NIEER
NATIONAL INSTITUTE FOR EARLY EDUCATION RESEARCH

UNIVERSITY of WASHINGTON

childcare quality & early learning
CENTER FOR RESEARCH AND PROFESSIONAL DEVELOPMENT
UNIVERSITY OF WASHINGTON

YEAR 1 REPORT: SEATTLE PRE-K PROGRAM EVALUATION

September 2016

Milagros Nores,Ph.D., Steve Barnett,Ph.D., Gail Joseph,Ph.D. Sara Stull,Ph.D. Alexandra Figueras-Daniel, Ph.D. & Janet S. Soderberg, Ph.D., The National Institute for Early Education Research & The Childcare Quality & Early Learning Center for Research & Professional Development

About the Authors

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

W. Steve Barnett, Ph.D. Dr. Barnett is Director of the National Institute for Early Education Research (NIEER) and 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.

Gail Joseph, Ph.D. Dr. Joseph is an associate professor in the area of Educational Psychology and Early Childhood and Family Studies 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 Director of the EarlyEdU Alliance and principal investigator and Director of the University of Washington's Childcare Quality and Early Learning Center for Research and Professional Development (CQEL).

Sara Stull, Ph.D. Dr. Stull is a Research Associate at the University of Washington's Childcare Quality and Early Learning Center for Research and Professional Development (CQEL). She leads CQEL's work on the SPP Evaluation Study and conducts research at CQEL on topics related to early childhood education and assessment. She has contributed to CQEL's research on kindergarten entry assessment and QRIS evaluation.

Alexandra Figueras-Daniel, Ph.D. Dr. Figueras-Daniel is an Assistant Research Professor at the National Institute for Early Education. Her research interests include classroom practices and professional development, dual language learners and classroom practices that support dual language acquisition, and measurement of quality in multicultural contexts. She has contributed extensively to NIEER’s research on DLLs and classroom quality.

Janet S. Soderberg, Ph.D. Dr. Soderberg is the Director of University of Washington's Childcare Quality and Early Learning Center for Research and Professional Development (CQEL). Dr. Soderberg's research includes exploration of the association between classroom quality and children’s development, QRIS program evaluation, refinement and support of kindergarten entrance assessments, and ongoing public dissemination of research results. Her interests include child development, assessment, childcare quality, and multi-systems alignment.

Grateful acknowledgment is made to Erica Johnson and all the support from Seattle's Preschool Program and Joelle Gruber from Third Sector Intelligence for their support on this project. In addition, the authors would like to thank Zilong Li for his invaluable statistical support for this report.

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

Suggested citation: Nores, M., Barnett, W.S., Joseph, G., Stull, S. Figueras-Daniel, A. & Soderberg, J.S. (2016). Year 1 report: Seattle Pre-k program evaluation. New Brunswick, NJ: National Institute for Early Education Research & WA: Childcare Quality & Early Learning Center for Research & Professional Development.

Table of Contents

Introduction

On November 4, 2014, Seattle voters approved a four-year, $58 million property tax levy to provide "accessible high-quality preschool services for Seattle children designed to improve their readiness for school and to support their subsequent academic achievement," subsequently giving birth to the Seattle Preschool Program (SPP). The city is investing SPP levy proceeds to achieve the following outcomes city-wide: Children will be ready for school, all students will achieve developmentally appropriate pre-academic skills, all students will develop both socially and emotionally, and the readiness gap will be eliminated for SPP participants. The city of Seattle's Department of Education and Early Learning (DEEL) launched SPP in the 2015–16 school year and will expand it over the following three years.

The four-year demonstration phase of SPP has three purposes. The first is to demonstrate that the approved structure is viable and capable of producing positive outcomes for Seattle's children. The second is to create, refine, and support a community infrastructure to improve the quality of preschool programs. The third is to create a process and norms that support continuous quality improvement (CQI) through evaluation. Like preschool programs that have demonstrated effectiveness, Seattle will use results from its initial years of evaluation to make course corrections to its programs.

In 2015, the Evaluation Team, composed by researchers at Third Sector Intelligence, Inc. (3SI), the National Institute for Early Education Research (NIEER) at Rutgers University, and the Childcare Quality & Early Learning (CQEL) Center for Research & Professional Development at the University of Washington, conducted a thorough review of the research on evaluation, supplemented with interviews of key leaders in program design and improvement. This review focused on studies of large-scale public preschool programs administered by cities and states, including some specifically identified by the city of Seattle as highly relevant to the SPP. The outcome of this process was an evaluation strategy including three distinct, but inter-related evaluations: (1) an impact evaluation to assess the extent to which SPP is increasing Kindergarten readiness of the students it serves; (2) a process evaluation to describe the implementation of the SPP program and evaluate quality and consistency amongst SPP preschool providers; and (3) a self-evaluation to help SPP providers, in partnership with and supported by DEEL staff, measure and improve their classrooms. While the strategy was designed so that the three support each other, each of the component evaluations is distinct and has its own timeline.

This report presents the first year (2015–16) findings from the impact evaluation. Its focus is on children's learning, the classroom quality they experienced, and how learning varied with differences in experiences. In addition to describing these findings for SPP, the report presents comparable findings from other preschool studies including the Head Start Family and Child Experiences Survey (FACES) to contextualize the results. The report also presents findings regarding children's attendance. Although the first year population served is very small and the data come from only 14 classrooms, we present findings for subgroups of students and classrooms as well as the full sample.

Overall, the report highlights that the first year of the SPP successfully enrolled a diversity of children across various groups of ethnic/racial groups and from the target of under 300% FPL. Children enrolled in SPP classrooms evidenced modest gains in vocabulary, literacy, math and executive functions; gains were larger than expected for their age in language and letter knowledge and smaller than expected in math. Attendance was consistently associated with higher outcomes, and bilingual children and those below the FPL had smaller gains than their

peers. Quality was lower than desired for the ECERS-3 and for the Instructional Support domain of the CLASS. Quality was relatively good for the Emotional Support and Classroom organization domains of the CLASS.

As with all evaluations, this one has limitations that must be acknowledged. It is important to understand that the first year of the impact evaluation was not designed to support causal claims about the effects of SPP on children's learning. In the start-up year it was not possible to create a randomized control group. The evaluation can describe the gains made by children, identify how outcomes vary with child and classroom characteristics, and compare SPP outcomes to those in other programs elsewhere. However, the evaluation cannot determine how much of the observed growth is due to SPP participation per se. Also, the first year sample is quite small, consisting of just 14 classrooms and the children who attended them. With such a small number of classrooms there is considerable uncertainty surrounding the estimates, which only increases for subgroups. Finally, information on children's family background and on attendance is incomplete. The report offers considerable depth into classroom quality observed and recommendations for the continuous improvement cycle.

