

YEAR 4 REPORT: SEATTLE PRESCHOOL PROGRAM EVALUATION

September 2019

*Milagros Nores, Ph.D.,
Steve Barnett, Ph.D.,
Kwanghee Jung, Ph.D.
Gail Joseph, Ph.D., &
Lea Bachman, Psy.D.,
The National Institute
for Early Education
Research & Cultivate
Learning*

About the Authors

Milagros Nores, Ph.D. Dr. Nores is Co-Director for 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.

Kwanghee Jung, Ph.D. Dr. Jung is an Assistant Professor at The National Institute for Early Education Research (NIEER) at Rutgers University. Her expertise is in quantitative data analysis and the effect of participation in childcare and early education on children's learning and development.

Gail Joseph, Ph.D. Dr. Joseph is the Bezos Family Distinguished Professor in Early Learning at the University of Washington. She teaches courses, advises students, provides service and conducts research on topics related to early care and education. She is the Founding Executive Director of Cultivate Learning at the University of Washington (previously known as the Childcare Quality and Early Learning Center for Research and Professional Development, CQEL).

Lea Bachman, Psy.D. Dr. Bachman is a Research Associate at Cultivate Learning (CL) at the University of Washington. She leads CL's work on the SPP Evaluation Study and conducts research on topics related to early childhood education and assessment. She is a psychologist with significant experience in classroom observation, data collection, and management.

Grateful acknowledgment is made to Holly Campbell, Erica Johnson, and Seattle's Preschool Program for their support on this project. The authors are also grateful to Lea Bachman and Ran Guo for their assistance in producing this report.

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

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

Suggested citation: Nores, M., Barnett, W.S., Jung, K., Joseph, G. & Bachman, L. (2019). Year 4 report: Seattle Preschool Program evaluation. New Brunswick, NJ: National Institute for Early Education Research & Seattle, WA: Cultivate Learning.

Table of Contents

Table of Contents	4
Executive Summary	5
Introduction.....	8
Study Methods	8
Sample.....	9
Measures and Procedures.....	10
Methods.....	12
Results.....	12
1. Who enrolled in SPP in 2018–19, and how do they compare demographically to children in Seattle more generally?.....	12
2. What was the observed quality of children’s SPP classroom experiences in 2018–19, and did it improve in relation to prior years?	13
Average CLASS Scores	13
CLASS Dimension Scores	19
Average ECERS-3 Results	21
ECERS-3 subscales.....	24
3. How does quality vary within SPP and do children from different backgrounds experience different quality?.....	27
Classroom quality for Classrooms and FCCs separately	27
Classroom quality by curriculum.....	28
Classroom quality by year of entry into SPP	29
Classroom quality for children from different backgrounds.....	31
Attendance for children from different backgrounds.....	33
Associations between program features and quality.....	33
4. How did children in SPP classrooms and family child care providers progress in 2018–19/2017–18, and how did it vary with classroom quality? Other program characteristics? How did it vary with child characteristics?	34
Summary	42
References.....	43
Appendices.....	46
Appendix A. Measures.....	47
Measures on Children	47
Measures on Classrooms.....	47
Appendix B. Child Scores, pre, post and gains.....	51
Appendix C. Sensitivity Analyses.	60
Appendix D. P-values for tests of differences in means.....	65

Executive Summary

Fourth Year Evaluation (2018–19) of the Seattle Preschool Program

This is the fourth and final report of a four-year evaluation of the Seattle Preschool Program (SPP). The evaluation was conducted by the National Institute for Early Education Research at Rutgers University and Cultivate Learning at the University of Washington.

Covering the 2018-2019 school year, this report presents the results of our evaluation in relation to our prior evaluations. Areas examined include enrollment, quality of classroom experience, how children experience quality, child progress, among others.

SPP grew from 14 classrooms in 2015-16 to 75 today. It added 11 family child care providers when they were incorporated into the program during the 2017-2018 school year.

Overall Quality

Classroom Assessment Scoring System (CLASS) scores have maintained their upward trend throughout the four-year demonstration phase. From year one to year four, gains were measured in emotional support (6.14 to 6.61) and classroom organization (5.67 to 6.23). Despite exhibiting greater variation, instructional support scores have also improved (2.65 to 3.17).

Early Childhood Environmental Rating Scale—Third Edition (ECERS-3) average scores increased steadily from year two (3.89), year three (3.99), and year four (4.24).

SPP's quality, as measured by CLASS and ECERS-3, surpasses several major city and state pre-K programs and/or childcare systems and is now comparable to New York City's and San Antonio's preschool programs, which are well-established and widely known.

If quality continues improving, eventually rivaling programs with the highest measured levels of quality, strong impacts on children's long-term learning and development should be expected.

Family Child Care Provider Quality

Compared to the classroom setting, family child care providers (FCCs) saw lower overall quality scores. Controlling for classroom characteristics, FCCs scored lower on CLASS for instructional support quality but received generally high scores for emotional support and classroom organization.

Variations in Classroom Quality

Score variations for classrooms that have been part of SPP for three or four years versus those for one or two years appear to suggest quality has more to do with processes each year, and less with accumulative experience in SPP. Factors may include personnel rotation as well as other supports experienced by individual programs. Continuous improvement efforts should seek to raise lower scores.

Significant differences in average quality by gender or language were not seen, but modest differences were observed by race, ethnicity and in relation to the federal poverty line.

Child Gains

Children gained in every measured domain with some gains (language, literacy, and mathematics) larger than would be expected based on maturation (increased age) alone. We mostly saw no statistically significant differences in gains by gender, race or ethnicity, or language. Children identified as Asian made smaller gains in executive functions relative to their White peers, when accounting for other child and school characteristics. Children identified as Black made smaller gains in math and executive functions relative to their White peers. Lower gains were found for lower income levels in relation to the federal poverty line.

Students with Hispanic or Latino teachers had larger gains in literacy, reinforcing the importance of teacher diversity observed the previous year.

Differences in classroom size or curriculum, except for math, were not found to relate to children's performance. Stronger gains in math were found for children experiencing High Scope.

Higher quality as measured by the ECERS was found to relate to one of the executive function measures. Better quality in classroom organization as measured by the CLASS was associated with stronger vocabulary gains. The previous years' evaluations indicated better classroom organization as related to literacy and math. Teacher qualifications as measured by SPP were not found to be associated with child gains.

Student Profile

Compared to Seattle's K-12 public schools, the population of children enrolled in SPP are somewhat more likely to identify as African American or Black, Asian, Hispanic/Latino or Multi-racial, and less likely to identify as White.

Nearly a third of the program's total enrollment (30 percent), dual language learners identified as Hispanic or Latino (14 percent), White (22 percent), Black or African American (26 percent), and Asian (18 percent). The percentage of children identified as dual language enrolled in the program increased by 9 percent since the last report.

Recommendations

We recommend the program build on its early success by continuing to promote improvements in programmatic quality as it the program continues to expand.

It should emphasize language and literacy, seek to integrate content across developmental domains, and strengthen processes across activity settings.

The program's continuous improvement system should continue to build on existing quality through sustained, reflective, and intentional planning; use of data to improve personal care routine; feedback loops; language modeling; concept development; learning activities and materials; and gross motor play.

The SPP evaluation was conducted by National Institute for Early Education Research at Rutgers University and Cultivate Learning at the University of Washington. This report focuses on the 2018–2019 school year, presenting results in relation to the prior three years. This report concludes the evaluation for the four-year demonstration period.

Introduction

The City of Seattle has concluded its fourth and final year in its demonstration phase for the Seattle Preschool Program (SPP). Initially, SPP was established in 2014 after voter approval for a four-year, \$58 million property tax levy. The levy's proposition was of "accessible high-quality preschool services for Seattle children designed to improve their readiness for school and to support their subsequent academic achievement." In 2018, SPP was further supported by voters with a new education levy apportioning \$341.8 million to fund investments in early learning, including, among other priorities, the expansion of SPP.

SPP was initiated in 2015 by the city of Seattle's Department of Education and Early Learning (DEEL). In this first year, the program was provided in 14 classrooms. By 2016–17, the program more than doubled, operating in 32 classrooms. In 2017–18 it expanded to 48 classrooms and 13 family child care providers, and by 2018–19 it provided high quality preschool in 75 classrooms and 11 family child care providers.

SPP included an evaluation component from the beginning of the demonstration period, to inform its viability and support quality improvement efforts. The National Institute for Early Education Research at Rutgers University and Cultivate Learning at the University of Washington partnered with SPP to provide yearly evaluations throughout the four-year demonstration phase of the Seattle Preschool Program (SPP).

This report summarizes the findings for the fourth and final year (2018–19) of the demonstration phase. The report examines classroom quality and children's learning for this last year and in relation to previous ones. The report includes information on the children served by the program, their learning and development during the school year, and two areas of program quality across the years. It also assesses differences across specific subgroups of children and types of classrooms and examines associations between SPP children's learning gains and their classroom experiences.

Study Methods

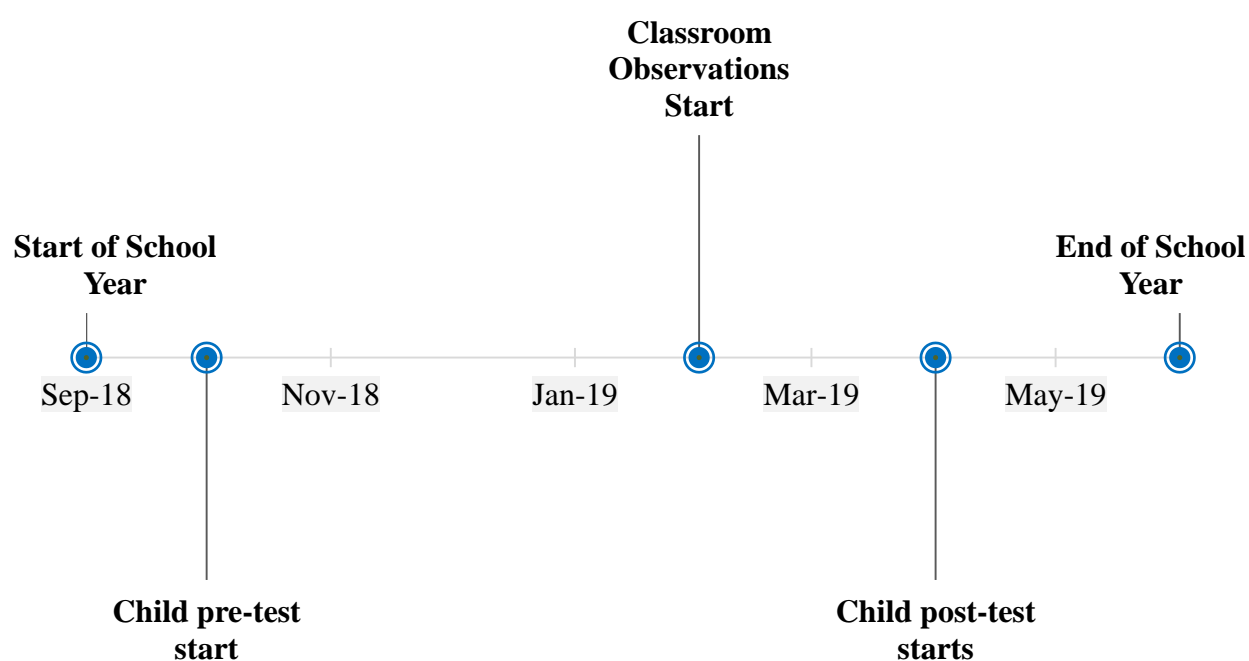
From its initiation, the SPP evaluation study was designed as a multi-site study blending various components so as to provide a comprehensive review of the program, its quality and how it served participating children through the demonstration period. The fourth year, as in previous years, included the collection of participating child and classroom information in order to address the following questions:

1. Who are the children enrolled in SPP in 2018–19, and how do they compare to children in Seattle more generally?
2. What was the quality of children's SPP classroom experiences in 2018–19, and did it improve in relation to prior years?
3. How does quality vary within SPP? Do children from different backgrounds experience different quality, on average?
4. How did children in SPP classrooms and family child care homes progress in 2018–19? How did quality vary in relation to classroom quality, other program characteristics, and child characteristics?

The SPP evaluation was framed to understand processes evolved over time. Initially, the evaluation included only the small number of classrooms that made part of the program. Over the years, new classrooms in centers and family child care providers have been incorporated. The research team measured learning and development at the beginning and at the end of the year, as well as classroom quality.¹ Measures and procedures used across all centers, the FCCs and for children are described below.

Children were assessed first early in the fall of 2018 and then assessed again (to measure progress) at the end of the 2018–19 school year (with a minimum of six months between assessments). Direct observations of classroom practices and processes were performed in February through early April 2019 to measure overall the quality of processes, practices and the early education environments, using observation protocols described further below.

Figure 1. Timeline.



Sample

In the 2018–19 school year, as part of this evaluation, the research team assessed 961 children in 75 SPP classrooms and 11 SPP family child care providers at pre- and post-test (840 with the full battery, 10 with the Spanish full battery and 111 only with the PPVT). To recruit children in the study, consent forms were distributed to families of all 1,078 children enrolled in the program. A total of 1,064 families consented to their children's participation in the study (processed at enrollment in the program by DEEL). From all the consenting children, we then randomly

¹ In the second year, the research team reiterated this process, and also recruited a non-equivalent comparison group. Last year, the research team measured learning and development at the beginning and end of the school year, as well as classroom quality in SPP classrooms and newly incorporated family child care providers (FCCs).

selected 12 children per classroom for the full battery and all consented children in the FCCs. A total of 24 children required language accommodations.²

The research team conducted classroom observations on the SPP classrooms and SPP family child care providers (FCC) in the spring of 2019. These are described in Table 1. Classrooms in SPP in Year 4 reported an average class size of about 17, and were distributed across seventeen agencies, with four classrooms on average per agency. FCCs are smaller in size with average class sizes of about 4 (not all preschool children), and all of them using Creative Curriculum. Teacher qualifications, race and ethnicity are also reported in Table 1. About 41% of teachers currently do not meet the program's qualifications.

Table 1. SPP Classroom and FCC characteristics in the 2018-19 school year, N=85

Classroom characteristic		SPP Classroom Frequency or Mean (SD ³)	SPP FCC Frequency or Mean (SD ¹)
Curriculum	Creative	30	11
	HighScope	45	-
Class Size^a		16.55 (3.32)	4.09 (2.59)
Agencies/Hubs		17	2
Teacher Qualifications	Unspecified	5.33%	100.00%
	Does not meet	41.33%	-
	Meets	34.67%	-
	Exceeds	18.67%	-
Teacher non-English Primary Language		33.33%	45.45%
Teacher Race and Ethnicity	Black or African American	12.00%	72.73%
	Hispanic or Latino	12.00%	9.09%
	White	45.33%	0.00%
	Asian	9.33%	0.00%
	Other/Unspecified	21.34%	18.18%
Average No. Classrooms per Agency/Hub		4.29 (5.54)	6.50 (0.71)

^aNumber of children in classroom as reported by director/roster in the Winter.

Measures and Procedures

Children were assessed with a measure of receptive language (the *Peabody Picture Vocabulary Test—Fourth Edition or PPVT-IV*; Dunn & Dunn, 2007), emerging literacy (the letter-word identification subtest from the *Woodcock-Johnson Psycho-Educational Battery—Third Edition*

² Language accommodations we made were to obtain verbal assent from children. All tests (except Spanish ones) were administered in English only. Teachers were the primary source for whether we needed to provide translation to obtain verbal assent from children. Most children understood enough English to be able to provide assent to do the activities.

³ SD stands for standard deviation, which is a measure of variation in the data. That is, it measures how close together or spread apart the classrooms are relative to the mean. The larger the value, the farther apart from the mean classrooms are, and the smaller the value, the closer to the mean classrooms are, in a specific indicator, such as classroom size.

or *WJ-III*; Woodcock, McGrew, Mather, & Schrank, 2001) and mathematics (the applied problems subtest from the *WJ-III*). In addition, children were assessed with two measures of executive functions, which capture children's inhibitory control, short term memory and attention. These are the *Dimensional Change Card Sort Task* (DCCS; Zelazo, 2006) and the *Peg Tapping Test* (PT; Diamond & Taylor, 1996). More detail on child measures is provided in Appendix A.

Classroom quality was captured using two observational instruments: the *Early Childhood Environment Rating Scale—Third Ed. (ECERS-3)*; Harms, Clifford & Cryer, 2014) and the *Classroom Assessment Scoring System Pre-K (CLASS Pre-K)*; Pianta, La Paro, & Hamre, 2008). Because a CLASS instrument does not exist for mixed aged groupings, Family child care providers were observed with three CLASS instruments using a *Combined CLASS Protocol* (Joseph, Feldman, Phillips & Jackson, 2010).⁴ More detail on classroom observation measures is provided in Appendix A.⁵

Data collection procedures were conducted by Cultivate Learning (CL) at the University of Washington. CL trained data collectors on standardized child assessments and classroom observation measures. Data collectors received a two-day training on the measures for child assessments, were given several days to practice, and were then tested for reliability on the assessments before starting data collection, and again halfway through the fall and spring data collection periods.

Classroom quality observations were conducted by trained and reliable observers on the observation protocol for the ECERS-3 and the CLASS protocols. ECERS-3 observers were trained by an ECERS-3 certified trainer and met the ERSI⁶ reliability requirements for observer certification. The trainee must complete score averages of 85% or above exact matches or one-away from the true score across three reliability observations. All data collectors met the ECERS-3 reliability requirements with agreement percentages ranging between 86–95%. CLASS observers were trained by a CLASS certified trainer and met the Teachstone reliability certification requirements. CLASS reliability⁷ agreement percentages ranged between 92–98%. Assessment procedures incorporated culturally sensitive attitudes, knowledge, interview skills, intervention strategies and evaluation practices specifically informed by the age of the children in the study.⁸

⁴ Protocol designed for Washington State's QRIS, Early Achievers. Also used in Oregon, see Tout, et. al, (2017).

⁵ Michigan School Readiness evaluation reported correlations between the High/Scope's Program Quality Assessment) PQA and ECERS scores at .86 (Xiang & Schweinhart, 2002). https://highscope.org/wp-content/uploads/2018/03/MSRP-through-age-10_2002-five-years-later.pdf

⁶ ERSI is the company that sells ECERS-3; for information on the tool and reliability go to <http://www.ersi.info/>

⁷ Teachstone is the company that sells CLASS products and manages CLASS certifications. All training activity is monitored and reported to them. <http://www.teachstone.com/about-teachstone/>.

⁸ Satisfaction surveys were delivered to providers after data collection which inquired about data collectors' adherence to prearranged procedures such as arriving on time, introducing themselves, being courteous and professional and whether the experience was positive. It also provided an opportunity for providers to offer open-ended comments. Fourteen responses were received: 100% agreed data collectors arrived when expected and checked in upon arrival, 93% said data collectors introduced themselves and thought it was a positive experience.

Methods

This report assesses classroom quality changes over time and differences across types of providers, using two tailed sample t-tests with unequal variances. We compared quality across the years, between FCCs and Classrooms and quality across selected subgroups. One-way ANOVAs, with Bonferroni multiple-comparison tests are used to test for differences in the quality of classroom processes and practices experienced by different subgroups of children (across race and ethnicity, by language indicators, and by FPL levels).⁹

Analyses of children's development over the school year, and in relation to previous years, is first described across subgroups and over the years (in terms of standard gains) and then explored using multivariate analyses. Multivariate analyses allow investigating the relationship between children's growth and child demographic information (and understanding whether there are subgroups of children gaining less from the preschool experience). These also permit exploring the association between children's fall to spring gains with school and classroom features.

Results

This section addresses each of the research questions stated earlier. As the last report of the demonstration phase, results are presented, when feasible, in ways that allow understanding change and improvements in the program over time. Results need to be particularly interpreted in a unique context of preschool expansion, where quality may be harder to increase upon. Main analyses draw from SPP classrooms. FCCs are incorporated later below in question 3, as well as in analyses in question 4. Questions 3 and 4 also incorporate characteristics of children enrolled in SPP classrooms and assessed in this study. All children in SPP were assessed in terms of receptive vocabulary. All other measures were only used in a random subsample of children.

1. Who enrolled in SPP in 2018–19, and how do they compare demographically to children in Seattle more generally?

Children's demographics for 2018–19¹⁰ are summarized in Table 2, below. Parallel demographics for children enrolled in Seattle Public Schools (who embody the SPP program target population) are also presented for comparison purposes. Children in the sample were mostly 4-year-olds (64%) and predominantly from English-speaking households (66%), with 30% speaking other languages, including Chinese, Vietnamese, Amharic, Somali, and Oromo, among others. Children more predominantly represented non-Whites than children in Seattle Public Schools, with 22% White, 26% Black, 18% Asian (decreasing from last year's 28%), 14% Hispanic, and 17% Multiracial/Other. About 76.5% of the children were under 300% of the Federal Poverty Level (FPL).

⁹ These categories are limited by what can be identified in this dataset. This is not indicative of importance over other categorizations, nor that there may not be important intersectional groupings as well.

¹⁰ Demographics were provided by DEEL.

Table 2. Child demographics for SPP study children relative to children in Seattle Public Schools

Child Characteristics	SPP Children 2018–19		Seattle Public Schools
	N	%	
Gender			
Female	580	48.8%	51.3% ^a
Male	606	51.0%	48.7% ^a
Age at Pre-Test			
3-Year-Olds	212	20.4%	-
4-Year-Olds	667	64.0%	-
Primary Language			
English	787	66.2%	-
Non-English	352	29.6%	21.7% ^a
Unknown	50	4.2%	-
Income			
20,000 or Less	365	30.7%	
21,000-40,000	224	18.8%	
41,000-60,000	205	17.2%	
61,000-80,000	126	10.6%	
81,000 or more	233	19.6%	
Unknown	36	3.0%	
FPL Percentage			
Less than 100%	405	34.1%	33.9% ^{a,c}
100 – 300%	504	42.4%	
≥ 300%	244	20.5%	-
Unknown	36	3.0%	
Race/Ethnicity			
White	258	21.7%	47.2% ^a
Black	309	26.0%	15.0% ^a
Asian	211	17.8%	14.0% ^a
Hispanic	167	14.1%	12.1% ^a
Multi-Racial/Other	198	16.7%	11.7% ^a

^aSeattle Public Schools as reported in http://www.seattleschools.org/district/district_quick_facts.

^bStudents attending Seattle Public Schools, as reported in Rivers (2016).

^cBased on Free and Reduce Lunch which is for families <185% FPL.

2. What was the observed quality of children’s SPP classroom experiences in 2018–19, and did it improve in relation to prior years?

Average CLASS Scores

Classrooms and FCC providers were observed using the CLASS Pre-K and combined CLASS protocol, respectively. Table 3 reports mean scores and standard deviations for three CLASS domains. Emotional Support and Classroom Organization experienced an increase relative to 2018 (0.40SD and 0.35SD, respectively). This increase was significant and similar to the increase experienced the year before. Statistically significant differences in the average scores compared to the previous year are indicated in Table 3 by an asterisk.¹¹

¹¹ Two-tailed two-sample t-test assuming unequal variances were used, P-values are reported in Appendix D.

