total of 107 of 139 teachers and 71 of 86 directors responded to the survey. If a director was also a lead teacher it was indicated that they should respond a director survey.

Teachers

Teachers in the program reported on average to be 37 years of age (between 23 and 68) with about 11 years of experience in early childhood, four years in the current program, and three years in the current classroom (and some teachers having just started). There appears to be a match between the student and the teacher ethnic and racial composition. Of responding teachers, 18% were White, 61% African American, and 16% Hispanic, and 12% reported speaking Spanish. There is a lot of variation on experience across all star levels. For teachers that answered the question on their annual salary, 78% reported a salary between 11 and 40 thousand a year.

Table 11. Annual Salary Teachers

	Freq.	Percent
$10,000 or less	-	-
$11,000-$20,000	13	13.13
$21,000-$30,000	33	33.33
$31,000-$40,000	31	31.31
$41,000 or more	6	6.06
Do not know	3	3.03
Do not wish to share	11	11.11
Total	**99**	**100.00**

Teachers were asked about their participation in professional development activities. In the survey, professional development (PD) was defined as training and assistance for individual growth. Most teachers (87%) reported three or more in-service training days. The most common modalities of PD reported were study groups of some kind, direct instruction from an outside consultant and peer observation and feedback, and outside training (Table 12). Teachers were also asked about the specific number of PD workshops and 70% reported 1-6 of these.

Table 12. Responses to: In which of the following staff development and training activities have you participated during the current academic year? (n=103)

Professional Development	Percent
Three or more in-service training days (training delivered at my program, by program leadership)	87.25
Workshops involving study groups or small-group problem solving	71.84
Direct instruction from an outside consultant on a specific topic	77.67
Peer observation and feedback	71.29
Follow-up support for a teacher trying out new skills and knowledge in the classroom	69.00
Visits to, or observations of, other schools	46.00
Release time for attending early childhood professional conferences	48.98
Enrollment in college or university courses	58.59
Workshops on computers and technology in the classroom	45.63
Training outside of my program, with participants from other programs	74.76
PD program that uses coaching/consultation	61.17
Other	7.47

Table 13. Responses to: How many PD workshops do you recall attending since January [2017]? (n=103)

	Freq.	Percent
1-3	42	40.38
4-6	31	29.81
7-9	13	12.50
10 or more	18	17.31
Total	**104**	**100.00**

The survey also inquired into the content of the PD (Table 14). The most common contents reported were child development, general curriculum, child assessment and health and safety. The least common areas were supports of DLL, classroom quality and nutrition.

Table 14. Responses to: Mark all the broad topic areas that were covered in professional development workshops that you attended in the last year (n=107)

Professional Development Workshops	Percent
Child Development	76.64
Supporting English Language Learners (ELLs)	34.58
General Curriculum	69.16
Family Engagement/Partnership	63.55
Classroom Quality	56.60
Classroom Space and Learning Materials	59.81
Child Assessment	75.70
Nutrition	52.34
Nutrition on the Philadelphia Nutrition Standards	31.78
Health and Safety	81.31
Early Childhood Mental Health/Social Emotional Development	57.94

In addition to asking about content, teachers were asked about how much the PD their perceptions of the PD they attended. There was a general agreement about the usefulness of the workshops, and although some staff found some content basic, there is agreement about alignment to everyday experiences with children (Table 15).

Table 15. For each of the following statements in relation to the workshops you attended, record if you Strongly Disagree to Strongly Agree (5-point Likert) (n=105)

Workshop Perceptions	Percent				
	Strongly Disagree	→			Strongly Agree
The majority of these workshops were highly useful and helped make you a more effective teacher.	-	1.90	14.29	34.29	49.52
The majority of these workshops were basic and you already knew most of the areas covered.	6.67	19.05	27.62	28.57	18.10
The majority of these workshops did not cover things that were central to your everyday experiences with children.	43.81	26.67	8.57	14.29	6.67
The majority of these workshops were aligned with your needs in relation to your everyday experiences with children.	1.90	0.95	15.24	44.76	37.14

Teachers were also asked about their participation in a professional development program that uses coaching or consultation and they were provided examples such as My Teaching Partner and Practice-Based Coaching. A total of 37.50 percent of teachers (with only 88 teachers responding to this question) responded they did. Some specific programs mentioned were Success by Six and DVAEC. Teachers were also asked what type of PD they would like to attend or have offered to them. A total of 66 teachers expressed interest in additional supports such as behavior management strategies and support with working with children with disabilities. In addition, teachers asked for support integrating STEM lessons, and additional support learning the creative curriculum.

Technical Assistance (TA) was defined in the teacher survey as training and assistance for programmatic growth. Teachers reported requesting on average 8 hours of TA and teachers received an average of 12 hours. Having said that, 42 teachers received the amount of TA they

requested, and 15 reported receiving less. The majority of teacher perceived all types of TA to be beneficial or highly beneficial (Table 16).

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. (5-point Likert) (n=95)

Technical Assistance	Percent					N/A
	Not Beneficial		→		Highly Beneficial	
Child Development	1.06	-	10.64	23.40	50.00	14.89
Supporting English Language Learners (ELLs)	1.08	7.53	8.60	17.20	24.73	40.86
General Curriculum	-	2.15	11.83	24.73	43.01	18.28
Family Engagement/Partnership	1.05	2.11	7.37	29.47	40.00	20.00
Classroom Quality	-	2.15	6.45	29.03	45.16	17.20
Classroom Space and Learning Materials	-	-	7.53	27.96	52.69	11.83
Child Assessment	-	1.05	10.53	26.32	46.32	15.79
Nutrition	-	4.40	6.59	27.47	35.16	26.37
Nutrition on the Philadelphia Nutrition Standards	-	3.30	10.99	23.08	31.87	30.77
Health and Safety	-	1.06	4.26	27.66	57.45	9.57
Early Childhood Mental Health/Social Emotional Development	-	1.09	7.61	23.91	42.39	25.00

Teachers were also asked about having received classroom materials from different sources. Teachers reported (n=94) having received them mostly from PHLpreK (73.40%), Merit grants (17.02%), and other reported having received materials from Success by six (5.32%) or other sources (4.26%). Of the $1800 in Classroom Materials provided by PHLpreK based on initial Classroom Environment Checklist from NIEER in the Spring of 2017, 64% reported these were highly beneficial, 31% reported these were somewhat beneficial or adequate and only 4% reported these were not beneficial (only 45 teachers answered this question). Similarly, in terms of the $2100 Creative Curriculum Kit, out of 73 teachers that answered this question, 68 reported it to be highly beneficial, 30% to be adequate or somewhat beneficial and only 1% reported it not to be beneficial.

Directors

Directors were provided with a similar survey that also inquired into their demographics, as well as PD and TA opportunities. Director's are on average 43 years old, 21% White, 60% African American, 12% Hispanic and only one of the 70 directors that responded the survey spoke Spanish. A large majority report a master's degree (43%) or a BA (39%) with only a small fraction reporting an AA (14%) or some college (4%). A total of 13% reported having a CDA and 21% 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 30 and 60 thousand a year (Table 17).

Table 17. Annual Salary Directors

	Freq.	Percent
$30,000 or less	11	16.41
$31,000-$40,000	15	22.39
$41,000-$50,000	15	22.39
$51,000 or more	14	20.89
Do not wish to share	12	17.91
Total	**67**	**100.00**

Like teachers, directors were also asked about professional development activities, their content, quantity and perceptions about their usefulness. Their responses in terms of modes of delivery (Table 18) are quite consistent with those of teachers, with directors reporting a variety of modalities for the most part. Directors do report a higher quantity of workshops than teachers (Table 19).