Study Methods

The SPP evaluation study is a multi-year, multi-site study that includes a combination of designs to assess the program quality and impact on children over time. In the first year of the study, the research team collected child, family and classroom information to answer the following six questions:

1. Who are the children enrolled in SPP classrooms in 2015–16 and how do they compare to the demographics of children in Seattle more generally?
2. How did children enrolled in SPP classrooms progress over 2015–16?
3. What are SPP child attendance rates for 2015–16 and how do these compare to national averages?
4. What is the overall observed quality of children's interactions with teachers, each other, and the physical environment in 2015–16?
5. How does quality vary within SPP across children and providers?
6. What activities do children engage in, and is there scope for their interests and active participation?

The main purpose of the SPP impact evaluation is to estimate the effects of SPP on children's learning and development. In Year 1, the research team measured learning and development at the beginning and at the end of the year. Measures and procedures are described below. Children were assessed as programs and classrooms were incorporated into SPP for a pre-test, and assessed again at the end of the school year. Because programs varied in how early the research team was able to assess them, some children were assessed early in the fall, while some children were assessed in the winter. All children where then assessed late in the spring semester. In addition classroom observations of classroom practices were conducted to assess overall quality, teacher-child interactions, and engagement. Classroom observations were conducted between the months of February through the end of May. Quality was assessed using observation protocols widely established in the field. Figure 1 (below) reports the data collection timeline for the school year of 2015-16.

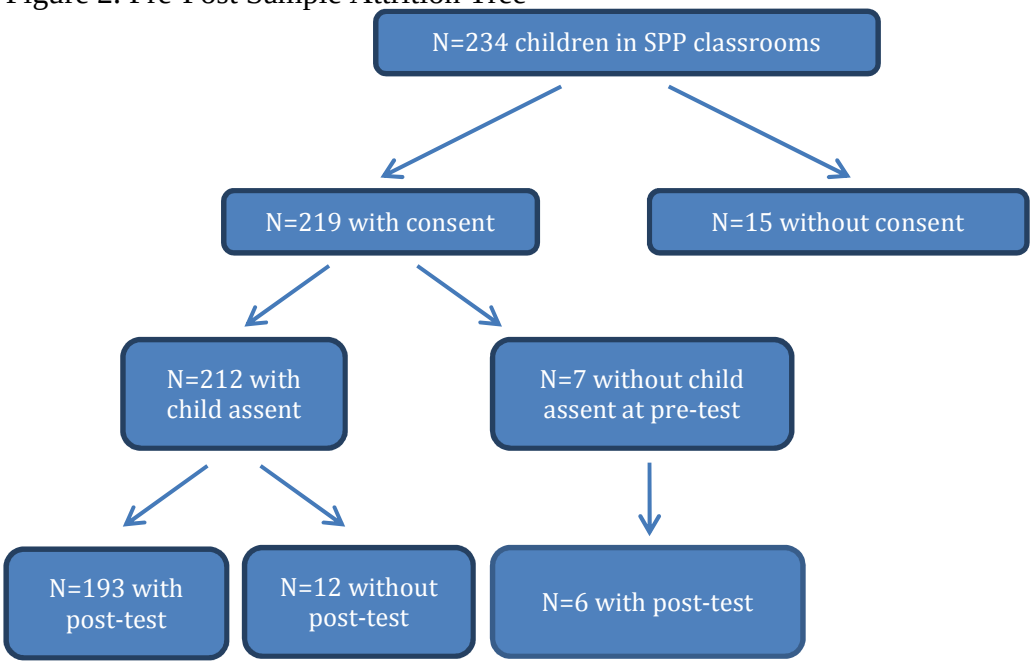
Figure 1. Data Collection Timeline

2015	
September	• Training for data collectors • Initial SPP site information gathered
October	• Parent consent form distribution • Fall assessment visit scheduling • Fall child assessment visits begin
November	• Fall child assessment visits continue
December	
2016	
January	• Calls to directors to discuss classroom observations (CLASS & ECERS-3)
February	• Unannounced CLASS & ECERS-3 observations (February through April)
March	• Family Survey distribution
April	• Spring assessment visit scheduling (early April) • Spring child assessment visits
May	
June	

Sample

We assessed 199 children in 14 SPP classrooms at pre- and post-test. Of the parents of the 234 children enrolled in these classrooms, 219 consented to participate in the study. Out of these, seven children declined participation, and only 193 were assessed at post-test (while of the seven that had declined participation, six did in fact assent at post-test for a total of 199 post-tests completed). Figure 2 below shows the study attrition tree. The analytic sample used in all analyses reported is of N=193, which are the children for which we have pre- and post-test data in at least one of the measures (not all children completed every measure).

Figure 2. Pre-Post Sample Attrition Tree



In addition, we conducted classroom observations on the 14 SPP classrooms from which we drew children. Classroom characteristics are described in Table 2. The classrooms in SPP in Year 1 used either a Creative Curriculum or a HighScope Curriculum, they reported an average class size of 18, they were distributed across five agencies, with about three classrooms per agency, and with average monthly attendance rates of 80 percent per classroom.

Table 1. Classroom characteristics, N=14

Classroom characteristic		Frequency or Mean (SD[1])
Curriculum	Creative	6
	HighScope	9
Class Size		17.64 (3.13)
Agencies		5
Average No. Classrooms per Agency		2.80 (1.48)
Average attendance rate		80.3 (9.04)

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. The PPVT is predictive of general cognitive abilities and is a direct measure of vocabulary size. The rank order of item difficulties is highly correlated with the frequency with which words are used in spoken and written language. The test is adaptive (to avoid floor and ceiling problems), establishing a floor below which the child is assumed to know all the answers and a ceiling above which the child is assumed to know none of the answers. The test is reliable based on reported split-half reliabilities or test-retest reliabilities. The PPVT has shown concurrent validity (e.g., Qi, Kaiser, Milan, & Hancock, 2006) and the results of these tests are found to be strongly correlated with school success (Blair & Razza, 2007; Early, et al., 2007).

The *Woodcock-Johnson Psycho-Educational Battery—Third Edition (WJ-III;* Woodcock, McGrew, Mather, & Schrank, 2001) includes multiple subtests. Only the *Applied Problems* and *Letter-Word Identification* subtests were used in this study. WJ-III was normed on a stratified random sample of 6,359 English-speaking subjects in the United States. Correlations of the WJ-R with other tests of cognitive ability and achievement are reported to range from 0.60 to 0.70. This measure has been used in numerous large-scale preschool studies (e.g., Early, et al., 2007; Wong, Cook, Barnett, & Jung, 2008).

Dimensional Change Card Sort Task (DCCS; Zelazo, 2006). This task engages reverse categorization where children must sort a set of cards based on different sorting criteria given by

[1] SD stands for standard deviation, which is a measure of variation in the data. That is, it measure 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.

the examiner. Generally, the test assesses attention-shifting. Scores on the DCCS reflect a pass/fail system on each of three levels of increasing difficulty. Raw scores range between 0 and 3, where a score of 0 means a child did not pass the first level which includes a color sorting task. At this first level, children are tasked with sorting two objects by color into a corresponding labeled box. A score of 1 means a child passed the color sort but failed the shape sort, which is the subsequent task and asks children to ignore color and instead sort objects by their shape. A score of 2 means a child passed shape sort but failed advance trials. Lastly, a score of 3 means the child passed advance trials, which ask children to ignore color or shape by adding a border to cards to indicate which attribute to sort by. There are no standard score equivalents. However, a study of test-retest reliability, means by age for children age 48 months or younger were 1.14 and for children 48-50 months means by age were 1.33, for 51-53 months they were 1.42, and for 54-56 months they were 1.58 (Meador et al., 2013).