Table 3. CLASS Dimension and Domain Means and Standard Deviations, spring semesters in 2016-2019.

CLASS Domains	Spring 2016 (N=14)		Spring 2017 (N=32)		Spring 2018 (N=48)		Spring 2019 (N=73) ^a		Spring 2019 – Including FCCs (N=84)	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Emotional Support	6.14	(0.53)	6.29	(0.47)	6.38	(0.57)	6.61	(0.36)*	6.54	(0.41)*
Classroom Organization	5.67	(0.74)	5.55	(0.76)	5.96	(0.77)*	6.25	(0.53)*	6.16	(0.59)*
Instructional Support	2.65	(0.71)	3.06	(0.88)*	3.42	(1.05)*	3.18	(0.85)	3.15	(0.82)*

^aTwo classrooms were not observed with the CLASS due to low enrollment of children above three years old in one of these (n=1) and delayed opening in the other classroom.

Linear trends on average scores in the CLASS domains over the evaluation years are depicted in Figure 2. Across all domains a positive trend is observed.

Figure 2. CLASS Domain linear trends on average scores, 2016-2019

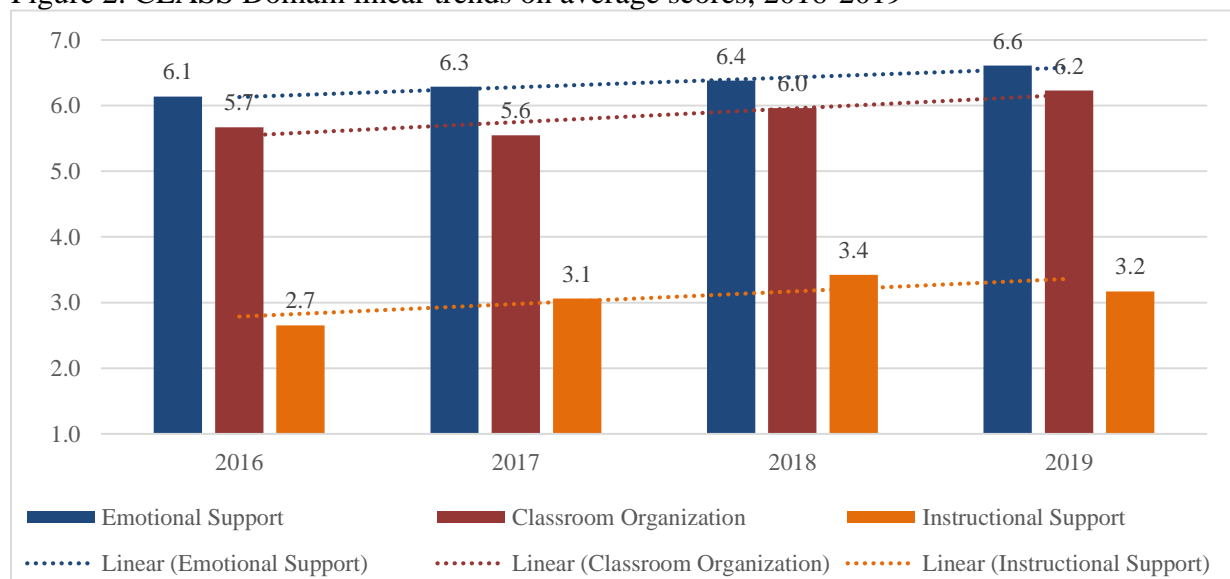


Table 4 and Figures 3 and 4 report the SPP CLASS scores in relation to other programs that have used the CLASS, and in relation to how these programs have performed over time. These has been reported for many high-quality programs.

Table 4. Classroom quality across the nation, and for selected programs

Study	Emotional Support	Classroom Organization	Instructional Support
SPP classrooms 2016 (N=14)	6.14 (0.53)	5.67 (0.74)	2.65 (0.71)
SPP classrooms 2017 (N=32)	6.29 (0.47)	5.55 (0.76)	3.06 (0.88)
SPP classrooms 2018 (N=48)	6.38 (0.57)	5.96 (0.77)	3.42 (1.05)
SPP classrooms 2019 (N=73)	6.61 (0.36)	6.25 (0.53)	3.18 (0.85)
Tulsa¹			
TPS pre-k (N=77)	5.23 (0.57)	4.96 (0.69)	3.21 (0.93)
CAP Head Start (N=28)	5.22 (0.78)	4.80 (0.84)	3.26 (0.94)
Boston² (N=83) (2009-2010)	5.63 (0.60)	5.10 (0.68)	4.30 (0.84)
NYC (N=555) (2012-13 to 2014-15)⁶	6.00	5.80	3.60
NYC (N=1,134) (2015-16)⁵	6.20	6.10	3.30
NYC (N=1,570) (2016-17)⁴	6.40	6.20	3.10
NYC (N=1,761) (2017-18)³	6.60	6.50	3.00
National Head Start Overview 2015⁷	6.03 (0.28)	5.80 (0.36)	2.88 (0.54)
Head Start FACES 2009⁸	5.30	4.70	2.30
EA Validation study (N=75) (2013-2014)⁹	5.96 (0.66)	5.26 (0.77)	2.34 (0.71)
NJ Abbott 2013-2014 (N=163)¹⁰	5.97 (0.63)	5.32 (0.89)	3.15 (0.96)
San Antonio (N=36) (2014)¹³	6.28 (0.35)	5.75 (0.60)	2.82 (0.82)
San Antonio (N=89) (2016)¹²	6.44 (0.51)	5.98 (0.81)	3.67 (1.23)
San Antonio (N=100) (2018)¹¹	6.65 (0.40)	6.35 (0.58)	3.85 (1.22)

¹Phillips et. al (2009); ²Weiland et. al (2013) (more recent analyses for a subsample of n=23 report ES scores of 5.71, CO scores of 5.57 and IS scores of 3.71; Boston K1KIDS, 2016); ³NYC Department of Education (2019); ⁴NYC Department of Education (2018); ⁵NYC Department of Education (2017); ⁶NYC Department of Education (2016); ⁷Office of Head Start. (2015); ⁸Aikens et. al (2013); ⁹CQEL (Unpublished); ¹⁰NIEER (2014), This is the only year CLASS was used; ¹¹EDVANCE (2018), ¹²EDVANCE (2016); ¹³EDVANCE (2014).

Figure 3. SPP 2019 CLASS scores by domain in relation to other programs

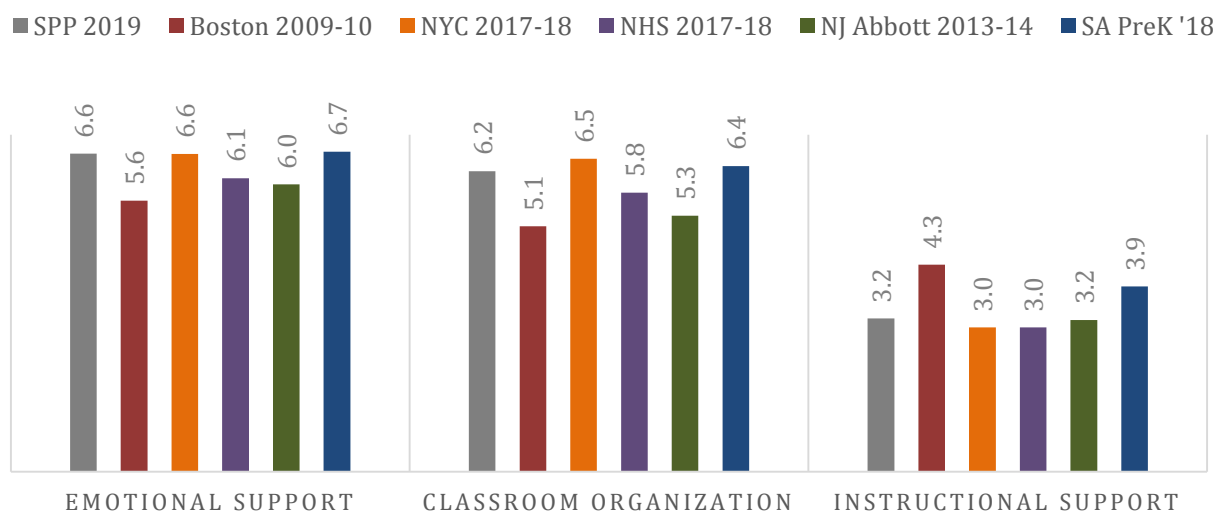
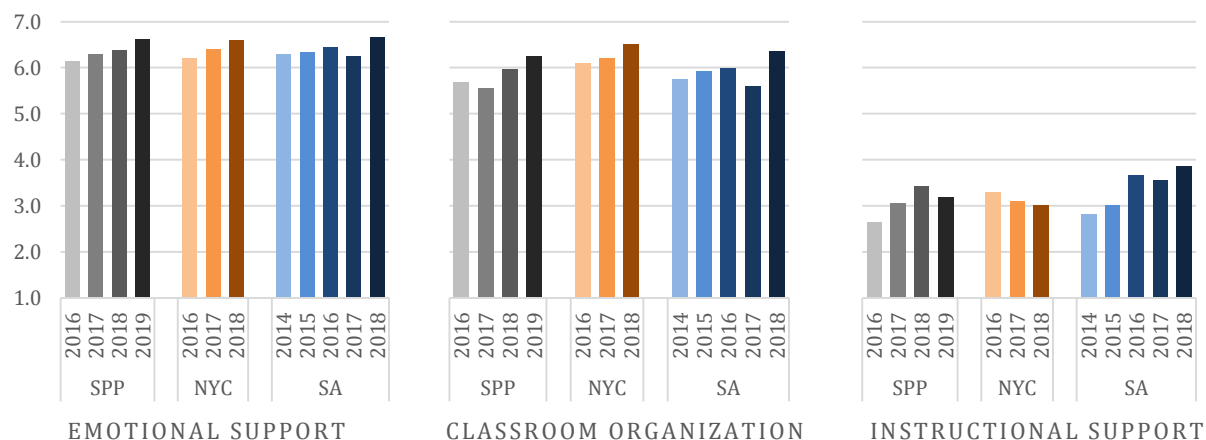


Figure 4. SPP CLASS trends in scores by domain in relation to trends in other recent city programs



Note: programs also showed longitudinally are NYC and San Antonio, for which CLASS has been assessed yearly as the programs grew.

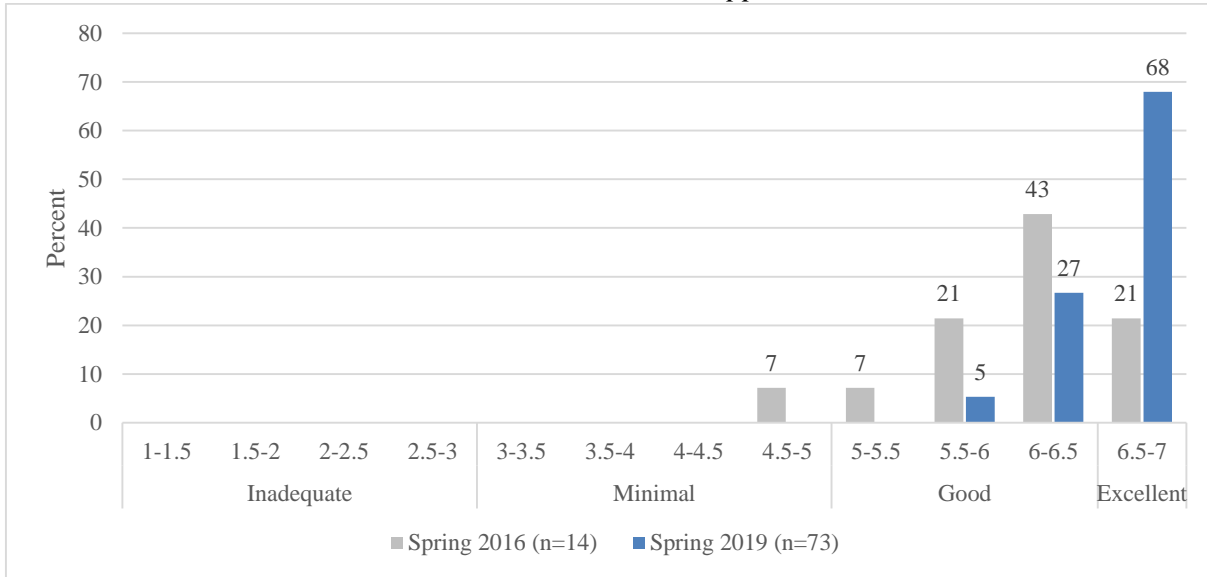
The CLASS distributions of classroom quality are depicted in Figure 5(a-c). These figures show histograms with the distributions for each CLASS domains for the spring of 2016 and 2019. Each bars represents the percentage of classrooms scoring within the interval. Classrooms scored exceptionally high on Emotional Support (ES) (100% percent scored above 5.5). Classroom Organization (CO) also had high scores, with a large portion of classrooms scoring above 5.5 (90%). Following national patterns, classrooms evidenced lower Instructional Support scores (IS), with 30% of the classrooms scoring above 3.5.

The 2018 CLASS score distributions show an increase in the number of classrooms with higher CLASS ES scores and a compression of the distribution towards higher scores overall. The distribution for CLASS CO (panel b) also shows a shift towards higher scores and a

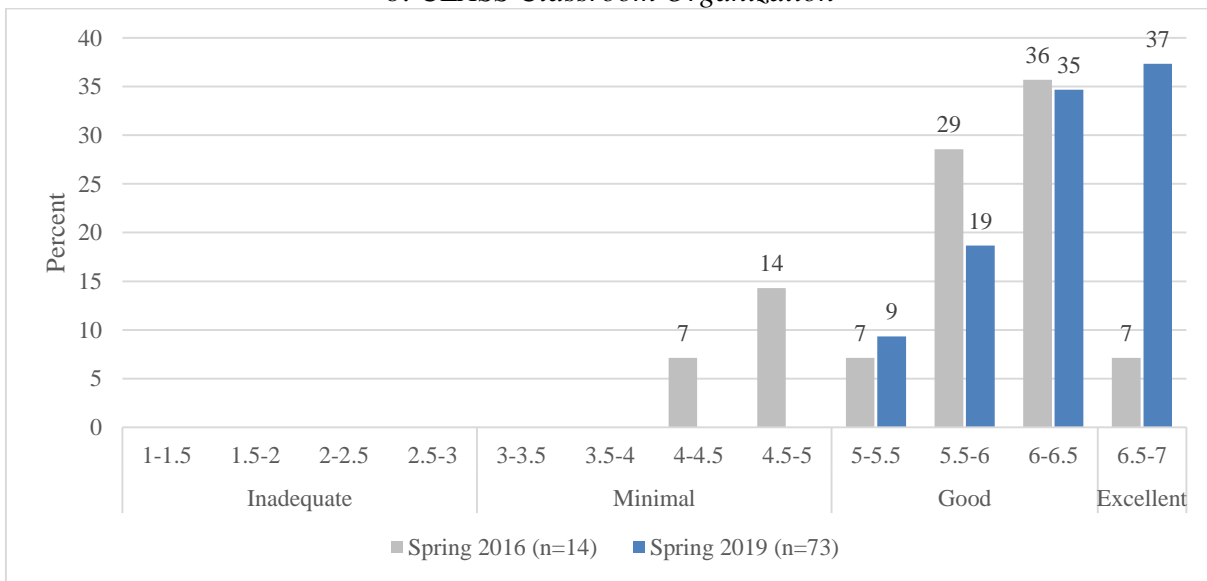
compression of the distribution within the 5-7 range. For CLASS IS (panel c) there is a shift towards the 2017 scores.

Figure 5. CLASS Domain distributions, 2016 and 2019

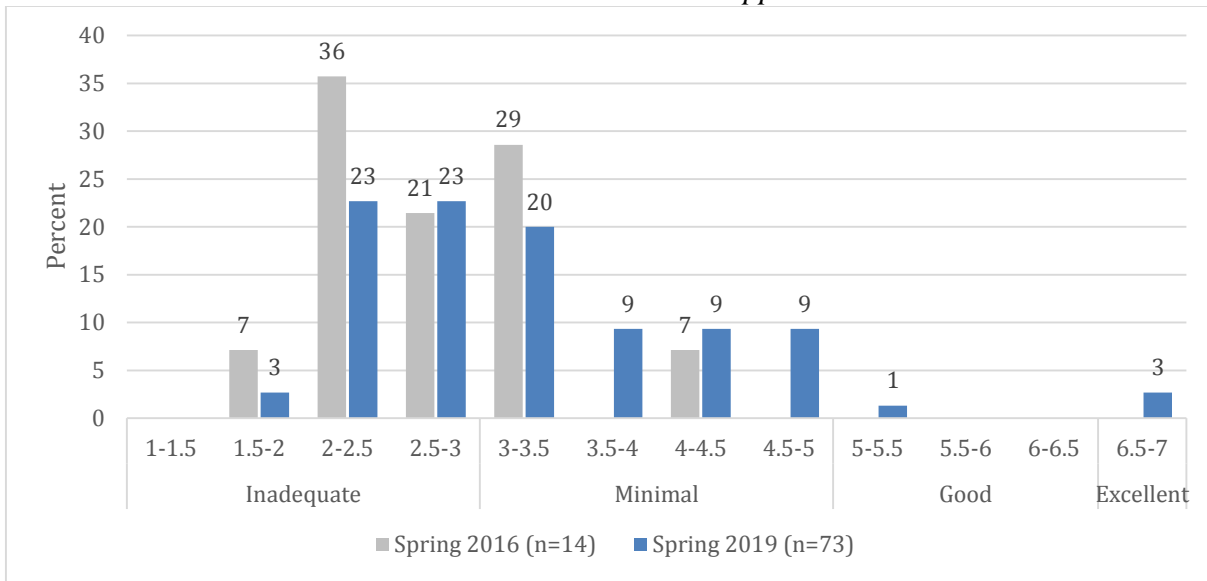
a. CLASS Emotional Support



b. CLASS Classroom Organization



c. CLASS Instructional Support



CLASS Dimension Scores

Table 5. CLASS Domains and Dimensions Means and Range by Item, 2016, 2017, 2018. & 2019 (excludes FCCs).

CLASS Dimensions and Domains	2016 Mean (Range) N=14	2017 Mean (Range) N=32	2018 Mean (Range) N=48	2019 Mean (Range) N=73
<i>Emotional Support</i>	6.14 (4.88-6.81)	6.29 (5.19-7.00)	6.38 (4.19-7.00)	6.61 (5.56-7)
1. Positive Climate	5.80 (4.25-7.00)	6.33 (5.25-7.00)	6.23 (3.00-7.00)	6.56 (6.00-7.00)
2. Negative Climate*	6.86 (5.75-7.00)	6.95 (6.63-7.00)	6.94 (5.00-7.00)	6.97 (6.00-7.00)
3. Teacher Sensitivity	5.91 (4.25-6.75)	6.04 (4.25-7.00)	6.23 (4.00-7.00)	6.45 (5.00-7.00)
4. Regard for Student Perspectives	5.96 (4.25-7.00)	5.96 (4.25-7.00)	6.04 (4.00-7.00)	6.40 (5.00-7.00)
<i>Classroom Organization</i>	5.67 (4.17-6.58)	5.55 (3.42-6.83)	5.96 (3.75-6.92)	6.25 (5.00-7.00)
5. Behavior Management	5.73 (3.75-7.00)	5.46 (3.50-6.75)	5.98 (3.00-7.00)	6.29 (4.00-7.00)
6. Productivity	6.05 (4.50-7.00)	5.91 (3.50-7.00)	6.06 (4.00-7.00)	6.36 (5.00-7.00)
7. Instructional Learning Formats	5.21 (3.50-6.50)	5.21 (3.00-6.75)	5.69 (3.00-7.00)	6.08 (4.00-7.00)
<i>Instructional Support</i>	2.65 (1.50-4.25)	3.06 (1.67-5.75)	3.42 (1.75-6.33)	3.18 (1.50-5.33)
8. Concept Development	2.07 (1.25-3.50)	2.64 (1.25-5.50)	2.63 (1.00-6.00)	2.16 (1.00-5.00)
9. Quality of Feedback	2.61 (1.50-4.25)	3.03 (1.50-5.50)	3.40 (2.00-6.00)	3.29 (2.00-6.00)
10. Language Modeling	3.29 (1.75-5.00)	3.57 (1.75-6.25)	4.19 (2.00-7.00)	4.03 (2.00-6.00)

Note: (*) The Negative Climate dimension was transposed so that on here, high represents “good”.

CLASS: Emotional Support Domain

Scores for dimensions under the three CLASS domains are reported in Table 5. The Emotional Support (ES) domain focuses on teacher’s promotion of nurturing and safe environments for children’s learning. Classrooms scored in all dimensions in this domain above 6. “Positive” and “Negative” climate dimensions assess the emotional connection between teachers and students.¹² Negative Climate scores are inverted for reporting. These continue to score highest (6.94, on average), demonstrating a lack of expressed negativity. The “Teacher Sensitivity” dimension

¹² Positive Climate “reflects the emotional connection between the teacher and students and among students and the warmth, respect, and enjoyment communicated by verbal and nonverbal interactions” (Pianta, La Paro & Hamre, p.23). Negative Climate “reflects the overall level of expressed negativity in the classroom” (p. 28).

focuses on teachers' capacity to anticipate problems, plan accordingly, provide individual support and problem resolution to support children effectively (averaging 6.47, increasing from 6.23 in 2018 and 6.04 in 2017). This high score implies consistency in teachers' awareness of children who need assistance or support, responsiveness to their needs, abilities, problems and emotions, understanding of children's non-verbal cues, and active encouragement to make children feel comfortable to seek support and share thoughts. "Regard for Student Perspectives" (average of 6.40, increasing from 6.04 in 2018 and 5.96 in 2017) measures the capacity of teachers to intentionally and consistently emphasize children's interests, motivations, opinions and perspectives. It includes assessing the degree to which teachers encourage student responsibility, autonomy, decision-making and participation. Additional opportunities for children to share their thinking and feeling, to choose their free play, to initiate activities and learning, and increased responsiveness and acknowledgement of their interests and choices, would further increase this score.

CLASS: Classroom Organization Domain

Classroom Organization assesses teachers' management and redirection of behavior, management of instructional time and routines, and management of activities to increase students' interests. In particular, the "Behavior Management" dimension considers teachers' ability to specify well-defined behavioral expectations and to enforce them consistently. This includes proactively preventing problems and effective redirection and reduction of misbehaviors by concentrating on positive aspects of children's behavior. "Productivity" is centered on teachers' management of time, particularly their capacity to maximize learning opportunities, to pace appropriately, and to be efficient with transitions throughout the day and across activities. This was the highest scoring dimension under CO, with no classrooms scoring below a 5. "Instructional Learning Formats" assesses teachers' facilitation of student learning during activities, including effective questioning, stating clear learning objectives, using modalities and materials to engage and expand children's learning, and including hands-on opportunities. These three dimensions increased in relation to 2017 and to 2018. On average, all three score above a 6. In all three dimensions, increases were of about 2/3 of a standard deviation or greater.