Table 18. Responses to: In which of the following staff development and training activities have you participated during the current academic year? (n=68)

Professional Development	Percent
Three or more in-service training days (training delivered at my program, by program leadership)	82.35
Workshops involving study groups or small-group problem solving	73.53
Direct instruction from an outside consultant on a specific topic	94.29
Peer observation and feedback	69.57
Follow-up support for a teacher trying out new skills and knowledge in the classroom	77.27
Visits to, or observations of, other schools	57.97
Release time for attending early childhood professional conferences	73.13
Enrollment in college or university courses	53.03
Workshops on computers and technology in the classroom	46.97
Training outside of my program, with participants from other programs	85.51
PD program that uses coaching/consultation	77.94
Other	38.89

Table 19. Responses to: How many PD workshops do you recall attending since January [2017]? (n=69)

	Freq.	Percent
1-3	15	21.74
4-6	22	31.88
7-9	11	15.94
10 or more	21	30.43
Total	**69**	**100.00**

Content areas covered in the PD workshops reported was quite varied (Table 20). Most common were child development, general curriculum, child assessment, kindergarten transition, and health and safety, which was in line with what was reported by teachers.

Table 20. Responses to: Mark all the broad topic areas that were covered in professional development workshops that you attended in the last year (n=70)

Professional Development Workshops	Percent
Child Development	70.00
Supporting English Language Learners (ELLs)	27.14
General Curriculum	77.14
Family Engagement/Partnership	58.57
Classroom Quality	48.57
Classroom Space and Learning Materials	27.14
Child Assessment	70.00
Nutrition	64.29
Nutrition on the Philadelphia Nutrition Standards	40.00
Kindergarten Transition	70.00
Business Practice	51.43
Supervision	54.29
Health and Safety	75.71
Early Childhood Mental Health/Social Emotional Development	52.86

Directors were also asked about their perceptions on the usefulness of the content covered by PD. There responses were quite aligned with those of teachers. The majority found them quite useful, and while some directors reported content may have been basic sometimes, there is a general agreement on the alignment to their needs as center directors. One fourth of directors though did reported neutral responses to these statements for most questions.

Table 21. For each of the following statements in relation to the workshops you attended, record if you Strongly Disagree to Strongly agree (5-Point Likert) (n=70)

Workshop Perceptions	Percent				
	Strongly Disagree	→			Strongly Agree
The majority of these workshops were highly useful and helped make you a more effective director.	1.43	2.86	21.43	30.00	44.29
The majority of these workshops were basic and you already knew most of the areas covered.	8.57	15.71	40.00	25.71	10.00
The majority of these workshops did not cover things that were central to your everyday experiences as a center director.	31.43	27.14	25.71	7.14	8.57
The majority of these workshops were aligned with your needs in relation to your everyday experiences as a center director.	-	4.35	23.19	37.68	34.78

Directors were asked about their participation in PD that used coaching, to which 54% (n=61) reported they had done so. They were also asked about their participation in Creative Curriculum© training (75% reported participating, Work Sampling (54%) and Ages and Stages Questionnaire (ASQ) (78%) (n=61). Among the later, 88% reported having receive a copy of the ASQ. When asked what type(s) of professional development opportunities have been offered to them and opportunities they would like to attend, 41 directors expressed interest in understanding the aspects of overseeing and supervising the operations of a business (management, bookkeeping, taxes, contracts, grants, hiring/recruitment, etc.). Directors also expressed interest in learning the Creative Curriculum, GOLD—Teaching Strategies, the Ages

and Stages Questionnaire, and work sampling. Other interests include, coaching and staff development.

Directors were also asked about TA opportunities, to which there was a low response rate. A total of 40/50 reported having received more or the same TA hours they had requested. Directors were asked about their perceptions of the benefits of the TA received on different content (Table 22). Many chose not to respond to these, and among respondents there was general agreement on the benefits, with the exception of TA on ELLs and nutrition.

Table 22. 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. (5-point Likert) (n=55)

Technical Assistance	Percent					N/A
	Not Beneficial →				Highly Beneficial	
Child Development	-	-	12.73	30.91	27.27	29.09
Supporting English Language Learners (ELLs)	2.04	2.04	6.12	14.29	14.29	61.22
General Curriculum	-	1.85	14.81	29.63	29.63	24.07
Family Engagement/Partnership	1.85	3.70	16.67	25.93	27.78	24.07
Classroom Quality	1.79	-	12.50	19.64	39.29	26.79
Classroom Space and Learning Materials	1.92	-	9.62	17.31	44.23	26.92
Child Assessment	1.82	-	16.36	25.45	30.91	25.45
Nutrition	3.70	7.41	16.67	12.96	31.48	27.78
Nutrition on the Philadelphia Nutrition Standards	4.35	4.35	6.52	17.39	32.61	34.78
Kindergarten Transition	1.82	5.45	14.55	20.00	34.55	23.64
Business Practice	-	3.70	12.96	31.48	25.93	25.93
Supervision	3.85	-	7.69	23.08	34.62	30.77
Health and Safety	3.64	3.64	10.91	21.82	34.55	25.45
Early Childhood Mental Health/Social Emotional Development	-	3.70	9.26	18.52	35.19	33.33

Only 39% of directors reported having attended the kindergarten workshop onsite at ChildCare programs for families (n=69). The training for early learning providers on kindergarten readiness, transition, and registration provide by the Philadelphia school district was attended by 64% (n=67).

The PD on Business Practices and Administration was attended by 64% (n=70). Business and Financial Strength Technical Assistance was utilized by only 49% of directors (n=68) with TA having been from UAC 67%, PHMC 26% both 7% (n=27 to this sub-question).

Directors were also asked if the program was participating in cohort style professional development, money for program improvement and intensive technical assistance to move the program to the next level in the QRIS, to which only 26% reported participating. (n=62)

Discussion of Findings

This is the second report on the Philadelphia's PreK program evaluation. The program has now finished its second year of operation, sustaining and solidifying partnerships with various programs across the city. The evaluation's main purpose is to support understanding the program's growth, it's strengths and weaknesses and support the design professional development (which has until now been truncated by the trial pending on the tax which funds the program), as well as feed into continuous improvement strategies to help the program mature in quality. Classroom observations in pre-K will continue going forward into the Spring of 2019. Pre-K classrooms in these programs are averaging high to moderate levels of quality as measured by the CLASS Emotional Support and Classroom Organization, and very low levels of quality on the CLASS Instructional support domain, with a statistically significant fall in the Emotional and Instructional support domains observed between year one and two. That is, classrooms are on average adequately nurturing and safe environments for children and adequately structured and organized, although the decrease in scores signal as a warning on the necessary structures to support these over time. Teachers' use of strategies and techniques for scaffolding and expanding children's learning and language are however infrequent, if present, with feedback loops, questions on thought processes, clarifications, back and forth conversations, elaborations on concepts, problem solving, and planning and production nearly absent in classrooms. The decrease in the already low scores in CLASS instructional supports should serve as an alarm.

The EduSnap observations showed that classrooms are generally effective at implementing a variety of activity settings. It also showed some improvement even if the general findings remain the same, with a third of the day spent in choice time, in contrast to about 20% of the day in transitions. There seems to be a balance in groupings but the lack of associations between choice and small group and higher CLASS scores (or the findings in children's gains discussed below) point to a lack of quality in the use of these groupings. Teachers carry out multiple content areas for about a fourth of the day, but they also carry out no content for a third of the day. Children experience a balance of didactic and scaffolded interactions with the latter being more present in higher quality classrooms, but children are never asked to explain and/or justify their thinking through metacognitive processes.