Peg Tapping Test (PT; Diamond & Taylor, 1996). In this game, children are asked to tap a peg twice when the experimenter taps once and vice versa. The task requires children to inhibit a natural tendency to mimic the experimenter while remembering the rule for the correct response. Sixteen trials are conducted with 8 one-tap and 8 two-tap trials in random sequence. The task requires both 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 the ability to exercise inhibitory control over one's proponent behavior, the natural tendency to mimic what the experimenter does. Common errors include: (1) complying with only one of the two rules, (2) tapping many times regardless of what the experimenter did, and (3) doing the same thing as the experimenter, rather than the opposite. Table 14 below describes the gain score results for the PT task by delegate. 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 and for children 48-50 months means by age were 4.57, for 51-53 months they were 6.02, and for 54-56 months they were 7.87, (Meador et al., 2013).

Lastly, we conducted family surveys to collect information about the child and the family. In particular, the parent survey asks families to provide information regarding the following:

- Basic demographics of the child and family such as family income, education, employment status, race/ethnicity, languages spoken at home, and family structure and size
- Learning activities in the home, and other types of care and education the child may receive outside the home
- Family perceptions of early education or child care programs, and family perspectives or the benefits of SPP including impacts on their child's learning and development

Parental response rate was low (53.1%). Consequently, we have included the information collected on families in tables in Appendix D, but did not include this information in any analyses, and instead, included information on families collected by DEEL for race and ethnicity, language at home, and in relation to the Federal Poverty Level (FPL) throughout all analyses.

Measures on Classrooms

Early Childhood Environment Rating Scale—Third Ed. (ECERS-3; Harms, Clifford & Cryer, 2014)

The ECERS-3 is an observation and rating instrument for preschool classrooms serving children aged three to five. The total ECERS-3 score represents an average of the scores on the 35 items under 6 domains. A rating scale between 1 and 7 is used, where 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. The most updated notes for clarification[2] were utilized when scoring all classrooms in this sample. A general description of each of the 35 items on the ECERS-3 is provided in Table 2.

Table 2. 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.

[2] Published online at http://ersi.info/ECERS-33_notes.html in August, 2015.

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 (CLASS; Pianta, La Paro, & Hamre, 2008)

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

Scores (codes) are assigned during various classroom activities, and then averaged across all cycles for an overall quality score. Interactions are measured through 10 dimensions, which are divided into 3 domains. The Emotional Support domain is measured by 4 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 3 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 instrument is broken down into domains, and then further broken down into dimensions, all of which are outlined in Table 3 below.

Table 3. CLASS Domains and Dimension Descriptions.

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

Procedures

The University of Washington hired and trained data collectors on the child standardized assessment and classroom observation measures. For child assessments, data collectors received a two-day training on the measures. Following the two-day training, data collectors were successfully shadowed by expert staff on two iterations of the assessments for reliability. After two iterations of assessments, each of the data collectors achieved 100% reliability.

For classroom observation measures, trained and reliable observers are necessary for observations of classroom quality. Initial training was provided in administering the observation protocol that includes the ECERS-3 and the CLASS for preschool classrooms. Training took place in separate full-day workshops. ECERS-3 observers were trained by an ECERS-3 certified

trainer and met the ERSI[3] reliability requirements for observer certification. The trainee must complete three observations with the trainer with 85% or above exact matches or one-away from the true score. All data collectors met the ECERS-3 reliability requirements with agreement percentages ranging between 91-95%. CLASS observers were trained by a CLASS certified trainer and met the Teachstone reliability requirements for observer certification. All data collectors met CLASS reliability[4] requirements with agreement percentages ranging between 86-93%. All observation score sheets were cleaned and entered at UW by trained staff. Assessment procedures integrated culturally sensitive attitudes, knowledge, interview skills, intervention strategies and evaluation practices specifically informed by the age of the child.

Results

We address the research questions through a combination of descriptive and statistical analyses, and each research question separately. These draw from the sample of children and classrooms described earlier.

1. Who are the children enrolled in SPP classrooms in 2015-16 and how do they compare to the demographics of children in Seattle more generally?

Children's demographics are summarized in Table 2 below. Children assessed were equally 3 and 4 year-olds. Children in this study were predominantly from English-speaking households (72.4%), and a small portion speaks Spanish at home (8.5%). Other languages spoken, albeit by a very small portion of children, included Vietnamese, Amharic, Mandarin, Somali, and Oromo (among others). About 71.2% of the children were under the 300% Federal Poverty Level (FPL). Children showed significant variation across parent-reported race and ethnicity, with the four major groups being White (58.6%), Black (24.1%), Asian (13.1%) and Hispanic (12.6%).

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

Child Characteristics	SPP Children 2015-16		Seattle Public Schools
	N	%	
Gender			
Male	99	49.7%	51.5%[a]
Female	100	50.3%	48.5%[a]
Age at Post-Test			
3-Year-Olds	11	5.5%	-
4-Year-Olds	92	46.2%	-
5-Year-Olds	96	48.2%	-
Primary Language			
English	144	72.4%	65.0%[b]
Spanish	17	8.5%	7.0%[b]
Vietnamese	6	3.0%	3.0%[b]
Amharic	3	1.5%	<1.0%[b]

[3] ERSI is the company that sells ECERS-3 products. More information about the tool, as well as reliability guidelines, can be found at http://www.ersi.info/

[4] Teachstone is the company that sells CLASS products and manages/sells CLASS observer trainings, certifications etc. All training activity is monitored and reported to them. http://www.teachstone.com/about-teachstone/.

Mandarin	3	1.5%	3.0%[b]
Somali	2	1.0%	4.0%[b]
Oromo	2	1.0%	<1.0%[b]
Other	9	4.5%	
Unknown	13	6.5%	-
FPL Percentage			
Less than 100%	59	29.6%	38.9%[a,c]
100 – 199%	31	15.6%	
200 – 299%	52	26.1%	-
>300%	43	21.6%	-
Missing	14	7.0%	-
Race/Ethnicity			
White	57	28.6%	45.6%[a]
Black	48	24.1%	16.4%[a]
Asian	26	13.1%	15.8%[a]
Hispanic	25	12.6%	12.4%[a]
Multi-Racial	11	5.5%	8.5%[a]
Other	2	1.0%	1.3%[a]
Unknown	30	15.0%	-

[a]Seattle Public Schools as reported in http://www.seattleschools.org/district/district_quick_facts.
[b]Students attending Seattle Public Schools, as reported in Rivers (2016).
[c]Based on Free and Reduce Lunch which is for families <185% FPL.