CLASS: Instructional Supports Domain

The domain of Instructional Supports measures the type and amount of interactions through which teachers enable high-order thinking skills and reasoning, provide feedback, encourage creativity, scaffold children and promote language development. This domain is particularly important in order to be able to have an impact on students' learning. In most programs assessed with CLASS, this domain has been the most challenging to increase. Research using the class has consistently found Instructional Support (IS) scores lag behind the other two domains. IS scores have overall increased on average in the four-year evaluation period, but there are some variations year to year.

"Concept Development" assesses how and to what degree teachers use discussions to stimulate reasoning, analysis, and understanding. This includes challenging children to think and communicate on the how and the why, and on generating solutions and ideas. This includes teachers' ability to ask open-ended questions that encourage children to assess, brainstorm, and plan, to connect concepts to themselves and their lives, to integrate information with what they

already know and have experienced, and to troubleshoot. Consistency and intentionality are key to concept development. Concept Development lagged behind the other two dimensions in this domain (average 2.16). Increasing scores requires consistent use of discussions and activities to promote problem solving, prediction, comparison, planning and real-world applications. Fostering child comparisons, contrasting, categorizations, coding, and communication are central mechanisms to raising children's understanding of concepts. "Quality of Feedback" (average 3.28) evaluates how and to what extent teachers scaffold children, engage in back-and-forth exchanges and follow-ups, utilize metacognitive approaches with children that expand on their thinking processes, and encourage children to persist. The "Language Modeling" dimension assesses both quality and quantity of teacher's language used to promote children's language development (average 4.02). This includes promoting peer conversations, back-and-forth exchanges between teachers and children, use of open-ended questions that engage children, extension and elaboration in conversations, as well as encouraging complex responses. Self-talk and parallel talk model language for children. This also includes the use of specific and advanced language.

Average ECERS-3 Results

ECERS-3 scores for 2016 through the spring of 2019 are reported in Table 6. The ECERS-3 is the most updated revision of the ECERS and looks at choice and access to materials for an hour, does not depend on teacher reports, and assess the role of teachers in children access and interaction with materials.¹³ Included in this table are the mean scores, standard deviations, as well as the minimum and maximum scores observed for the ECERS-3 subscales and overall for classrooms that participated in SPP. FCCs are excluded as ECERS is not an observation tool suitable for family child care providers. Average ECERS-3 scores increased relative to 2018 (a 0.40 SD increase) while the program continued to expand the number of classrooms. Classrooms scored in the range of 2.44 through 5.29. The increase in the overall score was driven by increases in space and furnishings (0.23 SD), personal care routines (0.69 SD), language and literacy (0.40 SD) and learning activities (0.47). Variation of scores among programs remained similar to previous years, and evidence that the expansion of number of programs did not bring in more classrooms at the lower end of the spectrum. Statistically significant differences in the average compared to the previous year are marked with an asterisk,¹⁴ as is the case for the increase between 2018 and 2019.

¹³ The most important differences between the ECERS-3 and the previous version, ECERS-R, are: (a) ECERS-3 considers only what is observed during the 3-hour time sample (additional time may be added ONLY to review materials or the safety features of the playground), but does not include teacher reports as ECERS-R did; (b) it can be used only in classrooms with children ages 3-5 (ECERS-R could be used in classroom with children 2½); (c) it has 35 Items with 6 Subscales, a reduction from the previous 43 Items with 7 Subscales; (d) Requires less attention to accessible materials and more attention to how the teachers use the materials to stimulate children's learning; (e) Major hazards for indoor and outdoor have been separated into separate indicators; and (f) Discards the concept of Substantial portion of the day and access to materials must now be observed for 1 hour of the 3-hour assessment.

¹⁴ Two-tailed two-sample t-test assuming unequal variances were used, P-values are reported in Appendix D.

Table 6. ECERS-3 Item, Subscale, and Overall Means and Standard Deviations, Spring semesters in 2016 through 2019.

ECERS-3 Item and Subscales	Spring 2016 (N=14)		Spring 2017 (N=32)		Spring 2018 (N=48)		Spring 2019 (N=73)*	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Overall	3.57	(0.46)	3.89	(0.55)*	3.99	(0.63)	4.24	(0.63)*
Space and Furnishings	3.88	(0.55)	3.94	(0.61)	4.25	(0.80)	4.43	(0.78)
Personal Care Routines	3.14	(0.65)	3.41	(0.86)	2.67	(0.85)	3.26	(0.78)
Language and Literacy	3.47	(0.83)	3.93	(0.82)	4.22	(0.92)	4.59	(0.96)
Learning Activities	2.87	(0.56)	3.26	(0.57)	3.45	(0.66)	3.76	(0.72)
Interaction	4.49	(0.90)	4.99	(1.07)	5.12	(0.99)	5.07	(1.00)
Program Structure	4.43	(0.97)	4.67	(0.88)	4.76	(1.01)	4.72	(0.99)

*Two sites were not observed with ECERS-3 due to low enrollment of children above age 3 in one classroom (n=1) and delayed opening in the other classrooms.

The linear trend for the ECERS-3 average scores over the four years of the study are depicted in Figure 6. The trend is positive and shows continued growth.

Figure 6. ECERS-3 average scores and linear trend, 2016-2018

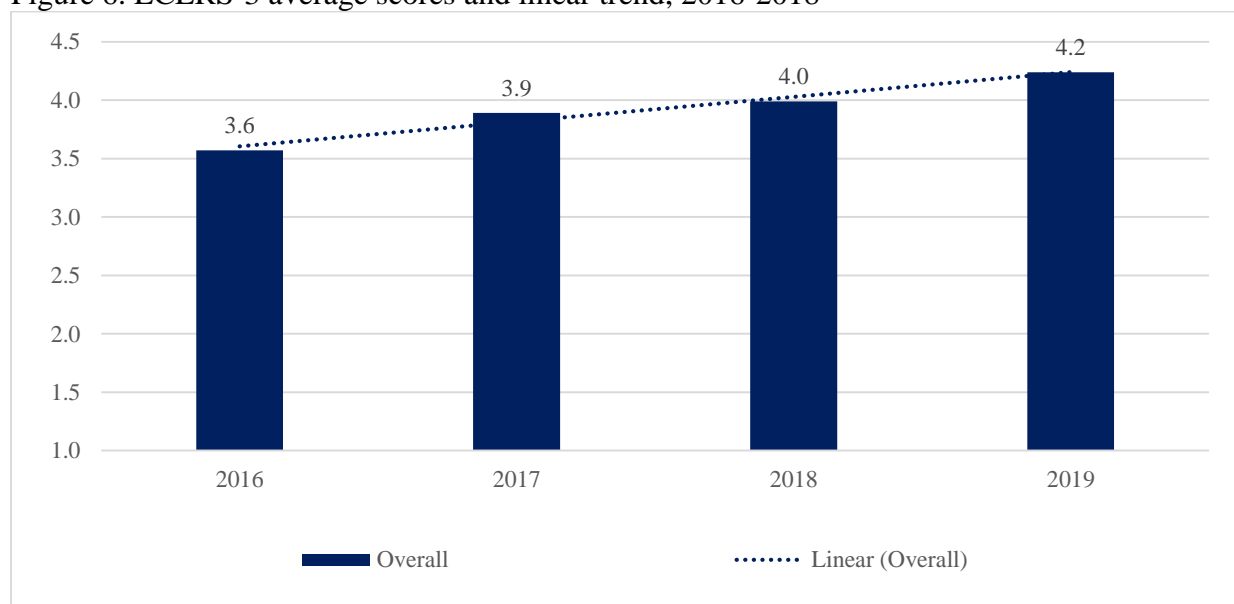


Table 7 and Figure 7 report the SPP ECERS-3 scores in relation with other programs/studies that have used the ECERS-3: GA, PA, WA state pre-K and childcare centers and NJ Abbott districts.¹⁵ Overall and subscale scores are included in Table 3 (standard deviations included when available). SPP classrooms score above GA, WA and PA, and closely to the last reported averages for NJ Abbott’s average score for Space and Furnishings, Interaction and Program Structure. Areas that score somewhat lower relative to NJ Abbott are

¹⁵ The ECERS-3 is still not as widely used as the ECERS-R, which does not allow for comparisons with many high-quality programs.

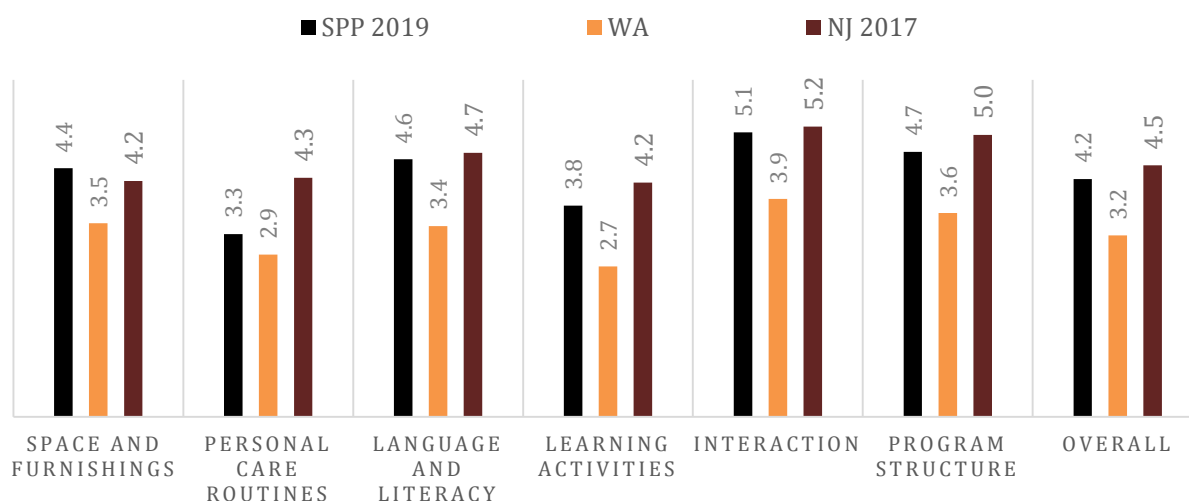
Personal Care Routines, and Learning Activities. Average scores are compared in table and Figure 7. SPP has grown consistently between 2016-2019, despite also going through expansion, in all areas but personal care routines and program structure.

Table 7. Studies with reported ECERS-3 scores

Study	Space/ Furnishing	Personal Care Routines	Language & Literacy	Learning Activities	Interaction	Program Structure	Average Total
SPP 2016 (N=12)	3.88 (0.55)	3.14 (0.65)	3.47 (0.83)	2.87 (0.56)	4.49 (0.90)	4.43 (0.97)	3.57 (0.46)
SPP 2017 (N=32)	3.94 (0.61)	3.40 (0.86)	3.93 (0.82)	3.26 (0.57)	4.99 (1.07)	4.67 (0.86)	3.89 (0.55)
SPP 2018 (N=48)	4.25 (0.80)	2.67 (0.85)	4.22 (0.92)	3.45 (0.66)	5.12 (0.99)	5.12 (1.01)	3.99 (0.63)
SPP 2019 (N=73)	4.43 (0.78)	3.26 (0.78)	4.59 (0.96)	3.76 (0.72)	5.07 (1.00)	4.72 (0.99)	4.24 (0.63)
GA¹	3.49	3.14	3.36	3.14	4.31	3.64	3.46
WA state pre-K & childcare (2013-14) (N=299)²	3.45	2.89	3.40	2.68	3.88	3.63	3.23
PA³	3.74	3.77	3.77	2.93	4.72	4.10	3.68
GA, PA, WA (2015-16) (N=1063)⁴	3.62	3.36	3.62	2.97	4.41	3.92	3.53
NJ Abbott: 2015-16 (N=293)⁶	4.43 (1.02)	4.36 (1.33)	4.86 (1.26)	4.22 (1.17)	5.26 (1.34)	5.20 (1.31)	4.61 (1.03)
2016-17 (N=300)⁵	4.20 (0.84)	4.26 (1.14)	4.70 (1.10)	4.17 (1.11)	5.17 (1.30)	5.02 (1.38)	4.48 (0.92)

¹Jenson (2015); ²CQEL (Unpublished); ³PAKEYS (Unpublished); ⁴Early et. al (2018), subscales estimated from paper ⁵NIEER (2017); ⁶NIEER (2016).

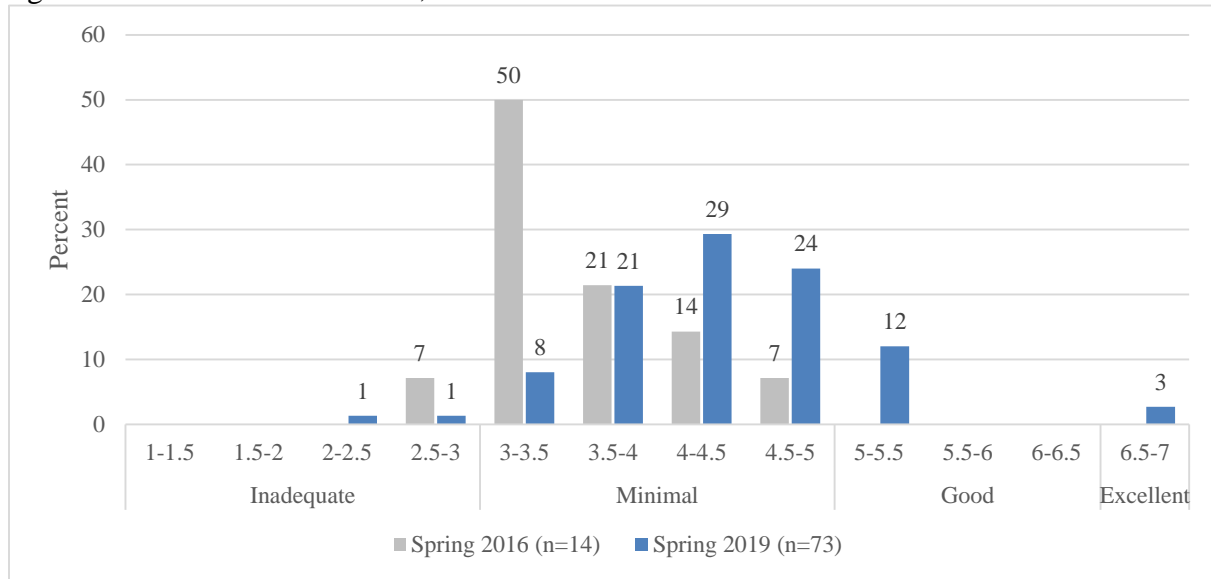
Figure 7. SPP ECERS-3 scores by subscale in relation to other selected programs



The ECERS-3 distributions of classroom quality throughout these expansion years in the demonstration phase are depicted in Figure 8. A histogram with the distributions for ECERS-3 are depicted for the spring of 2016 and 2019. Each bars represents the percentage of classrooms

scoring within the interval. Continuous improvement in classroom quality is observed across time. By the spring of 2019, classrooms scored on average slightly below the good quality threshold of 5 in the ECERS-3. The difference in means over time is statistically significant.¹⁶

Figure 8. ECERS-3 distributions, 2016 & 2019



ECERS-3 subscales

Scores for the ECERS-3 subscales and items are summarized in Table 8 for 2016, 2017, 2018, and 2019. These include average scores and the ranges for each item and subscale. The range is the minimum and maximum scores in classrooms.

¹⁶ Two tailed two-Sample ttest Assuming Unequal Variances, P-value=0.002 for differences between the last 2 years and P<0.001 for differences between earlier 2016, 2017 and 2019.

Table 8. ECERS-3 Item, Subscale, and Overall Means and Ranges by Item, 2016-2019

ECERS-3 Item and Subscales	2016 Mean (Range) N=14	2017 Mean (Range) N=32	2018 Mean (Range) N=48	2019 Mean (Range) N=73
<i>Space and Furnishings</i>				
1. Indoor space	6.43 (4-7)	5.47 (2-7)	5.40 (2-7)	5.71 (3-7)
2. Furnishings for care, play and learning	4.36 (4-7)	4.56 (3-7)	4.44 (3-7)	4.48 (2-7)
3. Room arrangement for play and learning	3.64 (2-7)	4.72 (2-7)	5.04 (2-7)	4.26 (2-7)
4. Space for privacy	4.14 (2-6)	4.53 (1-7)	4.63 (1-7)	5.56 (1-7)
5. Child-related display	3.36 (1-5)	3.09 (1-4)	4.29 (1-7)	3.88 (1-7)
6. Space for gross motor play	3.14 (1-4)	3.06 (1-6)	3.10 (1-4)	3.85 (1-7)
7. Gross motor equipment	2.07 (1-4)	2.13 (1-5)	2.81 (1-6)	3.25 (1-7)
<i>Personal Care Routines</i>				
8. Meals/ snacks	3.07 (1-4)	3.88 (1-7)	2.90 (1-5)	5.71 (3-7)
9. Toileting/diapering	2.21 (1-3)	3.19 (1-7)	2.79 (1-6)	4.48 (2-7)
10. Health practices	2.93 (2-4)	2.69 (1-5)	1.88 (1-5)	4.26 (2-7)
11. Safety practices	4.36 (2-7)	3.88 (1-7)	3.13 (1-7)	5.56 (1-7)
<i>Language and Literacy</i>				
12. Helping children expand vocabulary	3.50 (3-5)	3.63 (1-7)	4.63 (3-7)	4.78 (3-7)
13. Encouraging children to use language	4.36 (3-7)	4.84 (3-7)	5.15 (2-7)	5.71 (4-7)
14. Staff use of books with children	3.07 (1-6)	3.50 (1-6)	3.79 (1-7)	4.25 (1-7)
15. Encouraging children's use of books	4.21 (1-7)	4.41 (3-6)	4.08 (1-7)	4.47 (2-7)
16. Becoming familiar with print	2.21 (1-4)	3.25 (1-6)	3.44 (1-5)	3.73 (2-6)
<i>Learning Activities</i>				
17. Fine motor	4.36 (2-5)	4.47 (2-7)	4.88 (2-7)	5.04 (2-7)
18. Art	3.71 (2-6)	4.28 (1-7)	4.15 (1-7)	4.67 (1-7)
19. Music and movement	3.50 (2-5)	3.47 (2-6)	3.58 (1-5)	3.70 (2-7)
20. Blocks	2.00 (1-4)	2.97 (1-5)	3.13 (1-7)	2.84 (1-6)
21. Dramatic Play	2.79 (1-6)	3.50 (1-7)	3.77 (1-7)	4.23 (1-7)
22. Nature/science	2.50 (1-4)	2.28 (1-5)	2.73 (1-6)	2.95 (1-7)
23. Math materials and activities	1.71 (1-3)	2.25 (1-4)	2.42 (1-6)	2.93 (1-7)
24. Math in daily events	2.86 (1-5)	3.34 (1-5)	4.29 (1-7)	4.47 (1-6)
25. Understanding written numbers	1.29 (1-2)	1.69 (1-5)	1.44 (1-3)	1.96 (1-6)
26. Promoting acceptance of diversity	4.21 (3-6)	4.34 (2-6)	4.06 (3-6)	4.81 (3-7)
27. Appropriate use of technology*	N/A	N/A	N/A	N/A
<i>Interaction</i>				
28. Supervision of gross motor	3.71 (1-7)	4.56 (1-7)	4.67 (1-7)	4.36 (1-7)
29. Individualized teaching and learning	4.21 (3-7)	4.94 (2-7)	5.33 (3-7)	5.62 (4-7)
30. Staff-child interaction	4.93 (3-7)	5.66 (3-7)	5.96 (2-7)	5.82 (2-7)
31. Peer interaction	5.00 (3-7)	4.84 (1-7)	4.85 (1-7)	4.85 (1-7)
32. Discipline	4.57 (2-7)	4.97 (2-7)	4.77 (2-7)	4.70 (1-7)
<i>Program Structure</i>				
33. Transitions and waiting times	4.86 (3-7)	4.75 (3-7)	5.21 (2-7)	5.26 (1-7)
34. Free play	4.50 (3-6)	4.44 (2-7)	4.58 (3-7)	4.33 (2-7)
35. Whole group activities for play & learning	3.93 (2-5)	4.81 (2-6)	4.50 (2-6)	4.58 (2-7)

Note: (*) Only 2 classrooms received a score for #27, both were 1. All others were N/A.

The *Space and Furnishings subscale* incorporates indoor and outdoor space and furniture, the arrangement of the furniture (allows for learning, ample movement, routine care, exploration play and free choice). It also includes accessibility to at least five interest centers, as well allowing for adequate supervision. This subscale also includes child related display which should be mostly related to children’s interests, and which shows a decrease in relation to the previous year. The items for “gross motor equipment” and “space for gross motor play” were the lowest scoring in this subscale. These measure the space, the toys and equipment to interest the children, the lack of hazards, and that at least 30 minutes of gross motor play are provided. “Indoor space” scores increased in the last year to 5.71. Four items within this subscale scored with a range starting at 1, indicating some classrooms are exhibiting inadequate levels. However, this year all seven items under this subscale showed classrooms scoring at the excellent level (7).

The *Personal Care Routines* subscale addresses health, hygiene and safety practices in the classroom. It includes flexibility in the schedule, sanitary requirements, conversations during meals, individualized toileting, hygiene practices of teachers and children, and safety. All the items under this subscale show strong improvements relative to previous years. Only one of them, “Safety practices” has classrooms scoring at 1, the inadequate level.¹⁷ However, both “Safety practices” and “Meals/snacks” both score on average above 5 (5.56 and 5.71, respectively), the level for good.

Language and Literacy assesses how staff manage and support activities and materials in order to support children’s language and literacy development. All items in this subscale show an increase in relation to the previous two years. “Becoming familiar with print” is the lowest scoring item on average across classrooms (3.73).¹⁸ The item for “Staff use of books” averaged 4.25 this year (up from 3.79 in 2018 and 3.50 in 2017).¹⁹ Only one of the five items in this scale evidence classrooms scoring at “1” (inadequate), while in the other items minimums were at 2, 3 (minimal) and even 4. The low score of 4 occurred on the item “Encouraging children to use language,” which averaged the highest at 5.71 (good).

The *Learning Activities* subscale captures the presence, variety, and accessibility of learning materials in the classroom for children, intertwined with the extent to which teachers show interest, engagement and intentionality as children use different types of materials. The items for “Fine motor,” “Art,” “Math in daily events,” and “Promoting acceptance of diversity” scored higher within this scale, 5.04, 4.67, 4.47, and 4.81, respectively. In particular, the average for the fine motor item was above a 5 (good). In addition, they all increased relative to the previous year. Four items on average persist under the minimal score of 3: “Blocks,”

¹⁷ Four classrooms scored at 1. These classrooms had issues with indicator 1.1, In particular, important hazards were observed outdoor that included insufficient cushioning and/or inappropriate climbing equipment.

¹⁸ This item expects most visible print be combined with pictures, staff showing print as a useful tool and explaining how and why it is used, as well as staff writing down what children say in a way that is interesting and engaging. It also includes encouraging writing for advanced children.

¹⁹ A score in the good (5) to excellent (7) range on this item is attained when all children are observed to be actively engaged during story time, when there is accessibility to books for at least 1 hour, books are well organized and staff show positive interest in children’s independent book choices.