This report also looks into how children's gains differ among different type of children and what aspects of centers and teaching and learning, contribute to those gains. Results show minimal differences by gender. On the other hand, children identified as African American or Hispanic evidence lower gains in literacy and executive functions, and African American also evidence lower gains also in receptive vocabulary. Smaller gains are also observed for dual language children (executive functions) and for IEP children (receptive vocabulary and executive functions). Child absenteeism (larger for Hispanic children) shows associations with literacy, executive functions and socio-emotional development.

Star levels does not appear to be a strong indicator of quality as it relates to children's gains. No systematic differences were observed for children's gains in relation to either curriculum or teacher qualifications. CLASS CO levels is related to children's receptive vocabulary, literacy, math and socio-emotional development. The absence of systematic CLASS IS associations may be likely due to the low CLASS IS quality levels across the program.

As it relates to the use of groupings and use of whole group, findings indicate a quality problem in the use of groupings even though there is balance. This is aligned with the low CLASS IS scores found, as well as with the results on higher CO classrooms faring better for

children. The time teachers read to children is positively related to literacy and math gains and children's time reading is so with math scores. A very large and statistically significant association is observed for the time spent on vocabulary and receptive vocabulary gains, for the time spent on writing and literacy gains, and for the time spent on geometry and math scores. This indicates that time spent on content is important for children's performance in related content areas.

Overall, the drop in quality scores is cause of concern, and presses further the need for strong supports for teachers in classroom quality (which in fact is one of the PD areas reported to have been received the lowest). Increasing classroom quality for the program will require proposing actions that strengthen instructional supports (concept development, quality of feedback, language modeling, metacognition) but also support ways in which to incorporate these into choice, small group and group work activity settings effectively. Supporting these process and increasing content (time and quality of it) and further integration of content are areas supported by the findings. Teachers' reports of interest in further PD on content (e.g. STEM) and curriculum seem to indicate a positive motivation on which to build on.

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, which were key in supporting our work with program providers throughout the City. Finally, we are thankful for all the programs and teachers that opened their doors to us. The authors of this report are also grateful to Hebbah El-Moslimany and Zaire Ali for their coordination work with the centers and Elisa Valle for helping with this report.

References

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

Aikens, N., Klein, A. K., Tarullo, L., & West, J. (2013). Getting ready for kindergarten: Children's progress during Head Start. FACES 2009 Report. (OPRE Report 2013–21a) Office of Planning. Research and Evaluation, Administration for Children and Families. Washington, D.C.: US 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. (2011). Effectiveness of early educational intervention. Science, 333, 975-978.

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

Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), Quality measurement in early childhood settings (pp. 11–31). Baltimore, MD: Brookes.

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.

Duncan, G, and 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.

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

Hatfield, B. E., Lower, J. K., Cassidy, D., & Faldowski, R. (2015). Inequities in access to quality early care and education: Associations with funding and community context. Early Childhood Research Quarterly Special Issue: Quality Rating and Improvement Systems (QRIS) as Change Agents, 30, 316-326. doi:10.1016/j.ecresq.2014.01.00hjm

Hestenes, L. L., Kintner-Duffy, V., Wang, Y. C., La Paro, K., Mims, S. U., Crosby, D., & Cassidy, D. J. (2015). Comparisons among quality measures in child care settings: Understanding the use of multiple measures in North Carolina's QRIS and their links to social-emotional development in preschool children. Early Childhood Research Quarterly, 30, 199-214.

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.

Martinez-Beck, I. (2011). Introduction: why strengthening the measurement of quality in early childhood settings has taken on new importance, in: M. Zaslow, I. Martinez-Beck, K. Tout, & T. Hall (Eds). *Quality Measurement in Early Childhood Settings*. Baltimore, MD: Brookes.

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

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

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.

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

Appendix A. Child pre- and post-scores and gains.

Table A.1. Child absenteeism AY 2017–18.

		Valid N	Mean
Total		341	11%
Gender	Male	168	11%
	Female	173	12%
Age	3-year-old	108	11%
	4-year-old	233	11%
Ethnicity	White	49	13%
	Black	230	10%
	Hispanic	39	17%
	Other	21	9%
Language	English	299	11%
	DLL	41	10%
IEP	No	280	10%
	Yes	22	13%

Note. Absence rate information was collected by PHMC from March to June of 2018.

Table A.2. Frequency and percentage of lead teacher replaced in AY 2017-18.

		Lead teacher replaced			
		No		Yes	
		n	%	n	%
Total N=137		100	73.0%	37	27.0%
STAR Ratings	1 to 2	17	60.70%	11	39.30%
	3	48	77.40%	14	22.60%
	4	35	74.50%	12	25.50%
Number of PHL PreK classrooms	1	38	69.1%	17	30.9%
	2	28	70.0%	12	30.0%
	3	12	66.7%	6	33.3%
	4 or More	22	91.7%	2	8.3%
Lead Teacher Credentials	AA Degree	39	81.3%	9	18.8%
	BA Degree	35	79.5%	9	20.5%
	MA or MS	13	81.3%	3	18.8%
	Not listed/No information	9	37.5%	15	62.5%
	CDA or ECE Credits	4	80.0%	1	20.0%
PHL Prek Partner Agency	UAC	43	72.9%	16	27.1%
	PHMC	38	70.4%	16	29.6%
	1199C	12	92.3%	1	7.7%
	SDP	7	63.6%	4	36.4%
Curriculum	Creative	58	73.4%	21	26.6%
	Creative + Other	14	73.7%	5	26.3%
	Mother Goose Time	15	75.0%	5	25.0%
	Other	13	68.4%	6	31.6%
Success By 6	No	90	73.8%	32	26.2%
	Yes	10	66.7%	5	33.3%

Note. Changes of lead teachers are based on the information collected as of October 2017 and February 2018 and likely do not account for full turnover.

Table A.3. PPVT Raw score means and Gains by child characteristics

		PPVT Raw F17			PPVT Raw S18		PPVT Raw Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Gender	Total	465	61.13	24.17	72.94	24.03	11.81	14.38
	Male	230	59.15	23.89	71.43	23.67	12.28	15.24
	Female	235	63.07	24.34	74.43	24.34	11.36	13.51
Age	3	152	46.79	21.60	59.95	21.77	13.16	14.84
	4	314	68.04	22.22	79.22	22.49	11.18	14.12
Ethnicity	White	66	67.71	28.66	81.14	27.16	13.42	14.01
	Black	302	61.26	22.69	72.99	22.45	11.73	14.57
	Hispanic	59	53.42	23.39	64.39	22.89	10.97	15.84
	Other	37	61.03	25.79	71.65	28.07	10.62	11.00
Language	English	394	63.60	23.59	75.42	23.00	11.82	14.34
	DLL	71	47.44	22.86	59.41	25.10	11.97	14.73
IEP	No	410	62.34	23.92	73.96	24.09	11.62	14.00
	Yes	38	50.61	20.96	61.03	19.87	10.42	11.16