2. How did children enrolled in SPP classrooms progress over 2015-16?

This evaluation reports standardized measures of child outcomes in two content areas: receptive vocabulary (Peabody Picture Vocabulary Test), literacy (Woodcock-Johnson Tests of Achievement Letter-Word subtest) and math (Woodcock-Johnson Tests of Achievement Applied Problems subtest). In addition, the 2015-2016 evaluation included measures of executive functions (EF) as it did in the previous year: Dimensional Change Card Sort Game (DCCS) and Peg Tapping task (PT). These two assess two areas of EF in early childhood, which include the ability to inhibit automatic response tendencies that can interfere with achieving a task, and the capacity for set shifting.

We present descriptive results from the 2015-2016 evaluation showing gains for the whole sample, and then split out by various subgroups, by agency, comparing classrooms with class sizes under 18 with classrooms with class sizes above 18, and comparing lower quality to higher quality classrooms. Children's learning gains are compared to the gains reported by various other studies, as well. Finally, we report results from multivariate analyses that examines variations in outcomes with all of the child and program characteristics simultaneously. Receptive vocabulary measured by the PPVT is presented first, followed by early math (WJ-AP), and two measures of executive function, the Dimensional Change Card Sort Game (DCCS) and the Peg Tapping task (PT).

Receptive vocabulary results

Table 3 presents children's vocabulary scores results for the fall (pre-test) and spring (post-test) with change from fall to spring (the gain). The PPVT-4 measures children's receptive vocabulary

relative to English speaking peers. Standardized scores—which are adjusted for age—are reported in this section (raw scores are reported in Appendix A, Table A.1). Thus, 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 alone. The mean standard score for this measure is set at 100 which is another way of saying that the average child in the U.S. population is expected to score 100 at any age. The standard deviation is 15. Information on this table reflects the performance of all children regardless of language background. Here we report scores for all children with valid scores in both the fall and spring of the school year. In addition, children from Spanish speaking homes were tested in Spanish as well as in English and their Spanish language vocabulary scores are reported later.

Overall, children scored at average in the fall and slightly above the average in the spring. One year gains for the whole group of children were of 2.03 standard points, which is about close to half of the reported one-year gains for 4-year-olds in the FACES study of 4.5 standard points although Head Start children scored well below average before and after a year in the program (Table B.5a; Aikens, et. al, 2013). Minority children score considerably below average and make the largest gains. Children speaking languages other than English score the lowest and make the largest gains. Children below the poverty level score lower than their higher income peers but also make the largest gains and end up near the national average. By contrast, FACES reported larger 2009 PPVT-4 standard gains for four-year-olds of 3.4 for White children, 4.3 for Black children and 8.7 for Hispanic children (summarized in Barnett, 2013). However, these children start at a much lower level and even with the larger gains do not approach the national average.

Table 3. Receptive vocabulary means and gains by child characteristics

		PPVT 2015 Fall		PPVT 2016 Spring		PPVT Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=189)		100.31	17.03	102.34	15.94	2.03	11.04
Gender	Male (N=92)	99.17	18.13	101.66	15.73	2.49	10.68
	Female (N=97)	101.38	15.94	102.98	16.19	1.60	11.40
Age	Three Year Cohort (N=37)	97.43	13.24	100.32	16.85	2.89	10.84
	Four Year Cohort (N=152)	101.01	17.80	102.83	15.73	1.82	11.11
Ethnicity	White (N=55)	111.75	13.98	112.64	12.01	0.89	12.94
	Black (N=47)	96.34	13.63	97.77	15.22	1.43	10.11
	Asian (N=24)	91.21	17.92	96.25	16.48	5.04	12.13
	Hispanic (N=22)	94.23	16.33	97.32	14.07	3.09	9.74
	Other (N=13)	98.31	20.63	100.15	18.13	1.85	12.64
	Unknown (N=28)	98.00	15.54	99.96	15.51	1.96	7.55
Language	English (N=127)	105.09	16.21	106.35	14.80	1.27	12.25
	Spanish (N=13)	85.38	10.98	91.08	13.80	5.69	8.95
	Vietnamese (N=5)	80.40	6.58	83.80	9.47	3.40	4.28
	Other (N=19)	86.47	11.55	90.11	13.50	3.63	9.79
	Unknown (N=25)	98.28	15.66	100.80	15.71	2.52	6.19
FPL	<100 (N=55)	94.65	13.53	98.36	15.97	3.71	11.94
	100-300 (N=80)	99.58	17.76	99.66	15.09	0.09	10.55
	>300 (N=40)	110.60	15.36	113.07	12.53	2.48	11.87
	Unknown (N=14)	97.29	18.29	102.57	17.18	5.29	4.89

Table 4 reports children's pre and post vocabulary standard scores for selected center characteristics (raw scores are reported in Appendix A, Table A.2). Mean scores by agencies are

below the norm for two agencies (1 & 5) and above for three (2-5) at both pre- and post-test. However, lower scoring agencies show the higher gains between fall and spring. Higher quality classrooms, as measured by the ECERS and CLASS domains, evidence higher average gain patterns.

Table 4. Receptive vocabulary means and gains by center characteristics

		PPVT 2015 Fall		PPVT 2016 Spring		PPVT Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=189)		100.31	17.03	102.34	15.94	2.03	11.04
Agency	Agency 1 (N=41)	96.39	13.32	99.15	17.32	2.76	11.08
	Agency 2 (N=74)	102.23	17.15	105.20	15.16	2.97	10.02
	Agency 3 (N=15)	107.20	18.59	102.53	17.67	-4.67	14.53
	Agency 4 (N=23)	108.65	17.78	109.52	11.61	0.87	10.58
	Agency 5 (N=36)	92.61	15.88	95.42	14.85	2.81	11.26
Class Size	18 or Less (N=87)	101.16	15.32	103.16	14.03	2.00	9.62
	More than 18 (N=102)	99.58	18.41	101.64	17.44	2.06	12.17
Curriculum	Creative Curriculum (N=64)	100.80	16.06	102.88	16.21	2.08	10.85
	HighScope (N=125)	100.06	17.56	102.06	15.86	2.01	11.17
ECERS	Less than 3 (N=15)	106.07	20.53	105.53	16.66	-.53	10.99
	3 or More (N=174)	99.81	16.67	102.06	15.90	2.25	11.05
CLASS ES	Less than 5.5 (N=28)	104.25	18.36	102.82	16.11	-1.43	11.56
	5.5 or More (N=161)	99.62	16.75	102.25	15.96	2.63	10.87
CLASS CO	Less than 5.5 (N=51)	103.16	17.56	101.92	14.69	-1.24	11.13
	5.5 or More (N=138)	99.25	16.77	102.49	16.43	3.24	10.80
CLASS IS	Less than 3 (N=116)	100.88	16.36	101.98	14.86	1.10	9.91
	3 or More (N=73)	99.40	18.12	102.90	17.62	3.51	12.56

Literacy results

Table 5 reports children's WJ-III letter-word identification scores for the overall sample and by selected child characteristics. This test measures children's ability to identify letters and then read a list of words of increasing difficulty in isolation. Again, the test has a mean standard (i.e., age adjusted score) of 100 and a standard deviation of 15 (raw scores are reported in Appendix A, Table A.3). Scores for all children are included regardless of their home language background. Mean scores are reported for children with valid scores in both the fall and spring of the school year.