“Nature/science,” “Math materials and activities,” and “Understanding written numbers.” “Blocks” is the only item that did not show improvements relative to the previous year.²⁰

The *Interaction* subscale is centered on children’s supervision during gross motor time, teachers’ individualization of teaching and learning and interactions among children and between teachers and children. With the incorporation of additional classrooms, the average for three items showed decreases relative to 2018, although increasing relative to 2016. These were “Supervision of gross motor,”²¹ “Staff-child interaction,”²² and “Discipline.” The items for “Individualized teaching and learning” and “Staff-child interaction” score on average above 5 (good), with averages at 5.62 and 5.82, respectively. All items have some classrooms scoring at 7 (excellent).

The *Program Structure* subscale assesses general formats of the classroom, use of different activity settings (e.g. choice, whole-group, small-group, transitions) and how children spend their time. “Transitions and waiting times” and “Whole group” showed increases on average scores, now averaging 5.26 (up from 5.21 in 2018 and 4.75 in 2017) and 4.58 (up from 4.50 in 2017). “Free play” has minimally decreased over the course of the study.²³

3. How does quality vary within SPP and do children from different backgrounds experience different quality?

Classroom quality for Classrooms and FCCs separately

Table 9 describes center-based classrooms in relation to FCCs in terms of quality as measured by the CLASS. While there are slight differences in the measures due to the FCC measurements encompassing care for babies and toddlers, the structure of levels necessary for quality is consistent across all versions of the CLASS.²⁴ Table 10 reports average CLASS scores by domains for classrooms and FCCs. All domains are observed to be on average higher for SPP classrooms than for SPP FCCs and in most dimensions higher maximum scores are observed only in the classroom group. Statistically significant differences in mean scores across domains or dimensions are marked with an asterisk and when this is the case, higher scores are bolded.²⁵ Unlike the previous year, differences are observed in most dimensions in ES, all dimensions in CO, and for the quality of feedback dimension under IS.

²⁰ An increase in scores would require enough space, blocks, and accessories for 3 types of independent structures, materials to be stored mostly in open and labelled shelves, and with accessibility for at least 1 hour of the observation. Engagement and interest from teachers strengthen this item even further.

²¹ Increasing supervision of gross motor play requires ensuring safety through careful supervision, mostly or only positive interactions and staff engagement and interest.

²² Increases for the interactions items include frequent and positive staff-child interactions and consistency throughout the observation period. It also includes a relaxed and trustful atmosphere and warmth through appropriate physical contact. Positive guidance when discipline challenges occur is important for this item.

²³ This decrease is not statistically significant.

²⁴ We also estimated alphas for consistency within domains within the CLASS Pre-K used in the 73 classrooms and the CLASS combined used in the 11 FCCs. Both of these were closely consistent (with alphas between 78%-96%).

²⁵ Two-tailed two-sample t-test assuming unequal variances. P-values in Appendix D.

Table 9. CLASS Domain and Dimension scores for classrooms in centers and FCCs

	Classrooms in Centers				FCCs			
	Mean	(SD)	Min.	Max.	Mean	(SD)	Min.	Max.
<i>Emotional Support*</i>	6.61	0.36	5.56	7.00	6.04	0.39	5.15	6.60
1. Positive Climate*	6.56	0.50	6.00	7.00	5.98	0.41	5.00	6.50
2. Negative Climate^a	6.97	0.16	6.00	7.00	7.00	0.00	7.00	7.00
3. Teacher Sensitivity*	6.45	0.58	5.00	7.00	5.45	0.60	4.50	6.25
4. Regard for Student Perspectives*	6.40	0.59	5.00	7.00	6.05	0.49	5.25	6.75
<i>Classroom Organization*</i>	6.25	0.53	5.00	7.00	5.56	0.61	4.13	6.25
5. Behavior Management*	6.29	0.72	4.00	7.00	5.73	0.78	3.75	6.75
6. Productivity^{a,*}	6.36	0.51	5.00	7.00	5.68	0.59	4.25	6.25
7. Instructional Learning Formats^{b,*}	6.08	0.62	4.00	7.00	5.43	0.68	4.00	6.50
8. Facilitation of Learning & Dev.^c	n/a	n/a	n/a	n/a	2.95	1.21	0.00	4.75
<i>Instructional Support*</i>	3.18	0.85	1.50	5.33	2.29	0.69	1.50	3.75
9. Concept Development^a	2.16	0.80	1.00	5.00	2.07	0.70	1.25	3.25
10. Quality of Feedback*	3.29	1.07	2.00	6.00	2.68	0.62	1.75	3.50
11. Language Modeling	4.03	1.09	2.00	6.00	3.80	0.65	2.75	4.75

Note: ^aThe Negative Climate dimension was transposed so that here, high represents “good”. ^bThese three are scored only for pre-K children in the combined protocol used in FCCs. ^cFacilitation of Learning and Development is not included in the CLASS Pre-K, but is part of the CLASS Infant and Toddler measure.

Classroom quality by curriculum

Classrooms using Creative scored slightly higher on average than classrooms using High Scope across the ECERS total and subscale scores (Table 10). On the other hand, High Scope classrooms scored higher on CLASS ES and CLASS CO and most of the dimensions for these (Table 11). However, classrooms using Creative also scored higher for the CLASS IS domain and dimensions. Statistically significant differences are marked with an asterisk and for these cases, higher scores bolded.

Table 10. ECERS total and subscales score by curriculum used

	Creative				High Scope			
	Mean	(SD)	Min.	Max.	Mean	(SD)	Min.	Max.
<i>Overall*</i>	4.42	(0.50)	3.50	5.26	4.12	(0.68)	2.44	5.29
Space and Furnishings	4.43	(0.62)	3.43	6.00	4.43	(0.88)	2.71	6.29
Personal Care Routines*	3.59	(0.71)	2.50	5.75	3.04	(0.76)	1.75	5.00
Language and Literacy*	5.03	(0.92)	3.60	6.60	4.29	(0.87)	2.80	6.00
Learning Activities	3.88	(0.70)	2.20	5.30	3.69	(0.73)	2.00	5.10
Interaction	5.23	(0.92)	2.40	6.80	4.94	(1.04)	2.40	6.80
Program Structure	4.93	(0.88)	3.33	6.67	4.58	(1.05)	2.33	6.67

Table 11. CLASS Domain and Dimension scores by curriculum used

	Creative				High Scope			
	Mean	(SD)	Min.	Max.	Mean	(SD)	Min.	Max.
<i>Emotional Support*</i>	6.44	(0.43)	5.15	7.00	6.62	(0.37)	5.56	7.00
1. Positive Climate*	6.35	(0.53)	5.00	7.00	6.59	(0.50)	6.00	7.00
2. Negative Climate	6.98	(0.16)	6.00	7.00	6.98	(0.15)	6.00	7.00
3. Teacher Sensitivity*	6.10	(0.74)	4.00	7.00	6.50	(0.59)	5.00	7.00
4. Regard for Student Perspectives*	6.18	(0.55)	5.00	7.00	6.50	(0.59)	5.00	7.00
<i>Classroom Organization</i>	6.13	(0.63)	4.13	7.00	6.18	(0.56)	5.00	7.00
5. Behavior Management	6.20	(0.69)	4.00	7.00	6.25	(0.78)	4.00	7.00
6. Productivity	6.28	(0.64)	4.00	7.00	6.27	(0.50)	5.00	7.00
7. Instructional Learning Formats*	6.00	(0.64)	4.00	7.00	6.02	(0.66)	4.00	7.00
8. Facilitation of Learning & Dev.	2.90	(1.29)	0.00	5.00	-	-	-	-
<i>Instructional Support</i>	3.28	(0.78)	2.17	5.33	3.04	(0.84)	1.50	4.92
9. Concept Development*	2.33	(0.86)	1.00	5.00	2.00	(0.68)	1.00	4.00
10. Quality of Feedback	3.33	(1.05)	2.00	6.00	3.11	(1.08)	2.00	6.00
11. Language Modeling	4.10	(1.01)	2.00	6.00	3.91	(1.07)	2.00	6.00

Classroom quality by year of entry into SPP

We describe differences in quality between new classrooms in the program, and those two, three or four years in the program. These are summarized in Tables 12 and 13 and Figures 8 and 9 for ECERS-3 and CLASS scores by years in SPP. Classrooms with more years in the program scored slightly higher on the overall ECERS-3 score and most subscales, and scored slightly higher in the CLASS domains.

Table 12. ECERS-3 Subscale, and Overall Means and Ranges, 2017 (N=73)

ECERS-3 Item and Subscales	1 year in SPP (N=28)		2 years in SPP (N=18)		3 years in SPP (N=21)		4 years in SPP (N=6)	
	Mean	Mean	Mean	(SD)	(SD)	(SD)	Mean	(SD)
Overall	4.24	4.20	4.15	(0.69)	(0.64)	(0.60)	4.66	(0.46)
Space and Furnishings	4.45	4.41	4.28	(0.61)	(0.84)	(0.85)	4.84	(0.86)
Personal Care Routines	3.47	3.06	3.08	(0.98)	(0.72)	(0.70)	3.46	(0.29)
Language and Literacy	4.71	4.53	4.29	(0.92)	(1.07)	(0.90)	5.20	(0.80)
Learning Activities	3.77	3.65	3.81	(0.74)	(0.81)	(0.69)	3.92	(0.66)
Interaction	4.86	5.27	4.95	(1.13)	(0.81)	(1.03)	5.67	(0.62)
Program Structure	4.46	4.78	4.81	(0.95)	(1.02)	(1.00)	5.45	(0.75)

Figure 8. ECERS-3 subscales by cohort of program partnership with SPP

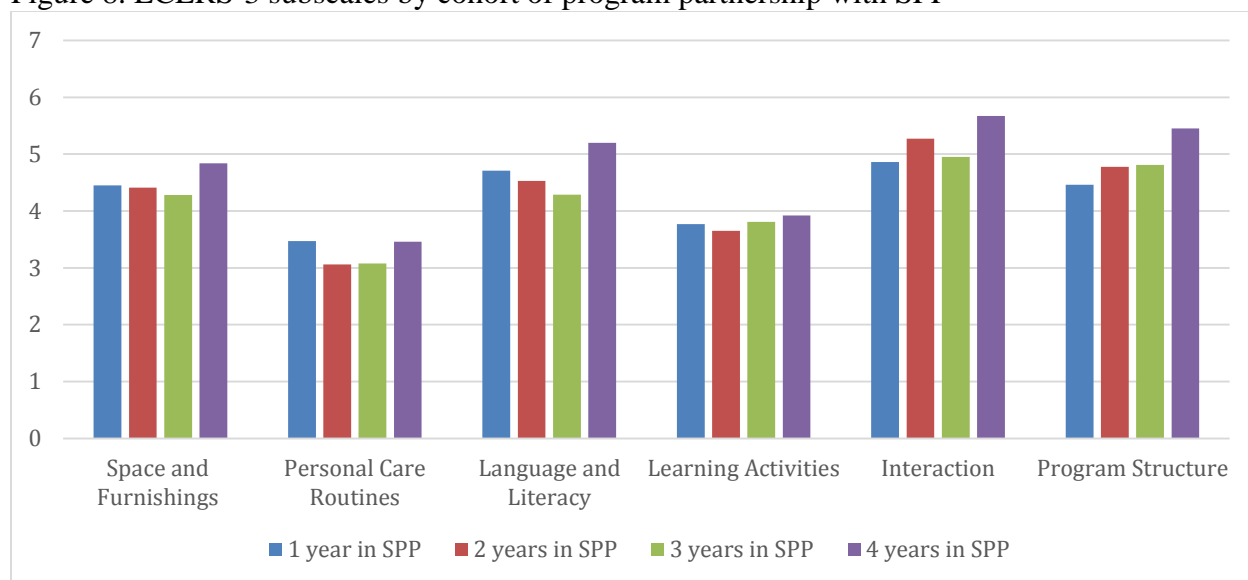
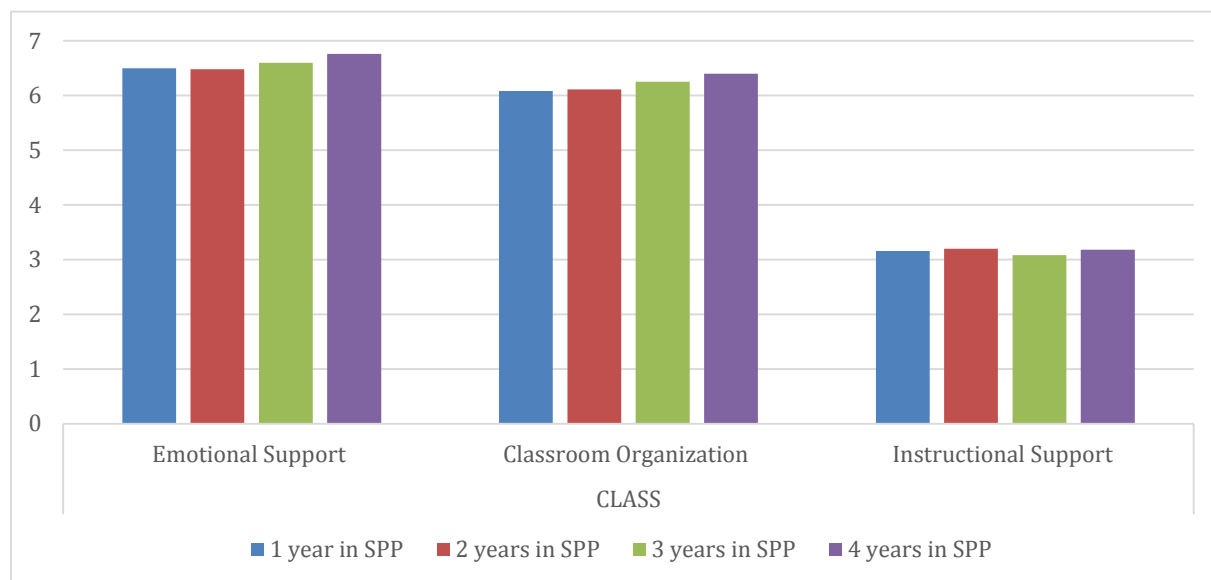


Table 13. CLASS Domain Means and Ranges, 2018 (N=84)

CLASS Domains	1 year in SPP (N=32)		2 years in SPP (N=27)		3 years in SPP (N=21)		4 years in SPP (N=6)	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Emotional Support	6.50	(0.48)	6.48	(0.40)	6.60	(0.34)	6.76	(0.24)
Classroom Organization	6.08	(0.64)	6.11	(0.61)	6.25	(0.52)	6.40	(0.44)
Instructional Support	3.16	(0.75)	3.20	(0.87)	3.08	(0.93)	3.18	(0.65)

Figure 9. CLASS domains by cohort of program partnership with SPP

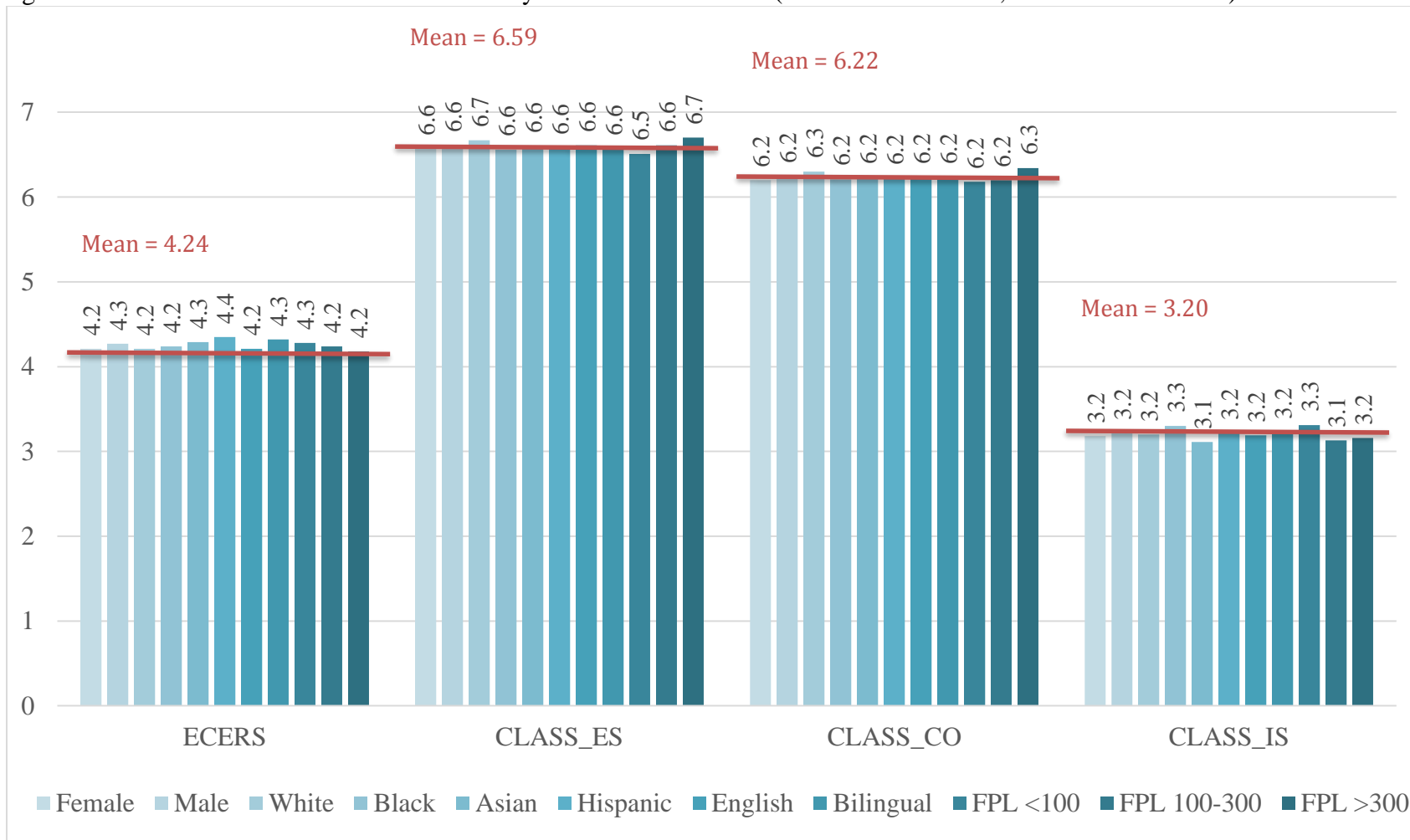


Classroom quality for children from different backgrounds

The quality of care by children's gender, ethnicity/race, language background and FPL for the SPP children in the sample is shown in Figure 10. The figure does not evidence strong differences in any of the measures across groups. Significance of differences between groups of children were run²⁶ and no statistically significant differences were found on quality by gender. Differences for dual language learners (DLL) were found only on the ECERS. In relation to the federal poverty level (FPL) a statistically significant difference was present for all CLASS domains, between families by FPL.

²⁶ One-way anova, with Bonferroni multiple-comparison tests for race/ethnicity, DLL and FPL, and Two-tailed t-test with unequal variances for gender. P-values in Appendix D.

Figure 10. ECERS and CLASS Domain scores by Child Characteristics (N=1129 for CLASS, N=1176 for ECERS)

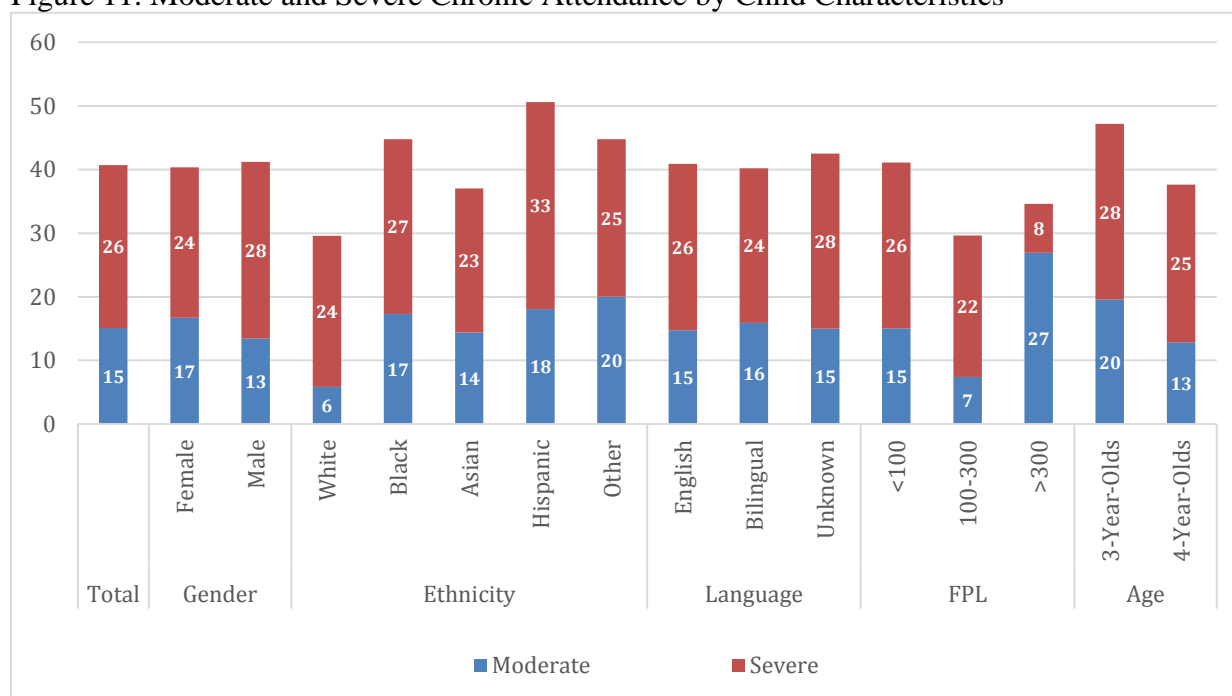


Note: Includes classrooms and FCCs (the latter only for the CLASS domains).

Attendance for children from different backgrounds

Chronic absence is defined as missing 10 percent or more of school days and has been shown to be related to later achievement and school progress.²⁷ We estimated moderate chronic absenteeism (missing greater than or equal to 10% and less than 20% of preschool) and severe chronic absenteeism (missing 20% or more of preschool) for children in the sample (Figure 11). Severe and chronic absenteeism combined varied between almost 30 and 50 percent depending on children’s characteristics for children in SPP. These rates were altogether higher for children identified as Black (44%) or Hispanic/Latino (51%) and for 3-year-olds (48%). For comparison, in Newark (NJ), preschool chronic absenteeism in 2017-18 was 48% (53% for Black children and 44% for Hispanic/Latino children; ACNJ, 2019). Chronic absenteeism in Chicago in preschool were 45% of three-year-olds and 36% of four-year-olds in Chicago (23-57% for Black children and 30-35% for Latinx children Ehrlich, Gwynne, & Allensworth, 2018). Rates in SPP, Newark and Chicago are larger than what has been found for other large cities nationally (22% in Baltimore: Connolly & Olson, 2012; 27% in Washington, DC: Dubay & Holla, 2016).

Figure 11. Moderate and Severe Chronic Attendance by Child Characteristics



Associations between program features and quality

We estimated the association between measured program features and levels of classroom quality. We used multi-level regression models that took into account that classrooms clustering at the agency level (Table 15). These associations are reported below for the ECERS-3, for CLASS pre-K (classrooms) and for CLASS pre-K and the CLASS combined (in classrooms and FCCs).