Table A.4. PPVT standard score means and Gains by child characteristics

		PPVT SS F17			PPVT SS S18		PPVT SS Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Gender	Total	465	94.26	17.22	95.95	15.24	1.69	13.05
	Male	230	92.96	17.82	95.54	15.24	2.59	13.46
	Female	235	95.54	16.54	96.34	15.26	.81	12.60
Age	3	152	92.91	18.30	95.40	16.34	2.49	14.20
	4	314	94.87	16.64	96.19	14.67	1.32	12.44
Ethnicity	White	66	99.41	21.47	102.59	17.24	3.18	11.87
	Black	302	95.02	15.58	95.82	14.29	.80	12.74
	Hispanic	59	86.81	15.11	90.15	14.41	3.34	15.56
	Other	37	90.84	20.83	94.32	16.52	3.49	13.15
Language	English	394	96.37	16.54	97.50	14.52	1.13	12.84
	DLL	71	82.58	16.29	87.49	16.30	4.92	13.79
IEP	No	410	95.31	16.40	96.75	15.31	1.44	12.32
	Yes	38	86.03	16.41	87.37	13.18	1.34	9.99

Table A.5. WJ-LW Raw score means and Gains by child characteristics

		LWIDNT Raw F17			LWIDNT Raw S18		LWIDNT Raw Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Gender	Total	465	6.80	5.92	9.40	6.75	2.60	3.62
	Male	230	6.56	6.21	8.99	6.80	2.43	3.95
	Female	235	7.03	5.62	9.80	6.70	2.77	3.25
Age	3	152	4.78	4.44	7.04	5.80	2.26	3.93
	4	314	7.77	6.28	10.53	6.89	2.76	3.45
Ethnicity	White	66	7.94	6.45	10.98	7.93	3.05	4.01
	Black	302	6.65	5.98	9.24	6.55	2.59	3.63
	Hispanic	59	4.66	3.68	6.51	4.10	1.85	2.88
	Other	37	9.65	5.95	12.76	7.49	3.11	3.82
Language	English	394	7.01	6.17	9.55	6.96	2.53	3.58
	DLL	71	5.63	4.11	8.55	5.48	2.92	3.83
IEP	No	410	6.88	6.10	9.53	6.97	2.66	3.66
	Yes	38	5.95	4.25	7.68	4.57	1.74	3.58

Table A.6. WJ-LW standard score means and Gains by child characteristics

		LWIDNT SS F17			LWIDNT SS S18		LWIDNT SS Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Gender	Total	464	93.32	13.70	92.36	13.71	-1.06	9.70
	Male	229	92.58	13.95	91.51	14.14	-1.29	10.49
	Female	235	94.04	13.44	93.20	13.25	-.84	8.88
Age	3	152	96.95	14.14	96.05	14.61	-.90	12.42
	4	313	91.53	13.13	90.53	12.91	-1.17	8.08
Ethnicity	White	66	97.11	12.86	96.88	10.90	-.23	9.91
	Black	301	93.34	13.91	92.39	14.25	-1.11	9.56
	Hispanic	59	86.31	11.10	84.93	10.88	-1.37	10.40
	Other	37	98.22	12.32	96.65	12.45	-1.57	9.39
Language	English	393	93.90	13.90	92.69	13.89	-1.33	9.49
	DLL	71	90.10	12.10	90.37	12.75	.27	10.84
IEP	No	409	93.40	13.68	92.60	13.86	-.92	9.42
	Yes	38	91.53	15.29	88.58	13.60	-2.95	13.20

Table A.7. WJ-AP Raw score means and Gains by child characteristics

		APPROB Raw F17			APPROB Raw S18		APPROB Raw Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Gender	Total	465	7.69	4.20	9.78	4.16	2.09	2.84
	Male	230	7.33	4.26	9.55	4.22	2.23	2.99
	Female	235	8.06	4.12	10.01	4.11	1.95	2.70
Age	3	152	4.93	3.59	7.57	3.84	2.64	2.87
	4	314	9.04	3.80	10.85	3.88	1.82	2.80
Ethnicity	White	66	10.26	4.85	11.94	4.25	1.68	3.06
	Black	302	7.36	3.76	9.45	4.03	2.09	2.72
	Hispanic	59	6.36	4.16	8.76	3.92	2.41	2.98
	Other	37	8.27	4.60	10.38	4.17	2.11	3.21
Language	English	394	8.01	4.09	10.01	4.08	2.00	2.72
	DLL	71	5.97	4.43	8.63	4.37	2.66	3.30
IEP	No	410	7.86	4.18	9.95	4.08	2.10	2.75
	Yes	38	5.74	3.96	7.58	4.84	1.84	3.48

Table A.8. WJ-AP standard score means and Gains by child characteristics

		APPROB SS F17			APPROB SS S18		APPROB SS Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Gender	Total	462	87.13	15.44	87.50	14.54	.23	11.95
	Male	227	85.98	15.78	87.00	15.26	.82	12.52
	Female	235	88.24	15.06	87.99	13.82	-.32	11.38
Age	3	152	85.69	17.04	88.43	16.17	2.56	13.19
	4	311	87.84	14.55	87.03	13.67	-.93	11.13
Ethnicity	White	66	97.26	17.75	96.94	13.11	-1.14	12.75
	Black	302	86.27	13.97	86.58	14.26	.06	11.62
	Hispanic	59	79.88	14.71	83.00	12.48	3.19	12.87
	Other	34	88.94	14.68	85.73	15.67	-1.56	10.98
Language	English	393	88.62	14.59	88.39	14.15	-.39	11.43
	DLL	69	78.65	17.44	82.84	15.36	4.27	13.35
IEP	No	409	87.74	15.05	87.91	14.26	.12	11.80
	Yes	37	78.22	17.91	80.79	17.91	1.82	13.34

Table A.9. DCCS Final score means and Gains by child characteristics

		DCCS Final F17			DCCS Final S18		DCCS Final Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Gender	Total	465	1.25	.58	1.45	.61	.20	.67
	Male	230	1.20	.62	1.42	.60	.22	.71
	Female	235	1.30	.54	1.49	.63	.19	.63
Age	3	152	1.01	.51	1.26	.56	.24	.66
	4	314	1.36	.57	1.55	.62	.18	.67
Ethnicity	White	66	1.42	.61	1.79	.64	.36	.72
	Black	302	1.21	.54	1.41	.56	.20	.63
	Hispanic	59	1.27	.64	1.24	.60	-.03	.74
	Other	37	1.22	.67	1.54	.77	.32	.67
Language	English	394	1.26	.57	1.48	.61	.22	.65
	DLL	71	1.17	.61	1.31	.62	.14	.76
IEP	No	410	1.26	.57	1.48	.61	.22	.66
	Yes	38	1.05	.66	1.24	.63	.18	.77

Table A.10. PegTap score means and Gains by child characteristics

		PT Final F17			PT Final S18		PT Final Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Gender	Total	465	4.06	5.61	7.06	6.22	3.00	5.43
	Male	230	3.60	5.53	6.64	6.27	3.04	5.22
	Female	235	4.52	5.66	7.48	6.15	2.96	5.64
Age	3	152	1.17	4.03	3.95	5.28	2.78	4.74
	4	314	5.49	5.74	8.60	6.08	3.11	5.73
Ethnicity	White	66	5.76	6.11	9.47	5.89	3.71	5.52
	Black	302	3.63	5.31	6.59	6.15	2.96	5.30
	Hispanic	59	3.07	5.41	5.34	5.97	2.27	5.41
	Other	37	6.38	6.44	9.38	6.28	3.00	6.29
Language	English	394	4.20	5.54	7.19	6.19	2.99	5.27
	DLL	71	3.51	6.04	6.58	6.38	3.07	6.28
IEP	No	410	4.39	5.70	7.37	6.19	2.98	5.46
	Yes	38	.95	3.46	3.45	5.28	2.50	3.58

Table A.11. C-TRF Total Problems Raw score means and Gains by child characteristics