Children scored on average above the norms in both the fall and the spring. One year gains for the whole group of children were of 2.60 standard points, which is about a half of the reported one-year gains for 4-year-olds in the FACES study of 5.0 standard points. As with the PPVT, Head Start children scored below the average of 100 (Table B.5a; Aikens, et. al, 2013). Across groups, 3-year-old, White, and English speaking children had largest gains. By contrast, FACES reported for Head Start in 2009 LW standard gains for four-year-olds of 4.3 for White children, 4.8 for Black children and 5.3 for Hispanic children (summarized in Barnett, 2013).

Table 5. Literacy means and gains by child characteristics

		WJ-LW 2015 Fall		WJ-LW 2016 Spring		WJ-LW Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=186)		101.99	13.30	104.59	13.48	2.60	10.10
Gender	Male (N=92)	100.64	11.35	103.43	12.00	2.79	11.03
	Female (N=94)	103.31	14.91	105.72	14.76	2.41	9.16
Age	Three Year Cohort (N=36)	98.75	11.39	103.47	11.91	4.72	13.94
	Four Year Cohort (N=150)	102.77	13.64	104.86	13.85	2.09	8.93
Ethnicity	White (N=56)	102.27	11.46	107.36	11.83	5.09	11.18
	Black (N=47)	101.47	15.25	103.96	15.86	2.49	9.77
	Asian (N=25)	104.48	10.52	105.08	9.75	0.60	6.54
	Hispanic (N=17)	92.53	10.04	93.59	12.67	1.06	8.24
	Other (N=13)	104.31	9.13	105.31	11.38	1.00	6.78
	Unknown (N=28)	104.75	16.74	106.04	14.08	1.29	12.72
Language	English (N=129)	101.02	12.57	104.43	13.01	3.41	9.97
	Spanish (N=8)	95.13	7.61	95.25	12.69	0.13	8.64
	Vietnamese (N=5)	103.80	8.81	100.80	8.61	-3.00	4.69
	Other (N=19)	107.32	13.35	109.00	15.20	1.68	6.95
	Unknown (N=25)	104.76	17.49	105.80	14.70	1.04	13.39
FPL	<100 (N=55)	100.31	11.91	102.91	13.33	2.60	8.37
	100-300 (N=76)	101.95	13.75	104.38	14.06	2.43	10.11
	>300 (N=41)	104.17	11.61	108.10	11.57	3.93	10.35
	Unknown (N=14)	102.43	19.89	102.07	15.41	-0.36	15.04

Table 6 reports children's pre- and post-test letter-word identification standard scores for children in the sample across selected center characteristics (raw scores are reported in Appendix A, Table A.4). Children's gains differ across classrooms and are higher when CLASS instructional quality is above 3.

Table 6. Literacy means and gains by center characteristics

		WJ-LW 2015 Fall		WJ-LW 2016 Spring		WJ-LW Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=186)		101.99	13.30	104.59	13.48	2.60	10.10
Agency	Agency 1 (N=41)	100.71	10.81	101.27	11.45	0.56	8.97
	Agency 2 (N=75)	101.83	13.55	105.72	15.03	3.89	9.84
	Agency 3 (N=14)	104.43	9.34	108.07	12.09	3.64	10.46
	Agency 4 (N=23)	107.52	17.66	108.96	14.28	1.43	8.45
	Agency 5 (N=33)	99.06	12.93	101.64	10.92	2.58	12.69
Class Size	18 or Less (N=87)	102.45	14.48	103.99	13.89	1.54	9.17
	More than 18 (N=99)	101.59	12.24	105.12	13.16	3.54	10.81
Curriculum	Creative Curriculum (N=64)	103.16	13.93	104.03	12.97	0.88	8.73
	HighScope (N=122)	101.38	12.98	104.89	13.78	3.51	10.67
ECERS	Less than 3 (N=15)	99.33	10.29	100.53	10.95	1.20	9.24
	3 or More (N=171)	102.22	13.53	104.95	13.65	2.73	10.19
CLASS ES	Less than 5.5 (N=28)	96.82	13.01	100.93	10.36	4.11	11.58
	5.5 or More (N=158)	102.91	13.18	105.24	13.88	2.34	9.83
CLASS CO	Less than 5.5 (N=51)	100.14	13.23	102.82	11.64	2.69	10.06
	5.5 or More (N=135)	102.69	13.31	105.26	14.09	2.57	10.15
CLASS IS	Less than 3 (N=114)	101.17	14.22	102.58	13.48	1.41	10.43
	3 or More (N=72)	103.29	11.69	107.78	12.93	4.49	9.32

Early math results

Table 7 shows children's pre- and post-test math scores, as measured by the applied problems subscale of the WJ-III. Again, the scale is normed with a mean of 100 and a standard deviation of 15. On average, children in the SPP sample scored above average in the fall and spring of the school year, with negative "gains" indicating that they lost ground relative to expectations for their higher age at post-test (average raw score gains are reported in table A.5 in appendix A, all of which are positive). This contrasts with one-year gains for four-year-olds in the FACES study of 2.2 standard points although Head Start children in such study scored below the norm throughout (Table B.5a; Aikens, et. al, , 2013). Among children in the sample, Asian, Hispanic, and Spanish-speaking children outperformed their peers. FACES reported 2009 AP standard gains for four-year-olds of 1.4 for White children, 0.6 for Black children and 4.2 for Hispanic children (summarized in Barnett, 2013).

Table 7. Math means and gains by child characteristics

		WJ-AP 2015 Fall		WJ-AP 2016 Spring		WJ-AP Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=186)		106.34	12.97	104.80	11.62	-1.54	9.43
Gender	Male (N=92)	104.66	14.36	104.36	12.46	-0.30	10.42
	Female (N=94)	107.99	11.29	105.23	10.79	-2.76	8.23
Age	Three Year Cohort (N=36)	103.67	11.42	104.44	12.12	0.78	8.08
	Four Year Cohort (N=150)	106.99	13.27	104.89	11.54	-2.10	9.67
Ethnicity	White (N=56)	112.88	11.80	110.66	10.55	-2.21	9.43
	Black (N=47)	101.38	13.72	100.60	9.95	-0.79	9.20
	Asian (N=25)	102.56	11.28	104.84	13.67	2.28	11.13
	Hispanic (N=17)	105.29	10.18	106.47	9.75	1.18	9.95
	Other (N=13)	107.23	14.46	101.69	10.92	-5.54	7.03
	Unknown (N=28)	105.21	11.57	100.54	11.31	-4.68	7.46
Language	English (N=129)	107.94	13.13	106.32	10.91	-1.62	9.44
	Spanish (N=8)	103.50	12.59	105.75	12.06	2.25	9.32
	Vietnamese (N=5)	94.60	7.64	95.80	11.92	1.20	9.78
	Other (N=19)	102.16	13.21	101.79	14.47	-0.37	11.80
	Unknown (N=25)	104.56	11.33	100.76	11.38	-3.80	7.22
FPL	<100 (N=55)	102.36	12.76	102.38	10.91	0.02	8.82
	100-300 (N=76)	105.64	13.31	103.87	12.99	-1.78	10.49
	>300 (N=41)	113.56	10.38	110.88	8.76	-2.68	8.58
	Unknown (N=14)	104.64	11.31	101.57	7.84	-3.07	8.01

We report children's pre and post standardized math scores and gains by selected center characteristics in Table 8 (raw scores are reported in Appendix A, Table A.6). Again, there is some variation between agencies, and for different quality levels, children in higher quality classrooms as measured by the ECERS make more progress over the year. This is also the case for children in classrooms with smaller class sizes and implementing HighScope.