²⁷ <https://www.attendanceworks.org/chronic-absence/the-problem/>

Creative curriculum showed positive association with ECERS, CLASS IS quality regardless of including the FCCs (relative to the reference group, which is High Scope). Class size was positively associated with CLASS IS. Classrooms with teachers who exceeded the required qualifications showed higher quality of CLASS CO and CLASS IS relative to classrooms with teachers that just met or did not meet required qualifications (the latter is the reference group). FCCs evidenced a negative association with quality as measured by the CLASS IS, after controlling for classroom characteristics.

Table 15. Association between classroom quality and program features

	ECERS	CLASS ES	CLASS CO	CLASS IS
Creative Curriculum	0.368** (0.14)	0.019 (0.08)	0.243 (0.13)	0.547** (0.17)
Class Size	0.021 (0.02)	-0.004 (0.01)	-0.011 (0.02)	0.035 (0.02)
Teacher Qual. Meets	0.208 (0.16)	0.122 (0.09)	0.268 (0.14)	0.116 (0.19)
Teacher Qual. Exceeds	0.365 (0.19)	0.205 (0.12)	0.506** (0.17)	0.720** (0.24)
Missing T. Qual.	0.753 (0.41)	-0.337 (0.25)	-0.069 (0.37)	1.360** (0.50)
Teacher Black	-0.322 (0.22)	-0.021 (0.12)	0.062 (0.18)	-0.028 (0.25)
Teacher Hispanic	-0.029 (0.22)	-0.077 (0.13)	-0.191 (0.19)	-0.594* (0.26)
Teacher Asian	-0.054 (0.23)	-0.289* (0.14)	-0.137 (0.21)	-0.443 (0.29)
Teacher Other	0.256 (0.27)	-0.069 (0.17)	0.080 (0.25)	0.254 (0.34)
FCC		-0.234 (0.30)	-0.760 (0.44)	-1.285* (0.60)
<i>N</i>	73	84	84	84

Note: Reference groups are teacher not meeting qualifications, teacher identifies as White and classroom is center-based. * p<0.05; ** p<0.01; *** p<0.001.

4. How did children in SPP classrooms and family child care providers progress in 2018–2017–18, and how did it vary with classroom quality? Other program characteristics? How did it vary with child characteristics?

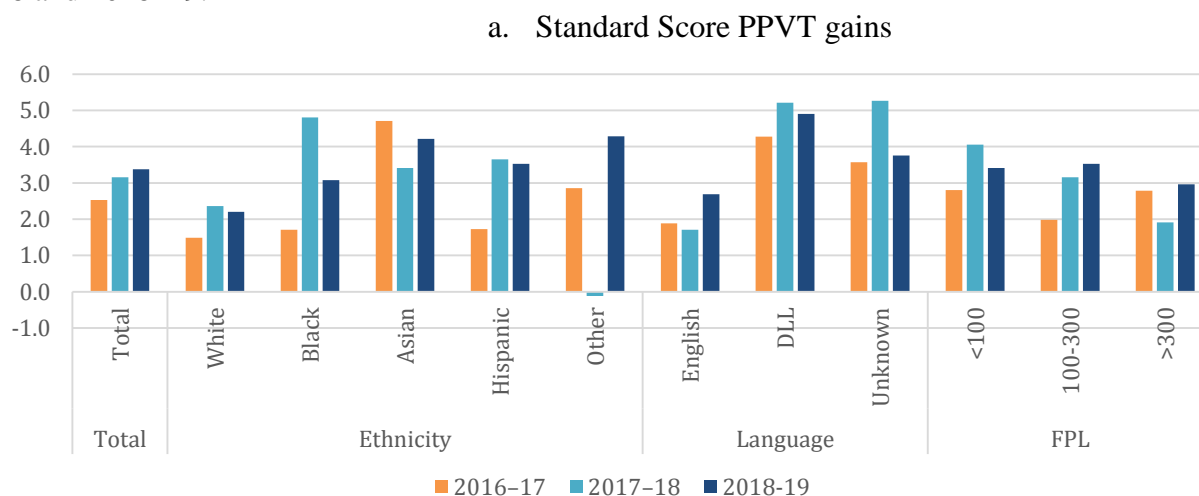
Child outcomes were measured in receptive vocabulary (with the Peabody Picture Vocabulary Test), literacy (with the Woodcock-Johnson Tests of Achievement Letter-Word subtest), and math (with the Woodcock-Johnson Tests of Achievement Applied Problems subtest). Executive functioning (EF) was also assessed using two measures: the Dimensional Change Card Sort (DCCS) and the Peg Tapping task (PT). These two assess a combination of short-term memory, the ability to inhibit automatic response tendencies that can interfere with achieving a task, and the capacity for set shifting.

The PPVT (vocabulary) and Woodcock-Johnson (literacy and math) assessments provide standardized scores that allow for comparisons to expected gains after controlling for age. Positive gains in these standard scores indicate that children gained more than other children after adjusting for age (that is, beyond the growth expected from child maturation alone). Child gains for the 2018–19 school year for all children in the SPP sample are depicted in Figure 12 and reported in detail in Appendix B. Depicted for comparison is also growth for 2016–17 and 2018–19. All children were assessed with PPVT and only a random sample was assessed with the rest of the battery.

Children’s gains in standard scores have increased over time in the PPVT, although not in the LW and AP measures. Children did also improve on the DCCS measure of executive functions. Additional trends observed are: (a) gains across all measures in relation to the standards (age alone), (b) larger fall to spring gains for children without an identified ethnicity or race and DLs, relative to other subgroups of children, (c) smaller vocabulary and letter word gains in relation to the previous year for children identified as Black, DLL and under 100% FPL.

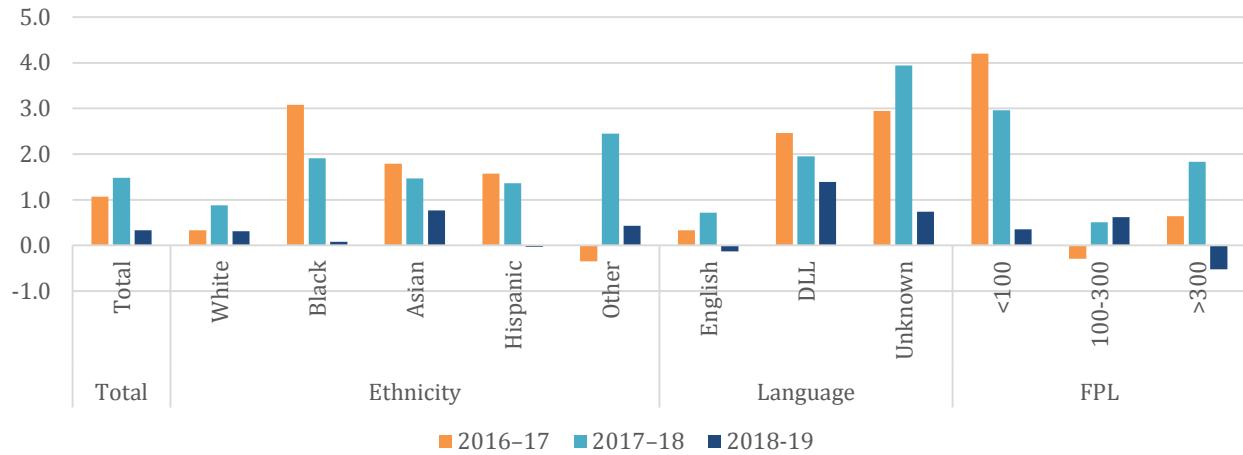
On average, the 2018-19 cohort scored slightly lower than the previous year’s cohort, therefore differences in gains only slightly contribute to differences between cohorts. About 18 percent of the variance in scores is due to classroom differences and the rest are within classrooms differences. This implies that variations in children and teachers within classrooms is what explain overall gains for children. Some variation year to year is also the result on differences in the rates of children that report identifying of a particular ethnic or racial group, or even the as dual language learners. Increases or changes in the composition of subgroups (particularly likely due to program expansion) would also contribute to variations across the years.

Figure 12. Child gains across the different measures by child demographics for 2016-17, 2017-18 and 2018–19.



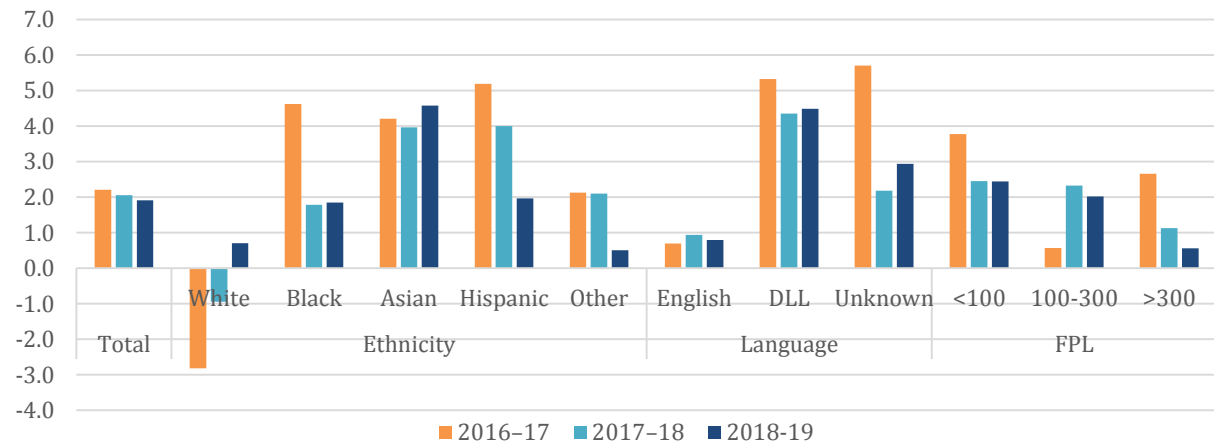
Note: Other includes children without an identified ethnicity, or that identified as from an ethnicity or race that was not White, Black/African American or Hispanic/Latino.

b. Standard Score LW gains



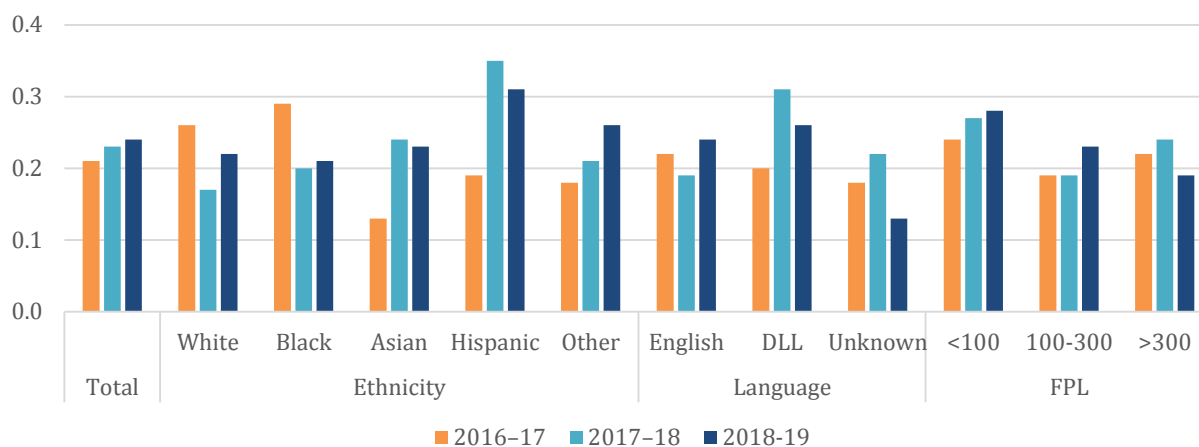
Note: Other includes children without an identified ethnicity, or that identified as from an ethnicity or race that was not White, Black/African American or Hispanic/Latino.

c. Standard Score AP gains



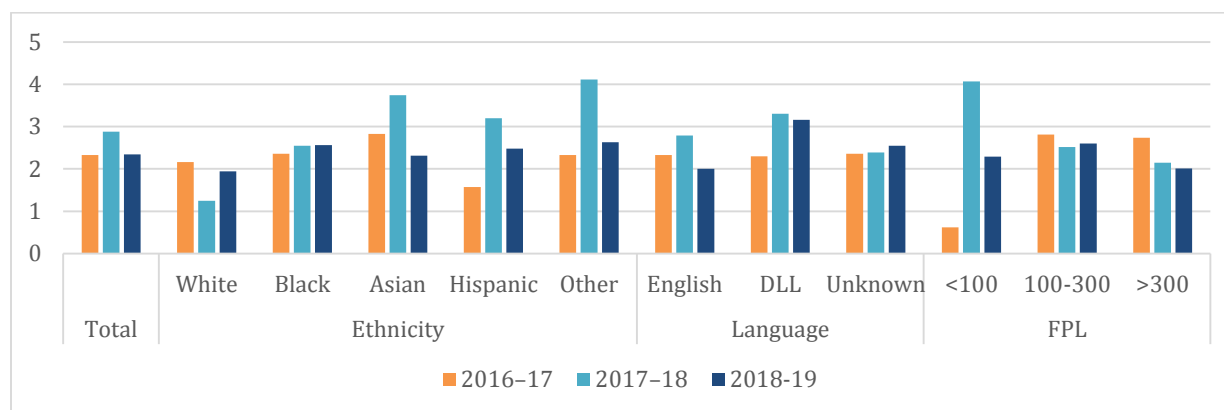
Note: Other includes children without an identified ethnicity, or that identified as from an ethnicity or race that was not White, Black/African American or Hispanic/Latino.

d. DCCS gains



Note: Other includes children without an identified ethnicity, or that identified as from an ethnicity or race that was not White, Black/African American or Hispanic/Latino.

e. PT gains



Note: Other includes children without an identified ethnicity, or that identified as from an ethnicity or race that was not White, Black/African American or Hispanic/Latino.

This next section focuses on assessing if differences in the school year trajectory existed across these subgroups of children through estimations that relate various children’s characteristics to children’s gains in the measures included in the study (controlling for school features).

Multivariate analyses also allow exploring whether there are associations between children’s learning gains and program features while taking into account children’s characteristics. We incorporate demographics on the children such as their age, gender, race and ethnicity, and home language, as well as household demographics such as income and, specifically how it relates to the Federal Poverty Level. Program features for SPP include class size²⁸, curriculum used (whether it is Creative or High Scope), teacher race and ethnicity, teacher

²⁸ Measured as a continuous variable.

degree and classroom quality. We also account for the fact that children in the same classroom or FCC program should not be considered to be independent of each other.

Tables 16-18 present the estimates of the associations of program features and child characteristics with children's development. We performed separate analyses with the two measures of quality, one controlling for quality as measured by the ECERS-3 (Table 16), and the other for quality as measured by the CLASS dimensions for classrooms in centers only (Table 17), as well as including FCCs (Table 18). Statistically significant results are highlighted in bold. For categorical variables, such as being female, the results need to be interpreted in relation to the omitted group (i.e. males).

In terms of children's characteristics, we find evidence of disadvantages for children identified as Black across child outcomes such as receptive vocabulary (standard and raw), math (standard and raw), and one of the executive function measures (DCCS). Children identified as Hispanic evidence lower math gains than their White peers. Asian children also showed lower gains in executive function measured by DCCS. Dual language children showed higher gain scores in literacy (standard score), math (standard score), and one of the executive function measures (Peg Tapping, or PT). However, when measured by raw scores, dual language children showed lower gains in receptive vocabulary. Children from lower income families (Household income less than \$20,000) evidence greater gains than their peers in math (standard and raw) and executive function measured by DCCS. Children from less disadvantaged families (below the Federal Poverty Level) showed lower gains in receptive vocabulary (standard and raw) and math (standard and raw) and executive function (PT).

In terms of program or classroom features, children in High Scope classroom showed higher gains in math (standard and raw) in comparison to children in Creative Curriculum. Larger class size was associated with executive function (DCCS). Lead teacher qualification was positively associated with receptive vocabulary gains (standard and raw). There are some positive associations between teachers who identify as Hispanic/Latino and children's literacy gains (standard and raw). ECERS-3 measure of quality was associated with higher literacy raw score and executive function (DCCS) gains. No associations are observed between the CLASS ES measure of quality and the different measures of child progress except negative association between CLASS ES and receptive vocabulary raw score. Positive associations were found between CLASS CO scores and PPVT standard and raw gains (see Appendix Table C.2 and C.3). Results are quite consistent in estimations with and without family child care providers (Tables 15 and 16).

Table 16. Multivariate analyses of children's 2018–19 standard score gains in relation to child and site or classroom characteristics and overall ECERS-3, excluding FCCs

	Rec.	Literacy	Math	Executive Function	
	Vocabulary (PPVT/TVIP)	(WJ/WM-LW)	(WJ/WM-AP)	DCCS	PT
3-year-olds	-0.982 (1.21)	-0.950 (1.14)	-0.373 (1.20)	-0.115 (0.07)	0.071 (0.65)
Asian	-2.036 (1.12)	0.505 (1.04)	-0.753 (1.09)	-0.147* (0.06)	-1.010 (0.58)
Black	-3.217** (1.04)	-0.119 (0.96)	-4.306*** (1.03)	-0.296*** (0.06)	-0.980 (0.55)
Hispanic	-0.900 (1.11)	-1.425 (1.03)	-2.077 (1.08)	-0.087 (0.06)	-0.322 (0.59)
Other	0.657 (1.00)	0.147 (0.93)	-1.788 (0.98)	-0.071 (0.06)	0.061 (0.53)
DLL	-1.420 (0.83)	1.211 (0.75)	1.448 (0.79)	-0.049 (0.05)	0.878* (0.42)
HH Income < 20k	3.741 (2.22)	1.427 (2.12)	4.507* (2.22)	0.261* (0.13)	0.748 (1.21)
HH Income 21-40k	1.044 (1.65)	-0.424 (1.58)	1.921 (1.66)	0.009 (0.10)	-0.158 (0.91)
HH Income 41-60k	1.068 (1.60)	1.809 (1.54)	1.272 (1.61)	0.026 (0.10)	0.026 (0.88)
HH Income 61-80k	0.355 (1.44)	1.528 (1.39)	1.817 (1.46)	0.113 (0.09)	0.024 (0.80)
FPL < 100	-5.871** (2.21)	-1.520 (2.12)	-5.234* (2.22)	-0.254 (0.13)	-2.042 (1.21)
FPL 100 to 300	-2.643 (1.49)	-0.762 (1.45)	-3.016* (1.51)	-0.109 (0.09)	-0.809 (0.82)
High Scope	-0.590 (0.83)	1.003 (0.79)	1.695* (0.80)	-0.003 (0.04)	-0.187 (0.39)
Class Size	-0.251 (0.13)	0.028 (0.12)	-0.108 (0.13)	0.012 (0.01)	0.051 (0.06)
Teacher Qual Exceeds	1.816 (1.08)	0.780 (1.02)	-0.052 (1.04)	-0.008 (0.05)	0.888 (0.50)
Teacher Qual Meets	1.809* (0.90)	-0.458 (0.86)	-0.335 (0.87)	-0.002 (0.05)	0.550 (0.42)
Teacher Black	1.970 (1.27)	0.805 (1.21)	-1.120 (1.23)	0.036 (0.07)	-0.381 (0.60)
Teacher Hispanic	-0.079 (1.17)	1.757 (1.13)	0.307 (1.14)	-0.043 (0.06)	0.654 (0.55)
Teacher Asian	0.718 (1.31)	-0.957 (1.25)	-0.538 (1.27)	0.020 (0.07)	0.541 (0.62)
Teacher Other	1.856 (1.56)	-0.213 (1.45)	0.791 (1.47)	0.005 (0.08)	0.919 (0.71)
ECER-3	1.193 (0.66)	1.018 (0.63)	1.179 (0.63)	0.067* (0.03)	-0.063 (0.31)
<i>N</i>	913	809	810	807	811

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are Males, White, English, FPL 300%+, Income>80 thousand, and Creative Curriculum. Other controls are pre-test, age in months, child returning to program, days between tests and an indicator for missing language, income, race, FPL, and teacher qualifications and race. Standardized scores are used for PPVT, and WJ or WM. Errors are clustered at the classroom level.

Table 17. Multivariate analyses of children's 2018–19 standard score gains in relation to child and site or classroom characteristics and CLASS dimensions, excluding FCCs

Variables	Rec. Vocabulary	Literacy	Math	Executive Function	
	(PPVT/TVIP)	(WJ/WM-LW)	(WJ/WM-AP)	DCCS	PT
3-year-olds	-0.786 (1.21)	-0.737 (1.14)	-0.056 (1.19)	-0.103 (0.07)	0.063 (0.65)
Asian	-1.976 (1.11)	0.552 (1.03)	-0.740 (1.07)	-0.149* (0.06)	-0.974 (0.58)
Black	-3.146** (1.04)	-0.122 (0.95)	-4.326*** (1.01)	-0.300*** (0.06)	-0.948 (0.55)
Hispanic	-0.832 (1.11)	-1.399 (1.02)	-2.112* (1.08)	-0.085 (0.06)	-0.311 (0.59)
Other	0.684 (1.00)	0.152 (0.93)	-1.818 (0.97)	-0.068 (0.06)	0.047 (0.53)
DLL	-1.505 (0.83)	1.255 (0.74)	1.594* (0.78)	-0.043 (0.05)	0.845* (0.42)
HH Income < 20k	3.522 (2.21)	1.590 (2.11)	4.801* (2.20)	0.261* (0.13)	0.745 (1.21)
HH Income 20-40k	0.924 (1.65)	-0.343 (1.58)	1.956 (1.65)	0.006 (0.10)	-0.154 (0.91)
HH Income 41-60k	1.097 (1.60)	1.817 (1.53)	1.218 (1.60)	0.019 (0.10)	0.038 (0.88)
HH Income 61-80k	0.243 (1.44)	1.375 (1.39)	1.579 (1.45)	0.105 (0.09)	-0.002 (0.80)
FPL < 100	-5.831** (2.21)	-1.678 (2.12)	-5.440* (2.21)	-0.241 (0.13)	-2.135 (1.21)
FPL 100 to 300	-2.594 (1.49)	-0.706 (1.45)	-2.860 (1.51)	-0.092 (0.09)	-0.869 (0.83)
High Scope	-0.254 (0.81)	1.246 (0.76)	1.970** (0.73)	-0.006 (0.04)	-0.081 (0.39)
Class Size	-0.214 (0.13)	0.027 (0.12)	-0.096 (0.12)	0.015* (0.01)	0.038 (0.06)
Teacher Qual Exceeds	1.119 (1.09)	0.085 (1.02)	-1.015 (0.97)	-0.030 (0.06)	0.803 (0.52)
Teacher Qual Meets	1.545 (0.88)	-0.445 (0.83)	-0.388 (0.80)	-0.006 (0.05)	0.580 (0.42)
Teacher Black	1.290 (1.20)	0.413 (1.14)	-1.541 (1.10)	0.008 (0.06)	-0.375 (0.59)
Teacher Hispanic	0.700 (1.17)	2.582* (1.13)	1.305 (1.07)	-0.016 (0.06)	0.762 (0.57)
Teacher Asian	0.505 (1.32)	-0.259 (1.25)	0.597 (1.20)	0.051 (0.07)	0.509 (0.64)
Teacher Other	1.499 (1.52)	-0.409 (1.39)	0.596 (1.33)	0.006 (0.08)	0.827 (0.71)
CLASS ES average	-2.422 (1.50)	0.970 (1.42)	2.266 (1.36)	0.081 (0.08)	-0.428 (0.72)
CLASS CO average	2.844* (1.28)	-0.064 (1.22)	-0.405 (1.17)	0.031 (0.07)	-0.115 (0.62)
CLASS IS average	0.298 (0.69)	1.294 (0.66)	1.587* (0.63)	0.024 (0.04)	0.265 (0.34)
<i>N</i>	913	809	810	807	811

* p<0.05; ** p<0.01; *** p<0.001. Notes for Table 14 are applicable here.