		C-TRF Raw F17			C-TRF Raw S18		C-TRF Raw Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Gender	Total	452	15.41	18.17	17.36	20.96	1.95	16.64
	Male	221	17.86	18.00	20.94	21.12	3.08	19.14
	Female	231	13.06	18.07	13.94	20.28	.87	13.78
Age	3	149	15.92	16.78	19.00	20.82	3.18	18.32
	4	304	15.11	18.83	16.46	20.98	1.35	15.72
Ethnicity	White	64	18.69	16.83	21.81	22.21	3.13	15.88
	Black	297	13.76	16.89	15.97	20.45	2.21	17.27
	Hispanic	59	21.17	25.18	19.78	21.92	-1.39	15.58
	Other	31	12.81	14.05	16.32	20.86	3.52	13.82
Language	English	382	15.06	17.90	17.63	21.54	2.57	16.82
	DLL	70	17.21	19.70	15.93	17.49	-1.29	15.29
IEP	No	399	14.14	16.96	16.39	19.79	2.26	16.03
	Yes	36	27.28	24.28	32.42	29.97	5.14	18.41

Table A.12. C-TRF Total Problems T score means and Gains by child characteristics

		C-TRF T F17			C-TRF T S18		C-TRF T Gain	
		Valid N	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Gender	Total	452	46.28	10.35	47.04	11.02	.74	9.34
	Male	221	47.05	10.40	48.47	10.89	1.42	9.62
	Female	231	45.54	10.27	45.66	10.99	.10	9.04
Age	3	149	46.87	9.96	48.21	10.75	1.38	10.28
	4	304	45.93	10.57	46.41	11.10	.46	8.85
Ethnicity	White	64	48.56	10.02	49.27	11.07	.70	7.95
	Black	297	45.42	9.81	46.20	10.95	.75	10.03
	Hispanic	59	48.69	12.76	48.73	10.96	.03	8.13
	Other	31	44.42	10.12	46.61	11.30	2.19	7.60
Language	English	382	46.02	10.22	47.13	11.12	1.09	9.46
	DLL	70	47.47	11.21	46.57	10.51	-.90	8.54
IEP	No	399	45.63	10.11	46.60	10.74	.96	9.40
	Yes	36	52.17	11.14	53.92	12.75	1.75	7.56

Appendix B. Child Estimations.

Table B.1. Multivariate analyses of children's 2017-18 Post (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	0.177	0.991	-1.314	0.066	-0.285	**-1.438~**
	(1.03)	(0.87)	(1.04)	(0.05)	(0.49)	(0.83)
Af.Am.	**-4.296***	**-3.305***	-2.453	**-0.333*****	**-2.106***	-1.788
	(1.78)	(1.49)	(1.86)	(0.09)	(0.85)	(1.51)
Hisp.	-2.454	**-5.310****	-0.666	**-0.376****	**-2.706***	-1.593
	(2.23)	(1.87)	(2.29)	(0.12)	(1.05)	(1.84)
Other Race/Ethn.	-2.883	-0.062	**-5.501***	-0.102	-1.148	0.037
	(2.55)	(2.13)	(2.65)	(0.13)	(1.20)	(2.06)
DLL	-2.553	1.295	-0.661	**-0.209***	-0.706	**-2.414~**
	(1.76)	(1.44)	(1.80)	(0.09)	(0.81)	(1.37)
IEP	**-3.444~**	-0.806	1.142	**-0.185~**	**-2.196***	1.566
	(1.94)	(1.61)	(2.10)	(0.10)	(0.91)	(1.59)
Absent. 4m-Rate	4.194	**-14.908****	-4.871	**-0.530~**	-2.827	**8.212~**
	(5.62)	(4.73)	(5.71)	(0.29)	(2.65)	(4.70)
Star 2	-5.588	**-9.065***	8.207	**0.475~**	-1.754	4.872
	(5.09)	(4.27)	(5.18)	(0.27)	(2.40)	(4.73)
Star 3	**-12.017***	**-8.780***	7.811	0.351	-0.898	3.920
	(4.94)	(4.13)	(5.01)	(0.26)	(2.33)	(4.60)
Star 4	**-11.735***	**-7.660~**	8.238	0.323	-0.506	5.900
	(5.19)	(4.35)	(5.27)	(0.27)	(2.45)	(4.83)
Creative+	2.169	0.151	2.859	-0.058	-0.639	1.765
	(1.90)	(1.59)	(1.93)	(0.10)	(0.89)	(1.80)
Mother Goose	**-6.266***	-3.291	**5.787***	0.076	1.204	0.074
	(2.66)	(2.23)	(2.68)	(0.14)	(1.25)	(2.51)
Other Curricul.	**-4.359***	0.252	-1.065	0.033	-0.274	**4.582***
	(2.02)	(1.68)	(2.04)	(0.11)	(0.95)	(1.90)
LT Turnover	**4.413****	-1.043	1.182	**-0.121~**	0.318	-1.363
	(1.39)	(1.16)	(1.41)	(0.07)	(0.65)	(1.30)
LT Associate	0.901	0.624	-1.469	0.019	0.551	**3.406~**
	(2.00)	(1.67)	(2.03)	(0.10)	(0.94)	(1.88)
LT Bachelor	-0.182	2.034	0.020	-0.079	0.009	1.136
	(1.93)	(1.61)	(1.95)	(0.10)	(0.91)	(1.82)
LT Master	1.542	0.628	-0.502	0.057	-1.745	1.449
	(2.81)	(2.35)	(2.86)	(0.15)	(1.32)	(2.63)
LT No response	1.300	0.513	-6.206	-0.128	**-3.115~**	2.013
	(3.80)	(3.18)	(3.84)	(0.20)	(1.79)	(3.62)
LT Af.Am.	1.210	-0.864	1.635	0.010	-0.060	2.776
	(2.05)	(1.71)	(2.08)	(0.11)	(0.96)	(1.91)
LT Hisp.	-2.990	**-3.578~**	**4.816***	**-0.255***	0.697	3.347
	(2.37)	(1.98)	(2.41)	(0.12)	(1.12)	(2.23)
LT Other	-0.725	**-7.171***	4.586	0.090	-0.846	**7.231***
	(3.71)	(3.13)	(3.74)	(0.19)	(1.74)	(3.52)
LT No response	-1.006	-1.035	**7.181~**	0.091	**3.500~**	0.925
	(4.09)	(3.42)	(4.13)	(0.21)	(1.93)	(3.87)
CLASS ES	**-5.383****	-1.718	**-4.398***	0.001	-0.501	**3.036~**
	(1.86)	(1.55)	(1.89)	(0.10)	(0.87)	(1.74)