Table 8. Math means and gains by center characteristics

		WJ-AP 2015 Fall		WJ-AP 2016 Spring		WJ-AP Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=186)		106.34	12.97	104.80	11.62	-1.54	9.43
Agency	Agency 1 (N=41)	102.95	11.14	100.98	10.27	-1.98	7.19
	Agency 2 (N=75)	107.12	13.59	105.27	12.34	-1.85	10.16
	Agency 3 (N=14)	110.93	10.66	107.86	12.63	-3.07	11.82
	Agency 4 (N=23)	113.13	12.22	110.43	12.25	-2.70	8.23
	Agency 5 (N=33)	102.12	12.83	103.27	8.96	1.15	9.91
Class Size	18 or Less (N=87)	106.01	13.15	104.72	12.06	-1.29	9.51
	More than 18 (N=99)	106.64	12.87	104.87	11.29	-1.77	9.41
Curriculum	Creative Curriculum (N=64)	106.61	12.46	104.38	11.85	-2.23	7.52
	HighScope (N=122)	106.20	13.28	105.02	11.55	-1.18	10.31
ECERS	Less than 3 (N=15)	109.40	14.64	104.73	12.48	-4.67	8.13
	3 or More (N=171)	106.08	12.83	104.81	11.59	-1.27	9.51
CLASS ES	Less than 5.5 (N=28)	106.00	13.68	103.43	10.26	-2.57	9.59
	5.5 or More (N=158)	106.41	12.89	105.04	11.86	-1.36	9.43
CLASS CO	Less than 5.5 (N=51)	105.76	13.33	102.96	11.63	-2.80	9.29
	5.5 or More (N=135)	106.56	12.88	105.50	11.59	-1.07	9.48
CLASS IS	Less than 3 (N=114)	105.75	13.61	104.29	10.99	-1.46	9.80
	3 or More (N=72)	107.28	11.93	105.61	12.60	-1.67	8.89

Executive functions

We used two measures of executive functions. The DCCS is an attention shifting test which uses short term memory. Table 9 shows children's pre- and post-test DCCS scores by selected child characteristics. The full sample exhibited gains of .17 in the DCCS which may seem small, but is meaningful as it is a quarter of standard deviation. With exception of the Vietnamese group, which is really a very small group of children, all subgroups of children showed gains between fall and spring.

Table 9. DCCS means and gains by child characteristics

		DCCS 2015 Fall		DCCS 2016 Spring		DCCS Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=192)		1.53	0.60	1.70	0.62	0.17	0.61
Gender	Male (N=94)	1.48	0.58	1.67	0.59	0.19	0.59
	Female (N=98)	1.58	0.62	1.72	0.64	0.14	0.63
Age	Three Year Cohort (N=38)	1.16	0.59	1.26	0.64	0.11	0.73
	Four Year Cohort (N=154)	1.62	0.57	1.81	0.56	0.18	0.58
Ethnicity	White (N=56)	1.79	0.53	1.91	0.51	0.13	0.63
	Black (N=47)	1.38	0.64	1.45	0.58	0.06	0.60
	Asian (N=25)	1.52	0.51	1.68	0.63	0.16	0.47
	Hispanic (N=23)	1.52	0.51	1.78	0.60	0.26	0.62
	Other (N=13)	1.23	0.60	1.85	0.69	0.62	0.77
	Unknown (N=28)	1.43	0.69	1.57	0.69	0.14	0.52

Language	English (N=129)	1.57	0.61	1.74	0.58	0.18	0.61
	Spanish (N=14)	1.50	0.52	1.79	0.58	0.29	0.61
	Vietnamese (N=5)	1.20	0.45	1.00	0.00	-0.20	0.45
	Other (N=19)	1.53	0.51	1.63	0.90	0.11	0.81
	Unknown (N=25)	1.44	0.71	1.60	0.58	0.16	0.47
FPL	<100 (N=57)	1.40	0.68	1.54	0.66	0.14	0.64
	100-300 (N=80)	1.44	0.57	1.68	0.65	0.24	0.64
	>300 (N=41)	1.85	0.42	1.98	0.35	0.12	0.56
	Unknown (N=14)	1.64	0.63	1.64	0.63	0.00	0.39

No norms exist for the measure. As a reference, the Learning-Related Cognitive Self-Regulation School Readiness Measures for Preschool Children Study (aka 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; ranges which include the average ages at fall and spring testing in this study (53.4 months in the fall and 58.9 in the spring). Children in SPP show similar gain patterns. Within the different groups, four-year-old, Hispanic, Other ethnic/race and Spanish-speaking children evidence larger gain patterns than their peers.

Children's pre and post DCCS scores are shown in Table 10. There are apparent differences in gain between agencies. Children's gains on the DCCS are higher in smaller classrooms. Gains on the DCCS do not differ in the same way across the two quality measures.

Table 10. DCCS means and gains by center characteristics

		DCCS 2015 Fall		DCCS 2016 Spring		DCCS Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=192)		1.53	0.60	1.70	0.62	0.17	0.61
Agency	Agency 1 (N=41)	1.41	0.71	1.51	0.60	0.10	0.54
	Agency 2 (N=76)	1.59	0.61	1.74	0.57	0.14	0.58
	Agency 3 (N=15)	1.80	0.56	1.93	0.70	0.13	0.83
	Agency 4 (N=23)	1.65	0.49	1.96	0.56	0.30	0.63
	Agency 5 (N=37)	1.35	0.48	1.57	0.65	0.22	0.63
Class Size	18 or Less (N=87)	1.48	.64	1.72	.60	.24	.65
	More than 18 (N=105)	1.57	.57	1.68	.63	.10	.57
Curriculum	Creative Curriculum (N=64)	1.50	.64	1.67	.62	.17	.58
	HighScope (N=128)	1.55	.59	1.71	.62	.16	.62
ECERS	Less than 3 (N=16)	1.63	.62	1.69	.60	.06	.44
	3 or More (N=176)	1.52	.60	1.70	.62	.18	.62
CLASS ES	Less than 5.5 (N=29)	1.41	.57	1.62	.62	.21	.62
	5.5 or More (N=163)	1.55	.61	1.71	.62	.16	.61
CLASS CO	Less than 5.5 (N=52)	1.40	.60	1.63	.66	.23	.70
	5.5 or More (N=140)	1.58	.60	1.72	.60	.14	.57
CLASS IS	Less than 3 (N=118)	1.45	.62	1.67	.61	.22	.64
	3 or More (N=74)	1.66	.56	1.74	.62	.08	.54

Table 11 shows children's pre- and post-test Peg Tapping scores by selected child characteristics. Peg tapping is a measure of inhibitory control. No norms exist for this measure, either. Children in SPP across all subgroups gained between fall and spring of the school year, with an overall gain of 2.98 for the full sample. The Self-Regulation Measurement Study (Meador, et. al, 2013) also included this measure and authors reported average scores of 6.02 at

51-53 months and 8.80 at 57-59 months, with a difference of 2.78. SPP children fare well in comparison to their sample.