Table 18. Multivariate analyses of children's 2017–18 standard score gains in relation to child and site or classroom characteristics and CLASS dimensions, including FCCs

Variables	Rec.	Literacy	Math	Executive Function	
	Vocabulary (PPVT/TVIP)	(WJ/WM-LW)	(WJ/WM- AP)	DCCS	PT
3-year-olds	-0.945 (1.18)	-1.265 (1.13)	-0.049 (1.16)	-0.106 (0.07)	0.085 (0.63)
Asian	-1.837 (1.10)	0.518 (1.03)	-0.350 (1.06)	-0.133* (0.06)	-0.802 (0.57)
Black	-3.227** (1.01)	-0.105 (0.94)	-3.686*** (0.99)	-0.269*** (0.06)	-0.803 (0.53)
Hispanic	-1.168 (1.09)	-1.833 (1.01)	-1.847 (1.05)	-0.081 (0.06)	-0.406 (0.57)
Other	0.780 (0.99)	0.103 (0.93)	-1.623 (0.96)	-0.057 (0.06)	0.119 (0.52)
DLL	-1.546 (0.81)	1.518* (0.74)	1.528* (0.76)	-0.036 (0.05)	0.896* (0.41)
HH Income < 20k	3.460 (2.14)	1.656 (2.05)	3.887 (2.12)	0.253 (0.13)	0.847 (1.16)
HH Income 21-40k	0.750 (1.63)	-0.293 (1.58)	1.879 (1.63)	0.021 (0.10)	-0.147 (0.89)
HH Income 41-60k	1.060 (1.58)	1.703 (1.54)	1.274 (1.59)	0.025 (0.10)	0.070 (0.87)
HH Income 61-80k	0.253 (1.43)	1.422 (1.40)	1.646 (1.45)	0.111 (0.09)	0.027 (0.80)
FPL < 100	-5.710** (2.13)	-1.616 (2.06)	-4.664* (2.12)	-0.245 (0.13)	-2.280* (1.16)
FPL 100 to 300	-2.577 (1.48)	-0.629 (1.46)	-2.940* (1.50)	-0.095 (0.09)	-0.864 (0.82)
FCC	-6.565 (3.50)	0.154 (3.19)	-3.048 (3.13)	-0.027 (0.18)	1.197 (1.67)
High Scope	-0.264 (0.80)	1.190 (0.77)	1.991** (0.74)	-0.009 (0.04)	-0.110 (0.39)
Class Size	-0.230 (0.13)	-0.039 (0.12)	-0.128 (0.12)	0.014* (0.01)	0.035 (0.06)
Teacher Qual Exceeds	1.123 (1.08)	0.290 (1.03)	-0.924 (0.98)	-0.028 (0.06)	0.858 (0.52)
Teacher Qual Meets	1.522 (0.87)	-0.361 (0.84)	-0.296 (0.80)	-0.005 (0.05)	0.621 (0.42)
Teacher Black	1.034 (1.15)	0.189 (1.11)	-1.731 (1.07)	0.004 (0.06)	-0.163 (0.57)
Teacher Hispanic	0.896 (1.13)	2.441* (1.11)	1.115 (1.06)	-0.017 (0.06)	0.614 (0.56)
Teacher Asian	0.443 (1.30)	-0.563 (1.26)	0.223 (1.21)	0.042 (0.07)	0.423 (0.64)
Teacher Other	1.376 (1.51)	-0.524 (1.41)	0.465 (1.35)	0.005 (0.08)	0.824 (0.71)
CLASS ES average	-2.238 (1.46)	0.594 (1.41)	1.695 (1.35)	0.066 (0.08)	-0.543 (0.71)
CLASS CO average	2.550* (1.23)	-0.223 (1.19)	-0.231 (1.15)	0.042 (0.07)	-0.056 (0.60)
CLASS IS average	0.406 (0.67)	1.195 (0.66)	1.472* (0.63)	0.018 (0.04)	0.198 (0.33)
<i>N</i>	950	847	848	845	849

* p<0.05; ** p<0.01; *** p<0.001. Notes for Table 14 are applicable here.

Sensitivity Analyses

In addition to the main analyses reported above, we conducted three sensitivity checks to assess the robustness of findings. First, we repeated the analyses with raw scores because imperfections in the standardization could affect results. We did this for the ECERS-3 and CLASS estimations. Second, we replicated this including FCCs. Third, we replicated the analyses without FCCs but including fixed effects for agencies, which can be interpreted as understanding differences within agencies and capturing systematic characteristics common across all classrooms within each agency, when these are present.

The results of the three types of sensitivity analyses are summarized as follows.

(1) Results of analyses on raw scores for the PPVT, LW and AP measures (Tables C.1 using ECERS and C.2 using CLASS) are generally consistent with the standard score analyses.

(2) Results with FCCs and Raw scores are consistent with standard score analyses (Tables C.3).

(3) Analyses with agency fixed effects (Tables C.4 and C.5) revealed that on average a few agencies under or over performed in specific areas of development (not shown), while the majority seem to have no specific effects on children. That is, for the most part, children attending most agencies did not perform any differently than children attending other agencies. However, within agencies, ECERS scores were positively associated with the receptive vocabulary and DCCS measure. Further, CLASS Instructional Support differences showed a statistically significant positive association with letter-word identification and math changes in children.

Summary

The evaluation finds that SPP quality has generally maintained the positive trend observed in previous years, although a small decrease was observed in CLASS instructional support. This variation is not unique to this program and has been observed in other city programs. Quality improvement systems and intentional planning are key mechanisms to sustain improvements over time. SPP quality as measured by the ECERS-3 and CLASS is presently higher than what has been found in some major cities or state pre-K programs. Average quality for classrooms and family child care providers this year showed important differences. Family child care providers evidenced lower quality levels than classrooms, but maintaining relatively good levels relative to the quality observed in other established programs. No major differences were found in average quality as measured by the ECERS-3 and the CLASS instructional support by gender, some were found for dual language learners. Modest differences were observed present by race and ethnicity and by income. Having said this, all children were observed experiencing generally high levels of quality. Children in SPP made gains in all measured domains with gains in language, literacy and mathematics larger than expected based on maturation. Higher CLASS classroom instructional support scores were associated with stronger gains in math for children in the SPP program. Students of Hispanic/Latino teachers had larger gains in literacy, reinforcing the importance of teacher diversity found in previous years. The Seattle Preschool Program should continue closely monitoring quality and building on its positive trend with particular attention to instructional supports, attendance, integration of content across domains in children's activities, supports for metacognitive process that include sustained, reflective thinking processes with children.

References

- 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.
- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child development*, 78(2), 647-663.
- Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, 25(2), 166-176.
- Childcare Quality & Early Learning Center for Research & Professional Development (Unpublished). *Early Achievers Standards Validation Study*. Seattle: University of Washington
- Childcare Quality & Early Learning Center for Research & Professional Development (Unpublished). *Large Scale Psychometric Assessment of ECERS 3*. Seattle: University of Washington
- Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to “Do as I say, not as I do.” *Developmental psychobiology*, 29(4), 315-334.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.
- Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D. Cai, K., Clifford, R.M., Ebanks, C., Griffin, J.A & Henry, G. T. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child development*, 78(2), 558-580.
- Early, D. M., Sideris, J., Neitzel, J., LaForett, D. R., & Nehler, C. G. (2018). Factor structure and validity of the Early Childhood Environment Rating Scale–Third Edition (ECERS-3). *Early Childhood Research Quarterly*, 44, 242-256.
- Edvance Research (2014). *Pre-K 4 SA Evaluation Report. YEAR 1. Final Report Submitted to Early Childhood Education Municipal Development Corporation*. San Antonio, TX: Author.
- Edvance Research (2016). *Pre-K 4 SA Evaluation Report. YEAR 3. Final Report Submitted to Early Childhood Education Municipal Development Corporation*. San Antonio, TX: Author.
- Edvance Research (2018). *Pre-K 4 SA Evaluation Report. YEAR 5. Final Report Submitted to Early Childhood Education Municipal Development Corporation*. San Antonio, TX: Author.
- Harms, T., Clifford, R. M., & Cryer, D. (2014). *Early childhood environment rating scale*. Teachers College Press.
- 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.
- Jenson, D. (2015) *ECERS-3: One year out. Four States’ Experiences with Planning and Implementing Use of ECERS-3*. Presented at the QRIS National Meeting 2015. Available

- at http://www.qrisnetwork.org/sites/all/files/conference-session/resources/703ECERS3_0.pdf
- Joseph, G. E., Feldman, E., Phillips, J. J., & Jackson, E. (2010). The combined CLASS: Assessing the adult-child interactions in mixed age family childcare. A procedure manual. Designed for Washington State's QRIS, Early Achievers. Seattle, WA: Cultivate Learning.
- K1KIDS (2016) BPS K1DS: Piloting the Boston Public Schools' Prekindergarten Model in Community-Based Organizations. Boston, MA: Boston Public Schools.
- Meador, D. N., Turner, K. A., Lipsey, M. W., & Farran, D. C. (2013). Administering Measures from the PRI Learning-Related Cognitive Self- Regulation Study. Nashville, TN: Peabody Research Institute. Available at <https://my.vanderbilt.edu/cogselfregulation/files/2012/11/SR-Measure-Training-Manual-final.pdf>
- Nores, M., Barnett, W.S., Joseph, G., Stull, S., Jung, K. & Soderberg, J.S. (2017). Year 2 report: Seattle Pre-K program evaluation. New Brunswick, NJ: National Institute for Early Education Research & Seattle, WA: Cultivate Learning.
- Nores, M., Barnett, W.S., Joseph, G., Stull, S., Jung, K. & Soderberg, J.S. (2018). Year 3 report: Seattle Pre-K program evaluation. New Brunswick, NJ: National Institute for Early Education Research & Seattle, WA: Cultivate Learning.
- NIEER (2014). New Jersey Abbott Preschool Quality Evaluation Study. Summary Report. New Brunswick, NJ: National Institute for Early Education Research.
- NIEER (2016). New Jersey Abbott Preschool Quality Evaluation Study. Summary Report. New Brunswick, NJ: National Institute for Early Education Research.
- NIEER (2017). New Jersey Abbott Preschool Quality Evaluation Study. Summary Report. New Brunswick, NJ: National Institute for Early Education Research.
- NYC Department of Education (2018). Pre-K Program Assessments Classroom Assessment Scoring System (CLASS) and Early Childhood Environmental Rating Scale –Revised (ECERS-R) Release. New York: Author. Available at <https://infohub.nyced.org/docs/default-source/default-document-library/2016-17-program-assessment-results-summary-pdf-format.pdf>
- NYC Department of Education. (2018). Pre-K Program Assessments Classroom Assessment Scoring System (CLASS) and Early Childhood Environmental Rating Scale—Revised (ECERS-R) Release. New York: Author. Available at <https://infohub.nyced.org/docs/default-source/default-document-library/2017-18--program-assessment-results-summary-update.pdf>
- NYC Department of Education (2017). Pre-K Program Assessments Early Childhood Environmental Rating Scale –Revised (ECERS-R) and Classroom Assessment Scoring System (CLASS) Release. New York: Author. Available at <http://schools.nyc.gov/NR/rdonlyres/5FEA3D5B-E615-4E16-83A8-C4E58A4D6F02/0/201516ProgramAssessmentResultsSummary.pdf>
- Office of Head Start. U.S. A National Overview of Grantee CLASS® Scores in 2015. Washington, D.C.: Department of Health and Human Services. Available at <http://eclkc.ohs.acf.hhs.gov/hslc/data/class-reports/docs/national-class-2015-data.pdf>
- PAKEYS (Unpublished). What does the data tell us? The evolution of environment rating scale (ERS) use within QRIS. Available at <https://qrisnetwork.org/sites/all/files/conference-session/resources/651DataTellsUs.pdf>.

- Phillips, D. A., Gormley, W. T., & Lowenstein, A. E. (2009). Inside the pre-kindergarten door: Classroom climate and instructional time allocation in Tulsa's pre-K programs. *Early Childhood Research Quarterly*, 24(3), 213-228.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System: Manual Pre-K*. Education Review//Reseñas Educativas.
- Qi, C. H., Kaiser, A. P., Milan, S., & Hancock, T. (2006). Language performance of low-income African American and European American preschool children on the PPVT-III. *Language, Speech, and Hearing Services in Schools*, 37(1), 5-16.
- Rivers, N. M. (2016). *Seattle Public Schools and Housing Report*. Seattle: Seattle Public Schools. Available at <https://www.seattleschools.org/cms/one.aspx?portalId=627&pageId=15652>.
- Tout, K., Magnuson, K., Lipscomb, S., Karoly, L., Starr, R., Quick H., Early, D., Epstein, D., Joseph, G., Maxwell, K., Roberts, J., Swanson, C. & Wenner, J. (2017). *Validation of the Quality Ratings Used in Quality Rating and Improvement Systems (QRIS): A Synthesis of State Studies*. OPRE Report #2017-92. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Xiang, Z., & Schweinhart, L. J. (2002). *Effects five years later: The Michigan School Readiness Program evaluation through age 10*. Ypsilanti, MI: High/Scope Educational Research Foundation.
- Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly*, 28(2), 199-209.
- Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, 27(1), 122-154.
- Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. (2001). *Woodcock-Johnson III NU tests of achievement*. Rolling Meadows, IL: Riverside Publishing.
- Zelazo, P. D. (2006). The dimensional change card sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, 1, 297-301.

Appendices

Appendix A. Measures.

Appendix B. Child Scores, pre, post and gains.

Appendix C. Sensitivity Analyses.

Appendix D. P-values for tests of differences in means.

Appendix A. Measures

Measures on Children

The *Peabody Picture Vocabulary Test—Fourth Edition (PPVT-IV)*; Dunn & Dunn, 2007) is a 228-item test of receptive vocabulary in standard English, predictive of general cognitive abilities. The test is adaptive and can be used with population ages 2.5 and above. The test has proven reliability based on reported split-half reliabilities or test-retest reliabilities, as well as concurrent validity (e.g., Qi, Kaiser, Milan, & Hancock, 2006). Results on the PPVT have been found to be strongly correlated with school success (Blair & Razza, 2007; Early, et al., 2007). The test is standardized to a mean of 100 and a standard deviation of 15.

The *Woodcock-Johnson Psycho-Educational Battery—Third Edition (WJ-III)*; Woodcock, McGrew, Mather, & Schrank, 2001) includes several subtests. Two of these were used in this study: the *Applied Problems* and *Letter-Word Identification* subtests. WJ is also adaptive and for use with populations above the age of 3. The WJ has shown correlations with other tests of cognitive ability and achievement ranging between 0.60 and 0.70. This measure has been used in numerous large-scale preschool studies (e.g., Early, et al., 2007; Wong, Cook, Barnett, & Jung, 2008). The test is standardized to a mean of 100 and a standard deviation of 15.

The *Dimensional Change Card Sort Task (DCCS)*; Zelazo, 2006) engages children in reverse categorization by sorting a set of cards based on different criteria provided by the examiner. The test assesses attention-shifting, as well as short term memory. Scores on the DCCS reflect a pass/fail system on three levels of increasing difficulty, and raw scores range between 0 and 3 based on these levels. 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 were 1.33, for 51–53 months were 1.42, and for 54–56 months were 1.58 (Meador et al., 2013).

The *Peg Tapping Test (PT)*; Diamond & Taylor, 1996) asks children to tap a peg twice when the experimenter taps once and vice versa. Sixteen trials are conducted with 8 one-tap and 8 two-tap trials in random sequence. The task requires two abilities: (a) the ability to hold two things in mind—the rule to tap once when experimenter taps twice and the rule to tap twice when experimenter taps once, and (b) the ability to exercise inhibitory control over one’s proponent behavior, the natural tendency to mimic what the experimenter does. The final score for Peg Tapping is a sum of all the 16 items that comprise the test. Again, 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 were 4.57, for 51–53 months were 6.02, and for 54–56 months were 7.87 (Meador et al., 2013).

Measures on Classrooms

Early Childhood Environment Rating Scale—Third Ed. (ECERS-3); Harms, Clifford & Cryer, 2014). The ECERS-3 is an observation and rating tool for preschool and kindergarten classrooms measuring environmental factors and teacher-child interactions. It emphasizes the role of the teacher in relation to the classroom environment and children’s developmental gains. The overall ECERS-3 score is an average on 35 items under 6 domains, which are each rated in a scale between 1 and 7. A rating of 1 indicates inadequate quality, a rating of 3 indicates minimal quality, a rating of 5 indicates good quality, and a rating of 7 indicates excellent quality. A

general description of each of the 35 items on the ECERS-3 is provided in Table A.1. A recent validation paper (Early, et. al, 2018) reports a four-factor (Learning Opportunities, Gross Motor, Teacher Interactions, and Math Activities) structure to the ECERS-3, found moderate correlations with the three CLASS Pre-K domains, and positive associations with growth in children’s executive functions (while not with children’s cognitive measures). The ECERS-3 was only used in classrooms in center-based care.

Table A.1. ECERS-3 Subscale and Item Descriptions.

Subscale	Items	Description
Space for Furnishings	1. Indoor Space	Considers enough indoor space for children, staff, and basic furnishings for routines, play, and learning.
	2. Furnishings for care, play, and learning	Focuses on ample furniture for routine care, play, and learning, including convenient cubbies for individual use.
	3. Room arrangement for play and learning	Space is arranged so that classroom pathways generally do not interrupt play and supervision.
	4. Space for privacy	Considers an indoor space for privacy available and set up physically in the classroom to discourage interruptions.
	5. Child-related display	Focuses on appropriate materials displayed for children throughout the classroom, including simple pictures, posters, and artwork.
	6. Space for gross motor play	Gross motor area is spacious, generally safe, and easily accessible to children.
	7. Gross motor equipment	Equipment is age appropriate, accessible, and ample enough to interest every child.
Personal Care Routines	Meals/Snacks	Schedule and sanitary procedures are appropriate during meal times. Staff sit with children to encourage learning.
	Toileting/diapering	Proper sanitary procedures usually followed with pleasant supervision.
	Health practices	Proper sanitary procedures used consistently as needed, with a few lapses.
	Safety practices	Considers no more than 2 major safety hazards present indoors or outdoors.
Language and Literacy	Helping children expand vocabulary	Measures how frequent staff uses specific words for objects and actions and descriptive words as children experience routines and play.
	Encouraging children to use language	Assesses how frequent staff asks questions that children are interested in answering and that require longer answers. Includes many conversations during gross motor free play and routines.
	Staff use of books with children	Staff read appropriate books to children that relate to current classroom activities or themes, showing interest and enjoyment while doing so.
	Encouraging children’s use of books	Many books are accessible and organized in a defined interest center.
	Becoming familiar with print	Focuses on how most visible print is combined with pictures, relates to current classroom topics, and shows a variety of words.
Learning Activities	Fine motor	Focuses on the accessibility for children of fine motor materials, including interlocking building materials, manipulatives, puzzles, and art materials.
	Art	Art materials, including drawing materials, paints, 3D objects, collage materials, and tools, must be accessible for children.
	Music and movement	Measures how many music materials and activities are accessible for children during free play.
	Blocks	Enough space, unit blocks and accessories from 3 different categories for 2-3 children to build at once.

	Dramatic play	Many and varied dramatic play materials, including dolls, furniture, play food and dress-up clothes must be accessible for children during free play.
	Nature/science	At least 15 nature/science materials, including living things, natural objects, factual books, tools, or sand/water must be accessible for children.
	Math materials and activities	At least 10 different appropriate math materials accessible, including materials to count/compare quantities, measure/compare sizes, and familiarize children with shapes.
	Math in daily events	Assess how staff encourages math learning as part of daily routines.
	Understanding written numbers	At least 3-5 different materials should be present in the classroom that shows children the meaning of print numbers.
	Promoting acceptance of diversity	At least 10 examples of diversity accessible, including books, displayed pictures and materials.
	Appropriate use of technology	All observed materials used are appropriate and limited to 10-15 minutes per child during the observation.
Interaction	Supervision of gross motor	Focuses on careful supervision in order to ensure children's safety.
	Individualized teaching and learning	Many activities observed are open-ended and most allow children to be successful.
	Staff-child interaction	Evaluates frequent positive staff-child interactions, with no long periods of no interaction.
	Peer interaction	Captures positive peer interactions during at least half of the observation.
	Discipline	Children appear to be aware of classroom rules, and generally follow them with reasonable amount of teacher control.
Program Structure	Transitions and waiting times	Classroom transitions are usually smooth and productively engaging.
	Free play	Free play takes place for 1 hour during observation, including some time indoors and some time outdoors (weather permitting).
	Whole - group activities for play and learning	Staff are responsive and flexible in ways that maximize child engagement during whole group activities.

Classroom Assessment Scoring System Pre-K (CLASS Pre-K; Pianta, La Paro, & Hamre, 2008). The CLASS Pre-K is an observational tool that identifies the classroom interactions that promote children's development and learning. Observations consist of four 20-minute cycles, with 10-minute coding periods between each cycle, which are then averaged for an overall quality score. Interactions are measured through 10 dimensions in 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 3 dimensions: Productivity, Behavior Management, and Instructional Learning Formats. The Instructional Support domain is measured by three dimensions: Concept Development, Quality of Feedback, and Language Modeling. Each scale uses a 7-point Likert-type scale, for which a score of 1 or 2 indicates low quality, and a score of 6 or 7 indicates high quality. The CLASS domains and dimensions are outlined in Table A.2.

Because a CLASS instrument does not exist for mixed aged groupings, Family child care providers were observed with three CLASS instruments using a *Combined CLASS Protocol* (Joseph, Feldman, Phillips & Jackson, 2010),²⁹ which was designed to be used in any child care

²⁹ Protocol designed for Washington State's QRIS, Early Achievers. Also used in Oregon, see Tout, et. al, (2017).

facility in a home, with multiple age groups. This protocol integrates the dimensions from Infant, Toddler, and Pre-K CLASS. There are three dimensions that apply only to pre-K children: Productivity, Instructional Learning Formats, and Concept Development. All other dimensions apply to children of different age groups, depending on which children are present.³⁰ In addition, the combined protocol includes a new dimension, that of Facilitation of Learning and Development, from the CLASS protocols for children under 3. Observers using the combined protocol are trained and reliable in all three CLASS instruments, and the items on the combined protocol draw from the corresponding items in the Infant CLASS Manual, the Toddler CLASS Manual, and the Pre-K CLASS Manual, which are used by observers throughout the process. The protocol requires paying attention to children of all ages. Therefore, if differentiation by age does not adequately occur (e.g. adequate language modelling is observed for infants and toddlers but not for preschoolers), scores will reflect the average for the whole age-group served, rather than only preschool children.³¹ Teachstone recommendations allow for either using the protocol associated with the age range of the majority of the children, or rotating between the necessary protocols, depending on the needs of the study. The combined protocol is based on the second recommendation but applies all protocols across all cycles, using the necessary protocols in accordance with the age of the children. Further information for the combined protocol is provided in Table A.3.