CLASS CO	**4.866****	**2.397~**	**2.775~**	0.030	0.113	**-3.631***
	(1.61)	(1.34)	(1.63)	(0.08)	(0.75)	(1.51)
CLASS IS	0.120	1.817~	-0.936	-0.074	0.274	**2.666***
	(1.19)	(1.00)	(1.21)	(0.06)	(0.56)	(1.11)
Transitions	-11.438	**-15.837~**	-2.741	0.037	-2.845	2.590
	(10.82)	(9.02)	(10.88)	(0.57)	(5.08)	(10.15)
Whole Group	-9.537	**-21.871****	-1.852	**1.044***	-2.591	12.564
	(9.80)	(8.19)	(9.83)	(0.51)	(4.61)	(9.22)
Small Group	-9.899	3.207	**21.326~**	0.526	-1.993	-6.868
	(12.35)	(10.31)	(12.44)	(0.64)	(5.80)	(11.69)
Group Work	**-83.720~**	**-110.689****	-47.471	-2.452	-30.768	**96.914***
	(43.11)	(36.03)	(43.91)	(2.26)	(20.39)	(40.50)
Choice	-7.110	**-27.656*****	-4.536	**0.919***	-3.102	5.779
	(8.51)	(7.10)	(8.53)	(0.44)	(3.99)	(7.96)
Read To	0.786	**44.473****	**43.952***	-0.487	6.114	5.212
	(17.27)	(14.45)	(17.37)	(0.90)	(8.11)	(16.36)
Reading	9.293	7.314	**65.594~**	-1.004	-7.478	-18.642
	(38.68)	(32.46)	(39.10)	(2.02)	(18.22)	(36.27)
Reading Co	-5.051	**-59.308****	**-62.251***	1.186	-9.736	-9.774
	(24.86)	(20.84)	(25.13)	(1.29)	(11.69)	(23.29)
Word Ident	10.652	12.935	20.243	0.830	**11.307~**	-15.771
	(12.46)	(10.42)	(12.58)	(0.65)	(5.88)	(11.71)
Vocab.	**85.175*****	34.629	37.075	0.397	-11.848	5.290
	(25.65)	(21.45)	(26.39)	(1.34)	(12.06)	(24.49)
Writing	-0.690	**50.431~**	-2.124	0.521	-4.477	16.785
	(30.91)	(25.78)	(31.06)	(1.61)	(14.49)	(29.16)
Numbers	-6.286	14.001	2.313	-1.590	12.447	21.457
	(20.19)	(16.85)	(20.73)	(1.05)	(9.47)	(19.13)
Geometry	-6.900	-9.584	**30.946****	0.180	2.213	-11.586
	(10.93)	(9.13)	(11.10)	(0.57)	(5.14)	(10.35)
Operations	-19.239	22.730	-31.009	-0.681	9.414	-9.350
	(23.06)	(19.25)	(23.71)	(1.22)	(10.84)	(21.48)
Science	-1.509	1.998	-3.196	0.353	0.025	2.056
	(8.78)	(7.33)	(8.87)	(0.46)	(4.13)	(8.35)
Gross Motor	4.580	9.832	0.274	-0.210	2.620	-13.081
	(8.95)	(7.49)	(9.08)	(0.47)	(4.21)	(8.46)
Social Studies	-6.162	1.599	5.199	0.387	-3.249	-4.379
	(6.85)	(5.72)	(6.96)	(0.36)	(3.23)	(6.43)
Aesthetics	-2.634	-0.887	2.906	**-0.681***	5.550~	-6.503
	(6.57)	(5.48)	(6.59)	(0.34)	(3.09)	(6.15)
Collabor.	-6.073	2.143	-4.289	-0.045	**5.764~**	**-15.135***
	(6.38)	(5.33)	(6.50)	(0.33)	(3.01)	(6.01)
N	387	386	379	387	387	387

* 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 B.2. Multivariate analyses of children's 2017-18 Post (spring) standard score in relation to child and site or classroom characteristics including the CLASS (cutoff) and EduSnap.

	Receptive Vocabulary	Literacy	Math	DCCS Final	Peg Tapping	Socio- Emotional (inverted measure)
Female	0.470	1.181	-1.159	0.062	-0.253	**-1.383~**
	(1.04)	(0.87)	(1.05)	(0.05)	(0.49)	(0.83)
Af.Am.	**-3.996***	**-2.649~**	-1.991	**-0.327*****	**-1.906***	-1.365
	(1.79)	(1.50)	(1.86)	(0.09)	(0.84)	(1.54)
Hisp.	-2.950	**-5.173****	-0.731	**-0.360****	**-2.604***	-1.368
	(2.25)	(1.89)	(2.30)	(0.12)	(1.05)	(1.86)
Other Race/Ethn.	-2.328	0.444	**-5.147~**	-0.106	-1.056	0.263
	(2.57)	(2.14)	(2.66)	(0.13)	(1.20)	(2.06)
DLL	-2.374	1.242	-0.744	**-0.225***	-0.745	**-2.471~**
	(1.79)	(1.45)	(1.81)	(0.09)	(0.81)	(1.38)
IEP	**-3.742~**	-0.762	0.696	**-0.195~**	**-2.248***	1.635
	(1.96)	(1.62)	(2.11)	(0.10)	(0.91)	(1.60)
Absent. 4m-Rate	5.462	**-14.560****	-4.030	**-0.538~**	-2.886	7.613
	(5.68)	(4.75)	(5.72)	(0.29)	(2.64)	(4.76)
Star 2	-7.176	**-10.006***	7.446	**0.460~**	-1.771	6.359
	(5.09)	(4.25)	(5.15)	(0.26)	(2.38)	(4.89)
Star 3	**-12.497***	**-9.181***	7.382	0.319	-0.987	5.395
	(4.97)	(4.14)	(5.01)	(0.26)	(2.32)	(4.76)
Star 4	**-13.242***	**-8.268~**	7.348	0.296	-0.504	7.909
	(5.19)	(4.32)	(5.23)	(0.27)	(2.42)	(4.98)
Creative+	**3.301~**	1.421	2.953	-0.086	-0.576	2.207
	(1.88)	(1.56)	(1.89)	(0.10)	(0.88)	(1.83)
Mother Goose	**-5.424***	-2.012	**6.641***	0.096	1.592	-0.338
	(2.61)	(2.18)	(2.62)	(0.14)	(1.22)	(2.54)
Other Curricul.	-2.893	0.396	-0.227	0.023	-0.287	**3.324~**
	(1.98)	(1.65)	(2.00)	(0.10)	(0.92)	(1.93)
LT Turnover	**3.746****	-0.873	0.235	**-0.127~**	0.160	-0.660
	(1.37)	(1.14)	(1.39)	(0.07)	(0.64)	(1.33)
LT Associate	0.828	1.198	-1.743	0.032	0.488	**3.992***
	(2.03)	(1.69)	(2.05)	(0.10)	(0.94)	(1.98)
LT Bachelor	0.610	**2.733~**	0.291	-0.088	0.022	1.481
	(1.96)	(1.63)	(1.96)	(0.10)	(0.91)	(1.91)
LT Master	1.245	1.107	-0.922	0.060	-1.726	2.064
	(2.84)	(2.37)	(2.87)	(0.15)	(1.33)	(2.74)
LT No response	2.570	-0.429	-5.012	-0.113	**-3.156~**	-0.664
	(3.85)	(3.21)	(3.88)	(0.20)	(1.79)	(3.81)
LT Af.Am.	1.457	-1.182	1.980	0.030	-0.050	1.547
	(2.09)	(1.75)	(2.11)	(0.11)	(0.98)	(2.02)
LT Hisp.	-1.793	**-3.567~**	**5.760***	**-0.247~**	0.742	2.051
	(2.47)	(2.06)	(2.50)	(0.13)	(1.15)	(2.39)
LT Other	-1.101	**-6.583***	4.452	0.107	-0.575	**7.209~**
	(3.86)	(3.24)	(3.86)	(0.20)	(1.80)	(3.77)
LT No response	-2.665	-0.488	5.195	0.069	**3.389~**	3.196
	(4.06)	(3.38)	(4.06)	(0.21)	(1.89)	(3.97)
CLASS ES > 5.5	-0.632	-0.068	-2.505	-0.041	-0.388	-1.013
	(1.78)	(1.48)	(1.82)	(0.09)	(0.83)	(1.72)
CLASS CO > 5.5	2.192	**2.768~**	0.713	0.015	-0.148	1.835
	(1.70)	(1.42)	(1.72)	(0.09)	(0.79)	(1.67)
CLASS IS > 3.5	1.048	1.569	-4.245	-0.206	-0.766	1.100
	(3.09)	(2.60)	(3.08)	(0.16)	(1.44)	(3.00)