Table 11. Peg Tapping means and gains by child characteristics

		PT 2015 Fall		PT 2016 Spring		PT Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=193)		6.76	5.84	9.74	5.61	2.98	5.33
Gender	Male (N=95)	6.59	5.91	9.87	5.64	3.28	5.54
	Female (N=98)	6.92	5.78	9.60	5.59	2.68	5.13
Age	Three Year Cohort (N=38)	2.58	4.22	6.08	5.80	3.50	3.91
	Four Year Cohort (N=155)	7.78	5.73	10.63	5.19	2.85	5.63
Ethnicity	White (N=56)	8.38	5.67	11.59	4.86	3.21	5.39
	Black (N=47)	5.87	5.90	7.81	6.15	1.94	5.85
	Asian (N=25)	6.04	6.52	10.40	6.07	4.36	5.63
	Hispanic (N=24)	6.67	5.27	10.17	4.51	3.50	4.37
	Other (N=13)	7.38	5.74	10.46	5.13	3.08	6.02
	Unknown (N=28)	5.43	5.63	7.96	5.62	2.54	4.43
Language	English (N=129)	7.09	5.89	10.29	5.68	3.21	5.46
	Spanish (N=15)	5.67	4.95	7.87	3.94	2.20	4.89
	Vietnamese (N=5)	-0.20	1.79	5.00	6.12	5.20	6.69
	Other (N=19)	8.68	5.80	10.47	5.19	1.79	5.62
	Unknown (N=25)	5.64	5.62	8.36	5.75	2.72	4.51
FPL	<100 (N=58)	5.90	6.06	9.26	6.04	3.36	6.22
	100-300 (N=80)	6.80	6.00	9.41	5.57	2.61	4.47
	>300 (N=41)	8.34	5.34	11.90	4.62	3.56	5.79
	Unknown (N=14)	5.43	4.73	7.21	5.18	1.79	4.63

Pre- and post-test Peg-Tapping scores for children in the sample across selected center characteristics are shown in Table 12. There is some variation across agencies. The quality measures are not associated with gains in the expected direction.

Table 12. Peg-Tapping means and gains by center characteristics

		PT 2015 Fall		PT 2016 Spring		PT Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=193)		6.76	5.84	9.74	5.61	2.98	5.33
Agency	Agency 1 (N=41)	6.37	6.02	9.34	6.30	2.98	5.74
	Agency 2 (N=76)	6.91	5.79	9.71	5.22	2.80	5.55
	Agency 3 (N=16)	10.75	4.23	12.75	3.87	2.00	2.90
	Agency 4 (N=23)	7.13	5.64	10.52	5.92	3.39	5.98
	Agency 5 (N=37)	4.92	5.81	8.43	5.72	3.51	4.95
Class Size	18 or Less (N=87)	6.22	5.39	9.38	5.74	3.16	5.46
	More than 18 (N=106)	7.20	6.17	10.03	5.51	2.83	5.24
Curriculum	Creative Curriculum (N=64)	6.64	5.85	9.77	6.15	3.13	5.78
	HighScope (N=129)	6.81	5.85	9.72	5.34	2.91	5.11
ECERS	Less than 3 (N=16)	6.31	6.54	9.44	5.45	3.13	4.95
	3 or More (N=177)	6.80	5.79	9.76	5.63	2.97	5.37
CLASS ES	Less than 5.5 (N=29)	5.00	5.90	8.59	5.54	3.59	4.31
	5.5 or More (N=164)	7.07	5.79	9.94	5.61	2.87	5.49
CLASS CO	Less than 5.5 (N=52)	5.87	5.83	9.29	5.53	3.42	5.20
	5.5 or More (N=141)	7.09	5.82	9.90	5.64	2.82	5.39
CLASS IS	Less or Equal to 3 (N=118)	5.68	5.62	9.08	5.62	3.40	5.21
	More than 3 (N=75)	8.45	5.79	10.77	5.45	2.32	5.48

Multivariate Analyses

Multivariate analyses examine the association between children's learning gains and program features that include agency, curriculum and classroom quality while simultaneously controlling for age of children, children's characteristics and attendance rate. We include information on the age of children, gender, race and ethnicity, home language, FPL, and attendance rates, as provided by DEEL. Classroom level components include class size, curriculum, agency and classroom quality. The analyses also take into account that scores of children who are in classrooms together cannot be considered to be independent of each other. For children assessed Spanish, the Spanish scores are used (with a control variable to take this into account). We conduct separate analyses with the two measures of quality, one controlling for quality as measured by the ECERS-3 and the other for quality as measured by the CLASS.

Table 13 presents estimates of these associations with ECERS-3 as the measure of classroom quality. Statistically significant results are highlighted in boldface. Attendance was positively associated with gains in receptive vocabulary and math. A five percentage point increase in attendance would translate into a 0.75 standard score gain. Children's outcomes did not differ between boys and girls. Gains were slightly lower for Bilingual children (relative to monolingual) even after testing Spanish-speaking children in Spanish. One of the agencies seems to have an advantage over the others even after controlling for quality (which may be related to teacher experience and qualification, for which we had no information at the time of this study). Children in the 100-300 FPL group make smaller gains than others. Black children made smaller gains on the DCCS. Also, children for which race is unknown gained less in math than their White peers. Children for whom attendance is unknown seemed to perform better, which may be worth investigating. On the other hand, children with missing information on race/ethnicity or FPL gained less than others.

The indicator for classroom size is positively associated with letter-word recognition and DCCS scores, an odd result. As for classroom quality, no association was found between the ECERS-3 and children's performance taking into account all else including agency.