Table A.2. CLASS Domains and Dimension Descriptions.

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

³⁰ If a given age group is not present or sleeping during the observation, the particular age group will not be considered when scoring.

³¹ Although there may be benefits of mixed age-grouping that the CLASS is not designed to capture.

	Language Modeling	Captures the effectiveness and amount of teacher’s use of language-stimulation and language-facilitation techniques.
--	-------------------	--

Table A.3. Considerations on The Combined CLASS protocol

The protocol for using Combined CLASS manuals (Joseph, Feldman, Phillips & Jackson, 2010) integrates dimensions from all three CLASS tools (Infant, Toddler, and Pre-K) to allow for multi-age groupings, most often present in family child care homes. Each of the individual CLASS protocols contain differing numbers of dimensions i.e., Infant has 4, Toddler has 8, and Pre-K has 10. Therefore, some dimensions in the Combined CLASS process apply only to certain age groups. For example, three dimensions apply only to preschool children, four dimensions apply only to toddlers and preschoolers, and the remaining four dimensions apply to all age groups. When coding dimensions that span all age groups, consideration is given to how many children are present within an age group and the relative breadth/depth of interactions that impact each. For example, imagine half the attendees are infants and half are preschoolers. In such a scenario, if the caregiver provides appropriate language stimulation to infants but only provides low level language modeling for Pre-K children, the score on this dimension may fall in the mid-range when using the Combined Class process even though it may fall in the low range for Pre-K CLASS children. Only children present are counted and infants sleeping are not considered “present”. Please note this process is a hybrid model designed for Washington State’s QRIS and utilized in this study. For information about other Family Child Care CLASS models, please see “Using the CLASS Measure in Family Child Care Homes” (Vitiello, 2014) via Teachstone.com

Table A.4. ECERS and CLASS Dimension and Domain Means by Child Demographics, 2019

		ECERS			CLASS ES			CLASS CO			CLASS IS		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Gender	Female	551	4.21	0.63	575	6.57	0.36	575	6.20	0.54	575	3.18	0.84
	Male	575	4.27	0.60	598	6.61	0.36	598	6.24	0.54	598	3.22	0.82
Ethnicity	White	249	4.21	0.63	253	6.67	0.34	253	6.30	0.53	253	3.20	0.85
	Black	284	4.24	0.60	309	6.56	0.38	309	6.21	0.56	309	3.30	0.86
	Asian	207	4.29	0.57	210	6.60	0.35	210	6.23	0.51	210	3.11	0.83
	Hispanic	159	4.35	0.60	164	6.58	0.33	164	6.22	0.51	164	3.22	0.81
	Other	188	4.13	0.70	194	6.53	0.39	194	6.12	0.57	194	3.10	0.76
Language	English	743	4.21	0.64	777	6.61	0.36	777	6.23	0.54	777	3.19	0.81
	DLL	343	4.32	0.55	349	6.57	0.36	349	6.23	0.54	349	3.23	0.90
	Unknown	43	4.13	0.74	50	6.45	0.33	50	5.96	0.50	50	2.98	0.58
FPL	<100	381	4.28	0.60	401	6.51	0.40	401	6.18	0.57	401	3.31	0.87
	100-300	474	4.24	0.64	497	6.61	0.34	497	6.20	0.53	497	3.13	0.79
	>300	241	4.18	0.61	242	6.70	0.29	242	6.34	0.50	242	3.16	0.84
	Unknown	33	4.24	0.48	36	6.54	0.35	36	6.09	0.65	36	3.06	0.73

Appendix B. Child Scores, pre, post and gains.

Receptive vocabulary results

Table B.1 reports children’s receptive vocabulary scores for the fall (pre-test) and spring (post-test), as well as fall to spring gains. This section reports standard scores—which are adjusted for age (raw scores are reported in section further below). The mean standard score for the PPVT is set at 100 which represents the average child in the U.S. population at any age. The standard

deviation is at 15. Positive gains are an indication that children improved more over the course of the preschool year than is expected based on the change in age (maturation) alone.

Table B.1. Receptive vocabulary means and gains by child characteristics

		PPVT 2018 Fall			PPVT 2019 Spring			PPVT Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		1031	95.69	19.31	1104	98.78	18.22	951	3.38	10.52
Gender	Male	519	95.15	19.47	553	98.46	19.06	470	3.39	10.74
	Female	512	96.19	19.14	548	99.07	17.35	480	3.34	10.31
Age	3-Year-Old Cohort	291	91.44	17.24	356	94.03	18.37	262	2.98	10.87
	4-Year-Old Cohort	735	97.38	19.84	743	101.06	17.76	685	3.52	10.39
Ethnicity	White	225	107.69	19.23	245	109.89	17.59	213	2.20	10.69
	Black	271	87.82	15.41	285	90.65	14.91	249	3.08	10.07
	Asian	187	89.43	17.64	201	92.84	16.65	178	4.22	10.64
	Hispanic	142	94.08	19.57	162	96.51	17.38	137	3.53	10.10
	Other	168	101.32	17.17	182	104.64	16.32	153	4.29	11.18
Language	English	687	101.46	17.77	736	104.06	17.26	638	2.69	10.39
	DLL	305	83.44	16.50	323	87.44	14.98	279	4.90	10.57
	Unknown	39	89.77	19.62	45	93.91	16.53	34	3.76	11.48
FPL	<100	339	90.19	17.42	378	92.80	16.62	312	3.41	10.81
	100-300	438	93.94	18.33	475	97.35	17.03	414	3.53	10.11
	>300	222	108.21	18.53	231	111.59	16.62	209	2.96	10.91

Children's fall and spring receptive vocabulary standard scores for selected center characteristics are reported in Table B.2. (raw scores are reported further below). Few children are left with levels of CLASS ES under the threshold of 5.5 and are therefore not reported.

Table B.2. Receptive vocabulary means and gains by center characteristics

		PPVT 2018 Fall			PPVT 2019 Spring			PPVT Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		1031	95.69	19.31	1104	98.78	18.22	951	3.38	10.52
Curriculum	High Scope	609	96.35	19.15	660	99.15	18.18	565	3.26	10.63
	Creative Curriculum	422	94.74	19.52	444	98.24	18.30	386	3.54	10.37
ECERS	Less than 3	25	101.92	17.12	26	103.85	12.18	23	2.65	11.55
	3 or More	963	95.71	19.45	1024	99.00	18.36	891	3.49	10.49
CLASS ES	Less than 5.5	2	-	-	2	-	-	2	-	-
	5.5 or More	1029	95.71	19.32	1089	98.89	18.16	949	3.38	10.52
CLASS CO	Less than 5.5	106	93.58	17.47	112	93.66	17.37	93	1.38	10.53
	5.5 or More	925	95.93	19.50	979	99.47	18.15	858	3.59	10.50
CLASS IS	Less than 3	481	94.68	18.53	526	96.87	18.11	449	2.83	10.38
	3 or More	550	96.57	19.94	565	100.73	18.01	502	3.87	10.62

Literacy results

Similarly, Table B.3 reports children's WJ-III letter-word (LW) identification scores for the overall sample and by selected child characteristics. Like the PPVT, the LW subtest measures the ability of children to identify letters and subsequently read a list of words of increasing difficulty. The test has a mean standard (i.e., age adjusted score) of 100 and a standard deviation of 15 (raw scores are reported further below).

Table B.3. Literacy means and gains by child characteristics

		WJ-LW 2018 Fall			WJ-LW 2019 Spring			WJ-LW Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		918	101.52	14.55	842	102.19	14.44	840	0.33	8.80
Gender	Male	456	102.30	15.92	411	102.72	15.99	410	0.05	9.09
	Female	462	100.75	13.04	430	101.67	12.79	429	0.58	8.52
Age	3-Year-Old Cohort	259	100.50	14.62	231	102.61	12.70	229	1.96	10.06
	4-Year-Old Cohort	656	101.93	14.54	609	102.05	15.07	609	-0.28	8.22
Ethnicity	White	202	101.47	12.78	191	102.14	12.67	191	0.31	7.69
	Black	245	102.29	16.51	222	102.79	16.54	222	0.08	9.34
	Asian	165	105.29	14.71	157	105.92	13.83	156	0.77	8.12
	Hispanic	123	96.31	12.34	120	96.21	12.84	119	-0.03	10.09
	Other	150	101.37	13.95	136	102.09	14.18	136	0.43	8.72
Language	English	626	101.43	14.11	576	101.63	14.10	575	-0.13	8.83
	DLL	256	101.63	14.77	235	103.22	14.44	234	1.39	8.54
	Unknown	36	102.39	20.00	31	104.87	19.68	31	0.74	9.61
FPL	<100	311	99.91	14.83	284	100.77	15.02	282	0.35	9.22
	100-300	383	101.47	14.22	361	102.06	13.60	361	0.62	9.08
	>300	197	104.96	14.35	186	104.37	15.21	186	-0.52	7.18

Children's pre- and post-test letter-word identification standard scores across selected center characteristics are reported in Table B.4. (raw scores are reported further below).

Table B.4. Literacy means and gains by center characteristics

		WJ-LW 2018 Fall			WJ-LW 2019 Spring			WJ-LW Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		918	101.52	14.55	842	102.19	14.44	840	0.33	8.80
Curriculum	High Scope	538	101.88	14.92	497	102.71	14.40	497	0.66	8.56
	Creative Curriculum	380	101.01	14.01	345	101.45	14.48	343	-0.16	9.13
ECERS	Less than 3	23	102.35	20.71	21	97.95	9.88	21	-1.67	7.44
	3 or More	851	101.35	14.25	783	102.33	14.38	781	0.51	8.77
CLASS ES	Less than 5.5	2	-	-	2	-	-	2	-	-
	5.5 or More	916	101.45	14.44	840	102.13	14.29	838	0.33	8.80
CLASS CO	Less than 5.5	95	100.29	16.19	81	101.89	16.53	81	0.07	9.84
	5.5 or More	823	101.66	14.35	761	102.23	14.21	759	0.35	8.69
CLASS IS	Less than 3	432	101.25	15.36	401	101.62	14.76	400	0.01	8.66
	3 or More	486	101.76	13.80	441	102.72	14.13	440	0.61	8.93

Early math results

The WJ applied problems (AP) subscale of the WJ-III, which measures math ability, are shown in Table B.5. The AP subscale is normed with a mean of 100 and a standard deviation of 15.

Table B.5. Math means and gains by child characteristics

		WJ-AP 2018 Fall			WJ-AP 2019 Spring			WJ-AP Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		918	100.77	14.27	841	103.20	13.02	839	1.91	10.05
Gender	Male	456	100.13	14.72	410	102.95	14.01	409	2.10	10.12
	Female	462	101.36	13.80	430	103.39	12.00	429	1.72	10.00
Age	3-Year-Old Cohort	259	96.80	13.92	231	101.94	12.67	229	4.53	10.74
	4-Year-Old Cohort	656	102.31	14.11	608	103.67	13.15	608	0.96	9.60
Ethnicity	White	202	108.05	14.54	190	109.41	12.93	190	0.71	9.67
	Black	245	94.79	12.62	222	97.08	11.04	222	1.85	10.06
	Asian	165	99.00	13.78	157	103.68	12.03	156	4.58	9.87
	Hispanic	123	100.32	12.36	120	102.20	12.02	119	1.97	8.75
	Other	150	103.61	13.55	136	104.89	13.04	136	0.51	10.96
Language	English	626	103.02	14.08	575	104.42	13.32	574	0.80	9.45
	DLL	256	95.87	13.61	235	100.52	11.92	234	4.49	11.09
	Unknown	36	96.44	12.56	31	100.87	12.54	31	2.94	9.24
FPL	<100	311	97.44	13.87	284	100.38	12.68	282	2.44	10.73
	100-300	383	99.97	13.28	360	102.27	12.49	360	2.02	9.85
	>300	197	108.20	14.03	186	109.16	12.16	186	0.56	9.08

Table B.6. reports children's fall and spring standardized math scores and gains by selected center characteristics (raw scores are reported further below).

Table B.6. Math means and gains by center characteristics

		WJ-AP 2018 Fall			WJ-AP 2019 Spring			WJ-AP Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		918	100.77	14.27	841	103.20	13.02	839	1.91	10.05
Curriculum	High Scope	538	100.45	14.27	496	103.50	12.94	496	2.71	10.13
	Creative Curriculum	380	101.21	14.28	345	102.77	13.14	343	0.75	9.84
ECERS	Less than 3	23	100.43	12.16	21	100.62	10.22	21	-0.52	7.33
	3 or More	851	100.89	14.47	782	103.53	13.13	780	2.06	10.21
CLASS ES	Less than 5.5	2	-	-	2	-	-	2	-	-
	5.5 or More	916	100.79	14.28	839	103.20	13.04	837	1.88	10.04
CLASS CO	Less than 5.5	95	96.69	13.39	81	100.59	13.69	81	2.75	11.01
	5.5 or More	823	101.24	14.30	760	103.48	12.93	758	1.82	9.95
CLASS IS	Less than 3	432	99.44	13.89	401	101.92	12.62	400	1.94	10.15
	3 or More	486	101.94	14.51	440	104.37	13.28	439	1.88	9.97

Executive functions

Two measures of executive functions were included. The DCCS is an attention shifting test which taps into a child's short-term memory, attention and inhibitory control. Table B.7. reports

children’s fall and spring DCCS scores for selected child characteristics. As reference, the Learning-Related Cognitive Self-Regulation School Readiness Measures for Preschool Children Study (a.k.a. the Self-Regulation Measurement Study) (Meador, et. al, 2013) tested alternative measures of executive functions including the DCCS. The authors found average DCCS scores of 1.42 at 51–53 months and 1.62 at 57–59 months (an average difference of 0.20 between these two ages); ranges which include the average ages at fall and spring testing in this study (53.2 months in the fall and 59.3 in the spring). Table B.8. reports children’s spring and fall DCCS scores by selected center characteristics.

Table B.7. DCCS means and gains by child characteristics

		DCCS 2018 Fall			DCCS 2019 Spring			DCCS Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		917	1.39	0.61	839	1.64	0.63	836	0.24	0.60
Gender	Male	455	1.35	0.62	408	1.60	0.63	406	0.22	0.61
	Female	462	1.42	0.60	430	1.68	0.62	429	0.25	0.59
Age	3-Year-Old Cohort	258	1.05	0.53	231	1.34	0.53	228	0.29	0.60
	4-Year-Old Cohort	656	1.52	0.59	606	1.75	0.62	606	0.22	0.60
Ethnicity	White	202	1.68	0.64	189	1.93	0.61	189	0.22	0.65
	Black	244	1.13	0.54	222	1.34	0.56	221	0.21	0.60
	Asian	165	1.38	0.53	156	1.61	0.65	155	0.23	0.57
	Hispanic	123	1.37	0.62	120	1.66	0.59	119	0.31	0.61
	Other	150	1.45	0.57	136	1.74	0.56	136	0.26	0.57
Language	English	625	1.46	0.63	574	1.71	0.64	572	0.24	0.61
	DLL	256	1.24	0.55	234	1.50	0.58	233	0.26	0.59
	Unknown	36	1.22	0.54	31	1.39	0.56	31	0.13	0.43
FPL	<100	310	1.24	0.55	284	1.52	0.59	281	0.28	0.56
	100-300	383	1.34	0.61	359	1.58	0.64	359	0.23	0.63
	>300	197	1.72	0.59	185	1.94	0.59	185	0.19	0.60

Table B.8. DCCS means and gains by center characteristics

		DCCS 2018 Fall			DCCS 2019 Spring			DCCS Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		917	1.39	0.61	839	1.64	0.63	836	0.24	0.60
Curriculum	High Scope	537	1.44	0.61	496	1.67	0.64	495	0.22	0.61
	Creative Curriculum	380	1.31	0.61	343	1.60	0.61	341	0.27	0.58
ECERS	Less than 3	23	1.52	0.51	21	1.62	0.67	21	0.10	0.70
	3 or More	850	1.40	0.61	780	1.66	0.62	777	0.24	0.60
CLASS ES	Less than 5.5	2	-	-	2	-	-	2	-	-
	5.5 or More	915	1.39	0.61	837	1.64	0.63	834	0.24	0.60
CLASS CO	Less than 5.5	95	1.29	0.60	81	1.48	0.55	81	0.16	0.58
	5.5 or More	822	1.40	0.61	758	1.66	0.63	755	0.25	0.60
CLASS IS	Less than 3	431	1.36	0.61	399	1.60	0.62	397	0.23	0.57
	3 or More	486	1.41	0.61	440	1.68	0.63	439	0.25	0.63

Children were assessed also with the Peg Tapping (PT) measure. PT is a measure of inhibitory control and attention. Table B.9. reports children’s fall and spring scores for Peg Tapping for selected subgroups of children by child characteristics. This measure has not been

normed. The Self-Regulation Measurement Study (Meador, et. al, 2013) also included this measure. In such study average scores were of 6.02 at 51–53 months and 8.80 at 57–59 months, with a difference of 2.78. Table B.10. reports pre- and post-test Peg-Tapping scores for children in the sample across selected center characteristics.

Table B.9. Peg Tapping means and gains by child characteristics

		PT 2018 Fall			PT 2019 Spring			PT Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		917	5.86	5.86	842	8.30	5.87	840	2.34	5.34
Gender	Male	455	5.56	5.77	411	7.90	5.98	410	2.25	5.36
	Female	462	6.15	5.95	430	8.69	5.74	429	2.43	5.32
Age	3-Year-Old Cohort	258	2.05	4.41	231	5.05	5.32	229	2.86	5.34
	4-Year-Old Cohort	656	7.34	5.69	609	9.52	5.59	609	2.15	5.33
Ethnicity	White	202	7.93	5.92	191	10.18	5.25	191	1.94	5.66
	Black	245	3.73	5.43	222	6.30	5.77	222	2.56	4.96
	Asian	165	5.88	5.84	157	8.06	6.03	156	2.31	5.35
	Hispanic	122	5.85	5.30	120	8.31	5.90	119	2.48	5.62
Language	Other	150	6.55	5.99	136	9.20	5.71	136	2.63	5.18
	English	625	6.49	5.95	576	8.61	5.85	575	2.00	5.25
	DLL	256	4.60	5.44	235	7.73	5.90	234	3.16	5.46
FPL	Unknown	36	3.83	5.41	31	6.97	5.65	31	2.55	5.37
	<100	310	4.37	5.45	284	6.75	5.97	282	2.29	5.56
	100-300	383	5.45	5.78	361	8.11	5.68	361	2.60	5.06
	>300	197	9.12	5.54	186	11.12	5.07	186	2.01	5.45

Table B.10. Peg-Tapping means and gains by center characteristics

		PT 2018 Fall			PT 2019 Spring			PT Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		917	5.86	5.86	842	8.30	5.87	840	2.34	5.34
Curriculum	High Scope	537	6.22	6.05	497	8.61	5.85	497	2.32	5.42
	Creative Curriculum	380	5.36	5.56	345	7.86	5.87	343	2.37	5.22
ECERS	Less than 3	23	7.09	5.65	21	10.24	5.38	21	3.10	5.42
	3 or More	850	5.92	5.88	783	8.35	5.88	781	2.31	5.38
CLASS ES	Less than 5.5	2	-	-	2	-	-	2	-	-
	5.5 or More	915	5.87	5.86	840	8.31	5.87	838	2.34	5.34
CLASS CO	Less than 5.5	95	5.63	5.85	81	7.94	5.75	81	2.25	5.38
	5.5 or More	822	5.89	5.87	761	8.34	5.88	759	2.35	5.33
CLASS IS	Less than 3	432	5.95	6.07	401	8.31	5.97	400	2.16	5.22
	3 or More	485	5.78	5.68	441	8.29	5.78	440	2.51	5.44

Raw Scores**Table B.11. Receptive vocabulary raw score means and gains by child characteristics**

		PPVT 2018 Fall			PPVT 2019 Spring			PPVT Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		1031	65.71	27.29	1104	77.60	27.15	951	13.03	14.03
Gender	Male	519	65.38	28.07	553	77.50	28.52	470	12.99	14.18
	Female	512	66.20	26.46	548	77.67	25.74	480	12.83	13.11
Age	3-Year-Old Cohort	291	48.35	21.37	356	60.84	23.85	262	13.27	13.35
	4-Year-Old Cohort	735	72.64	26.27	743	85.52	24.95	685	12.78	13.77
Ethnicity	White	225	83.32	26.92	245	94.13	25.74	213	11.51	13.97
	Black	271	53.63	22.38	285	65.60	23.16	249	12.53	12.65
	Asian	187	58.18	24.11	201	69.32	24.93	178	13.38	13.83
	Hispanic	142	61.01	27.04	162	74.48	25.93	137	12.99	13.42
	Other	168	73.24	24.71	182	85.75	24.74	153	14.85	14.74
Language	English	687	73.66	25.89	736	85.36	25.77	638	12.45	13.78
	DLL	305	49.58	22.48	323	61.58	22.73	279	14.06	13.33
	Unknown	39	54.31	25.44	45	65.51	23.71	34	12.50	13.48
FPL	<100	339	56.88	24.24	378	67.85	24.93	312	12.84	13.85
	100-300	438	62.90	25.26	475	75.07	24.96	414	13.07	13.03
	>300	222	86.34	25.40	231	99.10	23.09	209	12.68	14.63

Table B.12. Receptive vocabulary raw score means and gains by center characteristics

		PPVT 2018 Fall			PPVT 2019 Spring			PPVT Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		1031	65.71	27.29	1104	77.60	27.15	951	13.03	14.03
Curriculum	High Scope	609	67.36	27.36	660	78.58	27.33	565	12.73	14.01
	Creative Curriculum	422	63.56	27.00	444	76.13	26.85	386	13.22	13.11
ECERS	Less than 3	25	79.12	22.72	26	90.38	17.04	23	12.26	15.87
	3 or More	963	65.76	27.45	1024	77.83	27.36	891	13.09	13.61
CLASS ES	Less than 5.5	2	46.50	14.85	2	49.00	15.56	2	2.50	0.71
	5.5 or More	1029	65.84	27.28	1089	77.80	27.09	949	12.95	13.65
CLASS CO	Less than 5.5	106	62.71	25.69	112	70.78	26.04	93	10.10	14.03
	5.5 or More	925	66.16	27.43	979	78.55	27.11	858	13.23	13.58
CLASS IS	Less than 3	481	64.62	26.90	526	75.25	26.94	449	12.06	13.53
	3 or More	550	66.84	27.57	565	80.08	27.05	502	13.71	13.72