Transitions	-14.735	**-15.632~**	-2.681	0.070	-2.235	8.411
	(11.02)	(9.16)	(11.02)	(0.57)	(5.13)	(10.68)
Whole Group	-10.142	**-19.524***	-0.894	**1.140***	-1.798	13.734
	(9.90)	(8.25)	(9.89)	(0.51)	(4.61)	(9.65)
Small Group	-10.117	5.420	**26.534***	0.779	-0.553	-3.774
	(13.29)	(11.07)	(13.29)	(0.68)	(6.17)	(13.00)
Group Work	-66.672	**-120.902*****	-22.360	-2.002	-28.078	55.885
	(43.24)	(36.00)	(43.69)	(2.25)	(20.27)	(41.95)
Choice	-6.879	**-23.571*****	-5.190	**0.869***	-2.214	9.338
	(8.25)	(6.86)	(8.23)	(0.43)	(3.84)	(8.00)
Read To	-4.459	**38.738****	**44.470***	-0.369	5.626	7.854
	(17.29)	(14.43)	(17.33)	(0.89)	(8.04)	(17.00)
Reading	-18.441	20.792	45.561	-0.974	-5.543	24.018
	(37.51)	(31.41)	(37.67)	(1.94)	(17.53)	(36.30)
Reading Co	2.424	**-56.510****	**-58.459***	0.957	-7.626	-18.940
	(25.19)	(21.02)	(25.47)	(1.30)	(11.73)	(24.36)
Word Ident	6.210	11.405	16.083	0.839	**10.457~**	-13.828
	(12.55)	(10.46)	(12.62)	(0.65)	(5.86)	(12.19)
Vocab.	**92.722*****	**37.766~**	41.714	0.145	-11.688	11.594
	(25.75)	(21.46)	(26.33)	(1.33)	(12.00)	(25.50)
Writing	15.306	50.918*	15.808	0.628	-2.168	2.100
	(30.57)	(25.43)	(30.48)	(1.58)	(14.20)	(29.77)
Numbers	1.362	19.948	3.459	-1.438	11.438	14.675
	(20.13)	(16.73)	(20.57)	(1.04)	(9.34)	(19.71)
Geometry	-1.852	-5.029	**32.826****	0.255	2.369	-14.521
	(10.88)	(9.06)	(10.98)	(0.56)	(5.07)	(10.68)
Operations	-16.836	19.335	-25.148	-0.543	9.762	-16.102
	(23.29)	(19.39)	(23.77)	(1.22)	(10.85)	(22.34)
Science	-1.339	-0.431	-1.907	0.376	-0.270	0.434
	(8.76)	(7.29)	(8.79)	(0.45)	(4.08)	(8.64)
Gross Motor	3.571	6.261	3.537	-0.155	3.058	-13.097
	(8.96)	(7.46)	(9.04)	(0.46)	(4.18)	(8.77)
Social Studies	-8.903	-1.812	4.829	0.425	-3.348	-5.871
	(6.94)	(5.78)	(7.03)	(0.36)	(3.24)	(6.72)
Aesthetics	-5.155	-3.905	0.747	**-0.727***	5.074	-7.105
	(6.65)	(5.52)	(6.64)	(0.35)	(3.10)	(6.42)
Collabor.	-7.725	1.256	-4.907	0.014	**5.378~**	**-14.162***
	(6.44)	(5.36)	(6.51)	(0.33)	(3.01)	(6.29)
N	387	386	379	387	387	387

* 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 B.3. Multivariate analyses of children's 2017-18 Post (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	0.577	0.281	-0.285	-3.010~
	(1.27)	(0.36)	(0.26)	(1.60)
Af.Am.	-3.425	-0.829	-0.275	-2.882
	(2.19)	(0.62)	(0.46)	(2.78)
Hisp.	-4.067	**-1.939***	0.041	-4.477
	(2.74)	(0.78)	(0.57)	(3.45)
Other Race/Ethn.	-2.493	0.065	-0.706	0.768
	(3.13)	(0.89)	(0.64)	(3.94)
DLL	-2.619	**0.997~**	-0.316	-3.901
	(2.17)	(0.60)	(0.44)	(2.64)
IEP	-3.435	-0.296	-0.681	4.042
	(2.38)	(0.67)	(0.50)	(3.05)
Absent. 4m-Rate	1.143	**-3.455~**	-1.554	**16.892~**
	(6.89)	(1.96)	(1.42)	(8.79)
Star 2	-5.361	**-4.512***	1.347	3.615
	(6.25)	(1.78)	(1.29)	(8.08)
Star 3	**-13.110***	-4.281*	1.134	4.689
	(6.07)	(1.73)	(1.25)	(7.85)
Star 4	**-12.012~**	**-3.868***	1.229	7.883
	(6.37)	(1.81)	(1.31)	(8.25)
Creative+	3.010	0.390	**0.905~**	2.189
	(2.33)	(0.66)	(0.48)	(3.03)
Mother Goose	**-8.366***	-1.451	1.490*	-3.954
	(3.26)	(0.93)	(0.67)	(4.24)
Other Curricul.	-2.013	0.040	-0.175	4.488
	(2.47)	(0.70)	(0.51)	(3.21)
LT Turnover	**4.910****	0.007	-0.018	-0.634
	(1.70)	(0.48)	(0.35)	(2.21)
LT Associate	1.271	0.385	-0.632	3.793
	(2.45)	(0.70)	(0.50)	(3.18)
LT Bachelor	0.711	1.074	-0.126	-0.487
	(2.36)	(0.67)	(0.49)	(3.07)
LT Master	2.515	0.380	-0.600	-0.220
	(3.45)	(0.98)	(0.71)	(4.48)
LT No response	1.712	0.482	-1.391	1.580
	(4.66)	(1.33)	(0.96)	(6.08)
LT Af.Am.	1.811	-0.478	0.169	5.161
	(2.51)	(0.71)	(0.52)	(3.26)
LT Hisp.	-2.008	-1.255	0.895	4.034
	(2.91)	(0.83)	(0.60)	(3.79)
LT Other	0.991	**-2.856***	0.760	7.364
	(4.55)	(1.30)	(0.94)	(5.92)
LT No response	-0.854	-0.601	0.736	1.783
	(5.02)	(1.43)	(1.03)	(6.53)
CLASS ES	**-5.192***	-1.024	**-0.804~**	4.787
	(2.28)	(0.65)	(0.47)	(2.95)
CLASS CO	**4.603***	1.233*	0.486	**-5.617***
	(1.97)	(0.56)	(0.40)	(2.56)
CLASS IS	-0.336	0.504	-0.168	**5.761****

	(1.46)	(0.42)	(0.30)	(1.89)
Transitions	-15.176	-2.779	-1.894	9.355
	(13.29)	(3.77)	(2.73)	(17.19)
Whole Group	-11.574	**-5.979~**	-0.412	14.282
	(12.02)	(3.42)	(2.47)	(15.62)
Small Group	-22.195	1.061	4.602	-3.213
	(15.14)	(4.30)	(3.10)	(19.80)
Group Work	**-89.221~**	**-43.619****	-14.042	106.453
	(52.94)	(15.05)	(10.89)	(68.60)
Choice	-8.527	-7.965**	-0.997	4.576
	(10.44)	(2.96)	(2.14)	(13.53)
Read To	11.451	**11.900***	**9.911***	1.456
	(21.20)	(6.04)	(4.35)	(27.60)
Reading	29.097	2.681	14.059	-0.805
	(47.49)	(13.53)	(9.76)	(61.91)
Reading Co	-9.917	**-16.594~**	**-14.437***	-5.932
	(30.50)	(8.69)	(6.25)	(39.58)
Word Ident	4.163	4.119	4.398	-16.206
	(15.30)	(4.35)	(3.14)	(19.84)
Vocab.	**94.167****	13.252	6.481	14.874
	(31.48)	(8.94)	(6.46)	(40.99)
Writing	1.555	**27.298***	0.309	2.119
	(37.97)	(10.78)	(7.75)	(49.08)
Numbers	-6.353	8.163	4.739	14.511
	(24.77)	(7.03)	(5.08)	(32.14)
Geometry	-9.744	-4.553	4.626~	-24.527
	(13.42)	(3.81)	(2.76)	(17.51)
Operations	-24.566	7.015	-5.628	1.917
	(28.30)	(8.03)	(5.84)	(36.58)
Science	-2.195	0.061	-2.138	2.328
	(10.77)	(3.06)	(2.21)	(14.02)
Gross Motor	1.977	1.135	-0.294	-10.746
	(10.98)	(3.12)	(2.26)	(14.27)
Social Studies	-7.797	-0.725	0.880	0.415
	(8.40)	(2.38)	(1.73)	(10.89)
Aesthetics	0.757	-0.208	0.157	-7.427
	(8.07)	(2.29)	(1.66)	(10.43)
Collabor.	-9.445	0.642	-1.401	**-27.848****
	(7.83)	(2.23)	(1.61)	(10.18)
N	387	387	387	387

* 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 B.4. Multivariate analyses of children's 2017-18 Post (spring) raw scores in relation to child and site or classroom characteristics including the CLASS (cutoff) and EduSnap.

	Receptive Vocabulary	Literacy	Math	Socio- Emotional (inverted measure)
Female	0.988	0.350	-0.242	**-2.759~**
	(1.27)	(0.36)	(0.26)	(1.61)
Af.Am.	-3.225	-0.470	-0.179	-1.959
	(2.20)	(0.62)	(0.46)	(2.84)
Hisp.	**-4.774~**	**-1.805***	0.031	-4.001
	(2.76)	(0.78)	(0.57)	(3.52)
Other Race/Ethn.	-2.135	0.317	-0.645	1.250
	(3.14)	(0.89)	(0.64)	(3.98)
DLL	-2.374	0.897	-0.326	-3.762
	(2.19)	(0.60)	(0.44)	(2.68)
IEP	-3.656	-0.351	-0.763	4.432
	(2.40)	(0.67)	(0.50)	(3.10)
Absent. 4m-Rate	2.206	**-3.278~**	-1.468	**15.012~**
	(6.93)	(1.96)	(1.42)	(8.93)
Star 2	-6.599	**-4.976****	1.240	5.950
	(6.23)	(1.77)	(1.28)	(8.36)
Star 3	**-13.294***	**-4.519****	1.064	6.873
	(6.07)	(1.72)	(1.25)	(8.15)
Star 4	**-13.303***	**-4.203***	1.089	11.148
	(6.34)	(1.80)	(1.30)	(8.52)
Creative+	**3.790~**	0.858	**0.895~**	3.482
	(2.30)	(0.65)	(0.47)	(3.10)
Mother Goose	**-7.562***	-0.821	**1.702****	-3.804
	(3.20)	(0.91)	(0.65)	(4.31)
Other Curricul.	-0.456	0.091	-0.026	2.479
	(2.43)	(0.69)	(0.50)	(3.27)
LT Turnover	**3.960***	0.021	-0.235	0.389
	(1.68)	(0.47)	(0.34)	(2.26)
LT Associate	0.801	0.707	-0.723	4.347
	(2.48)	(0.70)	(0.51)	(3.35)
LT Bachelor	1.522	**1.373***	-0.082	0.090
	(2.39)	(0.68)	(0.49)	(3.24)
LT Master	2.289	0.579	-0.653	1.064
	(3.48)	(0.99)	(0.71)	(4.69)
LT No response	4.328	0.035	-1.059	-2.694
	(4.71)	(1.34)	(0.96)	(6.42)
LT Af.Am.	2.661	-0.608	0.303	3.371
	(2.56)	(0.73)	(0.52)	(3.45)
LT Hisp.	0.223	-1.258	1.166~	2.276
	(3.02)	(0.86)	(0.62)	(4.09)
LT Other	1.715	**-2.688***	0.898	8.811
	(4.71)	(1.34)	(0.97)	(6.38)
LT No response	-2.847	-0.519	0.356	5.993
	(4.96)	(1.41)	(1.02)	(6.71)
CLASS ES > 5.5	1.285	-0.653	-0.283	0.659
	(2.18)	(0.62)	(0.45)	(2.94)
CLASS CO > 5.5	0.723	**1.532****	-0.047	1.413
	(2.08)	(0.59)	(0.43)	(2.84)
CLASS IS > 3.5	-1.214	-0.335	-1.037	4.194
	(3.78)	(1.07)	(0.77)	(5.09)

Transitions	-20.463	-2.166	-1.984	16.255
	(13.48)	(3.81)	(2.75)	(18.14)
Whole Group	-13.753	-4.521	-0.248	15.836
	(12.11)	(3.43)	(2.48)	(16.38)
Small Group	-21.159	3.434	5.811~	-1.429
	(16.23)	(4.58)	(3.31)	(22.05)
Group Work	-65.513	**-46.569****	-8.682	40.976
	(52.89)	(14.99)	(10.83)	(71.30)
Choice	-9.475	**-6.417***	-1.060	12.667
	(10.09)	(2.86)	(2.06)	(13.61)
Read To	4.757	10.448~	**9.789***	0.151
	(21.16)	(6.02)	(4.31)	(28.71)
Reading	-11.590	9.248	9.691	67.490
	(45.88)	(13.04)	(9.39)	(62.02)
Reading Co	1.313	**-17.103~**	**-13.128***	-10.469
	(30.79)	(8.76)	(6.29)	(41.57)
Word Ident	-2.897	3.491	3.331	-15.073
	(15.35)	(4.35)	(3.14)	(20.70)
Vocab.	**99.742****	14.632	6.868	21.589
	(31.49)	(8.92)	(6.44)	(42.80)
Writing	21.782	**28.480****	3.955	-21.396
	(37.40)	(10.60)	(7.61)	(50.33)
Numbers	0.010	11.141	4.837	2.986
	(24.63)	(6.96)	(5.02)	(33.25)
Geometry	-4.247	-2.435	**5.059~**	-28.222
	(13.31)	(3.77)	(2.72)	(18.06)
Operations	-23.711	6.600	-4.693	-12.358
	(28.49)	(8.07)	(5.85)	(38.21)
Science	-1.349	-0.781	-1.935	-2.043
	(10.71)	(3.03)	(2.19)	(14.53)
Gross Motor	3.123	0.083	0.488	-12.371
	(10.95)	(3.10)	(2.25)	(14.82)
Social Studies	-10.638	-2.081	0.793	-2.515
	(8.48)	(2.40)	(1.73)	(11.42)
Aesthetics	-0.693	-1.930	-0.174	-7.843
	(8.13)	(2.30)	(1.66)	(10.93)
Collabor.	-12.457	0.644	-1.684	**-28.806****
	(7.88)	(2.23)	(1.61)	(10.68)
N	387	387	387	387

* 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.