Table B.13. Literacy raw score means and gains by child characteristics

		WJ-LW 2018 Fall			WJ-LW 2019 Spring			WJ-LW Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		918	8.20	5.84	842	10.74	6.54	840	2.43	3.20
Gender	Male	456	8.66	6.80	411	11.22	7.50	410	2.40	3.32
	Female	462	7.74	4.67	430	10.29	5.45	429	2.45	3.09
Age	3-Year-Old Cohort	259	5.61	4.73	231	7.68	4.56	229	2.10	2.77
	4-Year-Old Cohort	656	9.21	5.93	609	11.89	6.81	609	2.55	3.35
Ethnicity	White	202	8.36	5.02	191	11.23	5.89	191	2.68	2.97
	Black	245	8.19	7.21	222	10.62	7.72	222	2.31	3.51
	Asian	165	9.73	5.77	157	12.29	5.96	156	2.65	2.97
	Hispanic	123	6.46	4.37	120	8.37	5.20	119	1.97	3.09
	Other	150	8.11	5.35	136	10.57	6.77	136	2.38	3.31
Language	English	626	8.21	5.64	576	10.67	6.39	575	2.33	3.21
	DLL	256	8.14	5.79	235	10.91	6.41	234	2.74	3.12
	Unknown	36	8.33	9.00	31	10.84	9.85	31	1.84	3.46
FPL	<100	311	7.36	5.83	284	9.74	6.56	282	2.20	3.09
	100-300	383	8.06	5.58	361	10.51	5.96	361	2.47	3.20
	>300	197	10.10	6.07	186	12.68	7.30	186	2.62	3.32

Table B.14. Literacy raw score means and gains by center characteristics

		WJ-LW 2018 Fall			WJ-LW 2019 Spring			WJ-LW Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		918	8.20	5.84	842	10.74	6.54	840	2.43	3.20
Curriculum	High Scope	538	8.48	6.02	497	11.14	6.68	497	2.61	3.20
	Creative Curriculum	380	7.80	5.56	345	10.17	6.30	343	2.17	3.18
ECERS	Less than 3	23	9.57	8.87	21	10.05	3.40	21	1.81	2.46
	3 or More	851	8.11	5.61	783	10.78	6.53	781	2.50	3.23
CLASS ES	Less than 5.5	2	-	-	2	-	-	2	-	-
	5.5 or More	916	8.17	5.77	840	10.72	6.49	838	2.44	3.20
CLASS CO	Less than 5.5	95	8.02	6.67	81	10.64	8.11	81	2.14	3.94
	5.5 or More	823	8.22	5.74	761	10.75	6.36	759	2.46	3.11
CLASS IS	Less than 3	432	8.19	6.16	401	10.57	6.69	400	2.25	3.08
	3 or More	486	8.20	5.54	441	10.90	6.41	440	2.59	3.31

Table B.15. Math raw score means and gains by child characteristics

		WJ-AP 2018 Fall			WJ-AP 2019 Spring			WJ-AP Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		918	10.13	5.32	841	13.24	5.02	839	2.89	3.29
Gender	Male	456	9.99	5.60	410	13.21	5.51	409	2.91	3.33
	Female	462	10.26	5.04	430	13.25	4.51	429	2.87	3.26
Age	3-Year-Old Cohort	259	6.22	4.16	231	9.74	4.32	229	3.28	3.13
	4-Year-Old Cohort	656	11.65	4.92	608	14.55	4.62	608	2.76	3.34
Ethnicity	White	202	13.05	5.12	190	16.02	4.38	190	2.64	3.13
	Black	245	7.60	4.80	222	10.53	4.58	222	2.77	3.25
	Asian	165	9.66	5.04	157	13.39	4.66	156	3.75	3.23
	Hispanic	123	10.09	4.67	120	12.97	4.60	119	2.92	2.91
	Other	150	11.12	5.19	136	13.84	5.14	136	2.42	3.68
Language	English	626	11.00	5.25	575	13.83	5.07	574	2.59	3.13
	DLL	256	8.41	5.04	235	12.07	4.67	234	3.62	3.57
	Unknown	36	7.31	4.65	31	10.97	4.64	31	3.06	3.12
FPL	<100	311	8.61	5.04	284	11.75	4.91	282	2.93	3.47
	100-300	383	9.76	5.05	360	12.83	4.81	360	2.94	3.23
	>300	197	13.51	4.85	186	16.27	4.30	186	2.62	3.06

Table B.16. Math raw score means and gains by center characteristics

		WJ-AP 2018 Fall			WJ-AP 2019 Spring			WJ-AP Gains 2018-2019		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Total		918	10.13	5.32	841	13.24	5.02	839	2.89	3.29
Curriculum	High Scope	538	10.23	5.35	496	13.49	5.05	496	3.12	3.30
	Creative Curriculum	380	9.99	5.29	345	12.87	4.96	343	2.57	3.26
ECERS	Less than 3	23	10.83	4.63	21	13.76	3.55	21	2.52	2.69
	3 or More	851	10.17	5.38	782	13.34	5.08	780	2.93	3.35
CLASS ES	Less than 5.5	2	-	-	2	-	-	2	-	-
	5.5 or More	916	10.14	5.32	839	13.25	5.03	837	2.89	3.29
CLASS CO	Less than 5.5	95	8.72	5.07	81	12.07	5.27	81	2.96	3.68
	5.5 or More	823	10.29	5.33	760	13.36	4.98	758	2.89	3.25
CLASS IS	Less than 3	432	9.71	5.29	401	12.85	4.97	400	2.90	3.43
	3 or More	486	10.51	5.33	440	13.59	5.05	439	2.89	3.16

Appendix C. Sensitivity Analyses.

Table C.1. Multivariate analyses of children’s 2018–19 raw score gains in relation to child and site or classroom characteristics and ECERS-3, excluding FCCs

Variables	Rec. Vocabulary (PPVT/TVIP)	Literacy (WJ/WM-LW)	Math (WJ/WM-AP)
3-year-olds	-1.031 (1.68)	-0.250 (0.44)	-0.118 (0.41)
Asian	-2.148 (1.55)	-0.052 (0.40)	-0.237 (0.38)
Black	-3.416* (1.45)	0.061 (0.37)	-1.378*** (0.36)
Hispanic	-0.592 (1.55)	-0.566 (0.40)	-0.689 (0.38)
Other	1.329 (1.39)	-0.140 (0.36)	-0.644 (0.34)
DLL	-2.892* (1.15)	0.444 (0.29)	0.451 (0.27)
HH Income < 20k	4.430 (3.09)	1.148 (0.82)	1.416 (0.77)
HH Income 21-40k	1.293 (2.30)	-0.035 (0.61)	0.708 (0.57)
HH Income 41-60k	1.205 (2.22)	0.845 (0.59)	0.531 (0.56)
HH Income 61-80k	0.663 (2.00)	0.526 (0.54)	0.634 (0.50)
FPL < 100	-7.190* (3.08)	-1.124 (0.82)	-1.745* (0.77)
FPL 100 to 300	-3.679 (2.07)	-0.375 (0.56)	-1.008 (0.52)
High Scope	-0.766 (1.14)	0.333 (0.30)	0.422 (0.27)
Class Size	-0.295 (0.18)	0.007 (0.05)	-0.026 (0.04)
Teacher Qual. Exceeds	2.766 (1.48)	0.375 (0.39)	-0.139 (0.35)
Teacher Qual. Meets	2.719* (1.23)	-0.095 (0.33)	-0.198 (0.29)
Teacher Black	2.655 (1.74)	0.360 (0.46)	-0.407 (0.42)
Teacher Hispanic	0.407 (1.60)	0.587 (0.43)	0.083 (0.39)
Teacher Asian	0.856 (1.80)	-0.331 (0.47)	-0.169 (0.43)
Teacher Other	2.497 (2.15)	-0.478 (0.55)	0.269 (0.49)
ECERS-3	1.332 (0.91)	0.493* (0.24)	0.394 (0.21)
N	913	811	810

* p<0.05; ** p<0.01; *** p<0.001. Notes for Table 14 are applicable here. Errors are clustered at the classroom level.

Table C.2. Multivariate analyses of children's 2018–19 raw score gains in relation to child and site or classroom characteristics & CLASS dimensions, excluding FCCs

Variables	Rec. Vocabulary (PPVT/TVIP)	Literacy (WJ/WM-LW)	Math (WJ/WM-AP)
3-year-olds	-0.811 (1.68)	-0.159 (0.44)	-0.006 (0.41)
Asian	-2.047 (1.54)	-0.038 (0.40)	-0.236 (0.37)
Black	-3.272* (1.44)	0.056 (0.37)	-1.386*** (0.35)
Hispanic	-0.514 (1.54)	-0.555 (0.39)	-0.704 (0.37)
Other	1.379 (1.39)	-0.139 (0.36)	-0.637 (0.34)
DLL	-3.037** (1.15)	0.470 (0.29)	0.497 (0.27)
HH Income < 20k	4.081 (3.08)	1.199 (0.81)	1.490* (0.76)
HH Income 21-40k	1.091 (2.30)	-0.012 (0.61)	0.699 (0.57)
HH Income 41-60k	1.244 (2.22)	0.838 (0.59)	0.501 (0.55)
HH Income 61-80k	0.520 (2.00)	0.465 (0.54)	0.556 (0.50)
FPL < 100	-7.148* (3.07)	-1.162 (0.82)	-1.765* (0.76)
FPL 100 to 300	-3.624 (2.07)	-0.334 (0.56)	-0.928 (0.52)
High Scope	-0.256 (1.11)	0.400 (0.29)	0.497* (0.25)
Class Size	-0.244 (0.18)	0.011 (0.05)	-0.015 (0.04)
Teacher Qual. Exceeds	1.900 (1.49)	0.111 (0.39)	-0.455 (0.33)
Teacher Qual. Meets	2.294 (1.21)	-0.089 (0.32)	-0.256 (0.27)
Teacher Black	1.839 (1.64)	0.166 (0.43)	-0.552 (0.38)
Teacher Hispanic	1.328 (1.59)	0.913* (0.43)	0.392 (0.37)
Teacher Asian	0.301 (1.80)	-0.066 (0.47)	0.184 (0.41)
Teacher Other	1.955 (2.08)	-0.544 (0.53)	0.217 (0.45)
CLASS ES average	-4.130* (2.05)	0.405 (0.54)	0.735 (0.46)
CLASS CO average	4.303* (1.74)	0.059 (0.46)	0.075 (0.40)
CLASS IS average	0.082 (0.94)	0.481 (0.25)	0.394 (0.22)
N	913	811	810

* p<0.05; ** p<0.01; *** p<0.001. Notes for Table 14 are applicable here. Errors clustered at the classroom level.

Table C.3. Multivariate analyses of children's 2018–19 raw score gains in relation to child and site or classroom characteristics & CLASS dimensions, including FCCs

Variables	Rec. Vocabulary (PPVT/TVIP)	Literacy (WJ/WM-LW)	Math (WJ/WM-AP)
3-year-olds	-0.880 (1.64)	-0.244 (0.43)	0.042 (0.40)
Asian	-1.889 (1.52)	-0.003 (0.39)	-0.117 (0.36)
Black	-3.453* (1.41)	0.012 (0.36)	-1.209*** (0.34)
Hispanic	-0.894 (1.51)	-0.675 (0.39)	-0.620 (0.36)
Other	1.589 (1.37)	-0.103 (0.35)	-0.582 (0.33)
DLL	-3.026** (1.12)	0.544 (0.28)	0.466 (0.26)
HH Income < 20k	4.480 (2.97)	1.298 (0.78)	1.324 (0.73)
HH Income_21-40k	0.886 (2.26)	0.008 (0.60)	0.670 (0.56)
HH Income 41-60k	1.265 (2.20)	0.814 (0.59)	0.518 (0.55)
HH Income 61-80k	0.525 (1.99)	0.487 (0.54)	0.572 (0.50)
FPL < 100	-7.407* (2.95)	-1.250 (0.78)	-1.639* (0.73)
FPL100 to 300	-3.567 (2.05)	-0.357 (0.56)	-0.962 (0.51)
FCC	-11.262* (4.78)	-0.106 (1.21)	-0.978 (1.06)
High Scope	-0.273 (1.09)	0.384 (0.29)	0.501* (0.25)
Class Size	-0.251 (0.18)	-0.008 (0.05)	-0.021 (0.04)
Teacher Qual. Exceeds	1.879 (1.46)	0.179 (0.39)	-0.426 (0.33)
Teacher Qual. Meets	2.203 (1.18)	-0.070 (0.31)	-0.232 (0.27)
Teacher Black	1.749 (1.57)	0.198 (0.42)	-0.560 (0.36)
Teacher Hispanic	1.609 (1.52)	0.877* (0.41)	0.333 (0.35)
Teacher Asian	0.342 (1.77)	-0.146 (0.47)	0.095 (0.41)
Teacher Other	1.842 (2.05)	-0.544 (0.53)	0.189 (0.45)
CLASS ES average	-3.798 (1.98)	0.271 (0.53)	0.598 (0.45)
CLASS CO average	4.113* (1.66)	0.116 (0.45)	0.143 (0.39)
CLASS IS average	0.150 (0.91)	0.400 (0.25)	0.351 (0.21)
<i>N</i>	950	849	848

* p<0.05; ** p<0.01; *** p<0.001. Notes for Table 14 are applicable here. Errors clustered at the classroom level.

Table C.4. Multivariate analyses of children's 2018–19 standard score gains in relation to child/site/classroom characteristics & ECERS-3 with Agency Fixed Effects, excluding FCCs

Variables	Rec.	Literacy	Math	Executive Function	
	Vocabulary (PPVT/TVIP)	(WJ/WM- LW)	(WJ/WM-AP)	DCCS	PT
3-year-olds	-1.118 (1.21)	-0.912 (1.15)	-0.672 (1.20)	-0.129 (0.07)	0.008 (0.65)
Asian	-1.755 (1.13)	0.435 (1.05)	-0.731 (1.10)	-0.148* (0.07)	-1.160 (0.60)
Black	-3.137** (1.05)	0.134 (0.97)	-4.084*** (1.04)	-0.296*** (0.06)	-0.895 (0.56)
Hispanic	-0.728 (1.13)	-1.182 (1.04)	-1.832 (1.10)	-0.084 (0.07)	-0.268 (0.60)
Other	0.944 (1.01)	0.131 (0.94)	-1.547 (0.99)	-0.075 (0.06)	0.124 (0.54)
DLL	-1.343 (0.87)	0.995 (0.78)	1.496 (0.82)	-0.064 (0.05)	0.973* (0.45)
HH Income < 20k	3.418 (2.22)	1.383 (2.12)	4.884* (2.23)	0.212 (0.13)	1.014 (1.22)
HH Income 21-40k	0.541 (1.65)	-0.617 (1.58)	2.168 (1.66)	-0.013 (0.10)	-0.136 (0.91)
HH Income 41-60k	0.667 (1.59)	1.759 (1.53)	1.205 (1.61)	0.024 (0.10)	-0.013 (0.88)
HH Income 61-80k	0.236 (1.43)	1.394 (1.39)	1.850 (1.45)	0.111 (0.09)	-0.070 (0.80)
FPL < 100	-5.162* (2.20)	-1.699 (2.11)	-5.797** (2.21)	-0.227 (0.13)	-2.288 (1.21)
FPL100 to 300	-2.146 (1.48)	-0.774 (1.44)	-3.080* (1.51)	-0.112 (0.09)	-0.808 (0.82)
High Scope	0.314 (1.74)	1.166 (1.66)	1.429 (1.75)	-0.155 (0.10)	0.633 (0.96)
Class Size	-0.065 (0.15)	0.096 (0.14)	-0.153 (0.14)	0.011 (0.01)	0.119 (0.08)
Teacher Qual Exceeds	2.159 (1.18)	0.626 (1.09)	-0.339 (1.15)	0.029 (0.07)	1.020 (0.63)
Teacher Qual Meets	2.005* (0.85)	-0.244 (0.79)	0.083 (0.83)	0.010 (0.05)	0.664 (0.46)
Teacher Black	3.352* (1.35)	2.865* (1.31)	-0.122 (1.37)	0.020 (0.08)	-0.393 (0.75)
Teacher Hispanic	-0.430 (1.14)	1.909 (1.11)	1.278 (1.16)	-0.065 (0.07)	1.169 (0.63)
Teacher _Asian	1.929 (1.22)	0.241 (1.14)	0.450 (1.19)	0.056 (0.07)	0.809 (0.65)
Teacher Other	0.913 (1.57)	-2.055 (1.42)	0.593 (1.49)	-0.101 (0.09)	0.922 (0.82)
ECERS-3	1.482* (0.65)	1.082 (0.59)	0.957 (0.62)	0.080* (0.04)	-0.322 (0.34)
<i>N</i>	913	809	810	807	811

* p<0.05; ** p<0.01; *** p<0.001. Notes for Table 14 are applicable here. Standardized scores are used for PPVT, and WJ or WM. Errors are clustered by classroom.

Table C.5. Multivariate analyses of children's 2018–19 standard score gains in relation to child/site/classroom characteristics & CLASS dimensions with Agency Fixed Effects, including FCCs

Variables	Rec.	Literacy	Math	Executive Function	
	Vocabulary (PPVT/TVIP)	(WJ/WM-LW)	(WJ/WM-AP)	DCCS	PT
3-year-olds	-1.340 (1.19)	-1.463 (1.13)	-0.503 (1.16)	-0.121 (0.07)	-0.026 (0.64)
Asian	-1.479 (1.12)	0.393 (1.04)	-0.423 (1.08)	-0.139* (0.07)	-0.902 (0.59)
Black	-3.044** (1.03)	0.090 (0.96)	-3.550*** (1.00)	-0.272*** (0.06)	-0.687 (0.54)
Hispanic	-0.937 (1.10)	-1.637 (1.02)	-1.665 (1.06)	-0.082 (0.06)	-0.364 (0.58)
Other	1.075 (0.99)	-0.012 (0.94)	-1.410 (0.97)	-0.065 (0.06)	0.171 (0.53)
DLL	-1.513 (0.85)	1.197 (0.77)	1.379 (0.80)	-0.056 (0.05)	1.004* (0.44)
HH Income < 20k	3.622 (2.15)	1.481 (2.07)	4.127 (2.14)	0.212 (0.13)	1.216 (1.18)
HH Income_21-40k	0.488 (1.63)	-0.496 (1.58)	2.043 (1.63)	0.001 (0.10)	-0.084 (0.90)
HH Income 41-60k	0.841 (1.58)	1.518 (1.54)	1.152 (1.59)	0.021 (0.10)	0.052 (0.87)
HH Income 61-80k	0.071 (1.42)	1.227 (1.40)	1.653 (1.44)	0.110 (0.09)	-0.048 (0.79)
FPL < 100	-5.369* (2.12)	-1.606 (2.06)	-5.016* (2.12)	-0.223 (0.13)	-2.560* (1.16)
FPL 100 to 300	-2.408 (1.48)	-0.562 (1.46)	-2.943* (1.50)	-0.096 (0.09)	-0.910 (0.82)
FCC	-15.206* (6.01)	-5.276 (5.41)	-10.301 (5.57)	-0.192 (0.34)	-1.040 (3.06)
High Scope	1.661 (1.89)	0.136 (1.82)	0.516 (1.86)	-0.243* (0.11)	1.164 (1.03)
Class Size	-0.097 (0.15)	-0.050 (0.14)	-0.241 (0.14)	0.009 (0.01)	0.100 (0.08)
Teacher Qual. Exceeds	2.012 (1.18)	1.417 (1.11)	0.206 (1.15)	0.074 (0.07)	0.970 (0.63)
Teacher Qual. Meets	2.328** (0.86)	0.314 (0.81)	0.418 (0.83)	0.025 (0.05)	0.852 (0.46)
Teacher Black	2.030 (1.29)	2.014 (1.26)	-0.693 (1.29)	-0.004 (0.08)	0.207 (0.71)
Teacher Hispanic	1.075 (1.16)	2.615* (1.14)	1.862 (1.16)	-0.055 (0.07)	1.081 (0.64)
Teacher Asian	1.279 (1.25)	0.647 (1.20)	1.083 (1.23)	0.095 (0.08)	0.724 (0.68)
Teacher Other	1.294 (1.54)	-1.952 (1.43)	0.509 (1.46)	-0.090 (0.09)	0.907 (0.80)
CLASS ES average	-2.443 (1.59)	2.146 (1.52)	2.680 (1.56)	0.188* (0.10)	-0.525 (0.86)
CLASS CO average	0.861 (1.37)	-2.393 (1.31)	-1.880 (1.35)	-0.053 (0.08)	-0.654 (0.74)
CLASS IS average	1.371 (0.72)	1.529* (0.69)	1.766* (0.71)	0.046 (0.04)	0.285 (0.39)
<i>N</i>	950	840	848	845	849

* p<0.05; ** p<0.01; *** p<0.001. Notes for Table 14 are applicable here.

Appendix D. P-values for tests of differences in means.

Table D.1. P-values for T-tests comparing distributions

P(T<=t) two-tail	16' vs. 17'	17' vs. 18'	18' vs. 19'	18' vs. 19' <i>including FCCs both yrs</i>
ECERS-3	0.049	0.442	0.001	n/a
CLASS ES	0.346	0.444	0.018	0.014
CLASS CO	0.620	0.021	0.032	0.021
CLASS IS	0.107	0.099	0.171	0.018

Table D.2. P-values for T-tests for comparisons of CLASS means between classrooms and FCCs

Domains and Dimensions	P-value
Emotional Support	0.000
1. Positive Climate	0.000
2. Negative Climate*	0.320
3. Teacher Sensitivity	0.000
4. Regard for Student Perspectives	0.054
Classroom Organization	0.000
5. Behavior Management	0.013
6. Productivity	0.000
7. Instructional Learning Formats	0.006
8. Facilitation of Learning & Dev.	n/a
Instructional Support	0.401
9. Concept Development	0.581
10. Quality of Feedback	0.053
11. Language Modeling	0.473

Table D.3. P-values for T-tests for comparisons of CLASS and ECERS by curriculum

CLASS and ECERS	P-value
CLASS Emotional Support	0.041
1. Positive Climate	0.035
2. Negative Climate*	0.946
3. Teacher Sensitivity	0.008
4. Regard for Student Perspectives	0.011
CLASS Classroom Organization	0.747
5. Behavior Management	0.757
6. Productivity	0.986
7. Instructional Learning Formats	0.874
8. Facilitation of Learning & Dev.	n/a
CLASS Instructional Support	0.165
9. Concept Development	0.057
10. Quality of Feedback	0.367
11. Language Modeling	0.405
ECERS-3 Total	0.046
1. Space and Furnishings	0.990
2. Personal Care Routines	0.003
3. Language and Literacy	0.001
4. Learning Activities	0.269
5. Interaction	0.232
6. Program Structure	0.144

Table D.4. P-values for T-tests and Bonferroni tests comparing quality across children subgroups, includes FCCs

	Ethnicity	Gender	FPL	DLL
	Bonferroni Prob>chi2	T-Test Pr(T > t)	Bonferroni Prob>chi2	Bonferroni Prob>chi2
ECERS-3	0.022	0.145	0.304	0.003
CLASS ES	0.001	0.129	0.000	0.257
CLASS CO	0.020	0.181	0.001	0.703
CLASS IS	0.085	0.423	0.009	0.296