Table 13. Multivariate analyses of children's 2015-16 gains in relation to child and site or classroom characteristics and ECERS-3

Variables	Rec. Vocabulary (PPVT/TVIP)	Literacy (WJ/WM-LW)	Math (WJ/WM-AP)	Executive Function	
				DCCS	PT
Pre Test	**0.620*****	**0.703*****	**0.598*****	**0.015***	**0.469*****
	(0.05)	(0.05)	(0.05)	(0.01)	(0.06)
Attendance	**0.173***	0.066	**0.157****	0.004	0.046
	(0.08)	(0.07)	(0.06)	(0.00)	(0.04)
Missing Attendance	**14.061***	11.006	**12.777***	0.528	**6.352***
	(6.74)	(6.16)	(5.29)	(0.38)	(3.05)
Days Between Tests	-0.134	**-0.192****	-0.080	-0.001	0.013
	(0.07)	(0.07)	(0.06)	(0.00)	(0.03)
Female	-0.214	0.286	-0.856	0.057	-0.434
	(1.36)	(1.27)	(1.08)	(0.08)	(0.63)
Black	-3.663	0.574	-1.524	**-0.274***	-1.803
	(2.22)	(2.04)	(1.76)	(0.12)	(1.02)
Asian	-3.097	-2.903	1.741	-0.170	0.667
	(2.58)	(2.30)	(1.99)	(0.14)	(1.16)
Hispanic	0.006	-3.605	3.468	0.062	1.005

	(2.78)	(2.77)	(2.31)	(0.16)	(1.26)
Other Race	-4.489	-3.285	-4.164	-0.065	-0.528
	(2.98)	(2.72)	(2.33)	(0.16)	(1.36)
Missing Race	-7.742	-0.984	**-7.537***	-0.383	-1.889
	(4.51)	(4.21)	(3.59)	(0.25)	(2.13)
Bilingual	**-4.911***	-1.313	-1.278	-0.144	-1.652
	(2.12)	(1.91)	(1.63)	(0.11)	(0.92)
Missing Language	2.909	1.586	3.511	0.296	1.641
	(4.51)	(4.25)	(3.62)	(0.26)	(2.12)
FPL <100	-1.682	-1.300	0.100	-0.181	-0.594
	(2.29)	(2.09)	(1.80)	(0.13)	(1.04)
FPL 100-300	**-4.523***	-1.811	-1.657	**-0.224***	-1.368
	(1.95)	(1.79)	(1.54)	(0.11)	(0.90)
Missing FPL	2.081	-9.298	0.144	-0.398	**-5.404***
	(5.18)	(4.77)	(4.07)	(0.29)	(2.35)
Agency 2	4.901	4.672	4.091	-0.021	-1.711
	(3.54)	(3.39)	(2.82)	(0.21)	(1.63)
Agency 3	-3.486	5.402	3.563	0.112	-0.083
	(3.42)	(3.19)	(2.68)	(0.20)	(1.56)
Agency 4	6.462	**9.290***	3.290	**0.528***	-0.098
	(4.59)	(4.32)	(3.64)	(0.27)	(2.10)
Agency 5	-1.638	0.173	2.560	-0.223	-1.715
	(4.12)	(3.95)	(3.25)	(0.24)	(1.89)
Class Size	0.847	**1.342****	0.119	**0.063***	0.102
	(0.53)	(0.48)	(0.41)	(0.03)	(0.24)
ECERS	4.089	-0.228	2.808	-0.114	-1.123
	(3.01)	(2.87)	(2.37)	(0.18)	(1.38)
N	189	191	193	193	193

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are Males, White, English, FPL 300+, and Agency 1. Other controls are age in months, test type for children tested only in Spanish (for the letter word test, and an interaction between test type for age, and an indicator for attendance, language or FPL information missing. Standardized scores are used for PPVT, and WJ or WM.

Analyses in Table 14 estimate these associations with the three CLASS domains as measures of classroom quality. Results are quite similar to those in the previous table. In terms of classroom quality, the dimensions of CLASS emotional support and classroom organization do not show any association with children's gains. Unexpectedly, instructional supports score is negatively associated with DCCS gains. With CLASS IS scores being so low, these results are not necessarily inconsistent with findings in the literature in that CLASS IS is positively associated with children's executive function gains only at higher quality levels (e.g. Weiland et. al, 2012).

Table 14. Multivariate analyses of children's 2015-16 gains in relation to child and site or classroom characteristics CLASS dimensions

Variables	**Rec. Vocabulary (PPVT/TVIP)**	**Literacy (WJ/WM-LW)**	**Math (WJ/WM-AP)**	**Executive Function**	
				DCCS	**PT**
Pre Test	**0.617*****	**0.695*****	**0.598*****	**0.013***	**0.467*****
	(0.05)	(0.05)	(0.05)	(0.01)	(0.06)
Attendance	**0.159***	0.078	**0.152***	0.003	0.043
	(0.08)	(0.07)	(0.06)	(0.00)	(0.04)
Missing Attendance	12.580	12.188	**12.362***	0.363	5.990

	(6.83)	(6.27)	(5.38)	(0.39)	(3.15)
Days Between Tests	**-0.163***	**-0.152***	-0.084	-0.002	0.015
	(0.07)	(0.07)	(0.06)	(0.00)	(0.03)
Females	-0.243	0.066	-0.954	0.062	-0.436
	(1.34)	(1.26)	(1.08)	(0.08)	(0.63)
Black	-3.589	0.772	-1.466	**-0.299***	-1.897
	(2.18)	(2.02)	(1.75)	(0.12)	(1.02)
Asian	-3.271	-2.556	1.752	-0.188	0.640
	(2.55)	(2.29)	(1.98)	(0.14)	(1.16)
Hispanic	-0.088	-3.535	3.425	0.025	0.897
	(2.73)	(2.74)	(2.30)	(0.16)	(1.26)
Other Race	-3.907	-2.541	-3.921	-0.044	-0.393
	(2.93)	(2.71)	(2.32)	(0.16)	(1.37)
Missing Race	-8.273	-1.355	**-7.846***	-0.426	-2.032
	(4.43)	(4.17)	(3.56)	(0.25)	(2.13)
Bilingual	**-5.168***	-1.402	-1.429	-0.170	-1.758
	(2.09)	(1.89)	(1.62)	(0.11)	(0.92)
Missing Language	4.215	2.003	3.910	0.363	1.870
	(4.46)	(4.24)	(3.61)	(0.26)	(2.14)
FPL <100	-1.210	-1.278	0.184	-0.152	-0.513
	(2.27)	(2.08)	(1.80)	(0.13)	(1.05)
FPL 100-300	**-3.916***	-1.491	-1.488	-0.177	-1.198
	(1.92)	(1.79)	(1.54)	(0.11)	(0.90)
Missing FPL	2.380	**-9.243***	0.342	-0.360	**-5.370***
	(5.09)	(4.72)	(4.04)	(0.29)	(2.35)
Agency 2	5.932	**7.323***	2.877	0.321	0.061
	(3.19)	(2.99)	(2.55)	(0.18)	(1.52)
Agency 3	-4.574	6.490	1.876	0.152	0.445
	(3.88)	(3.57)	(3.04)	(0.22)	(1.79)
Agency 4	**15.758***	11.552	4.223	**1.266****	2.464
	(7.51)	(6.99)	(6.03)	(0.42)	(3.52)
Agency 5	-3.079	2.586	0.668	-0.009	-0.271
	(3.05)	(2.89)	(2.45)	(0.17)	(1.48)
Class Size	**2.114***	1.136	0.399	**0.133****	0.260
	(0.84)	(0.78)	(0.68)	(0.05)	(0.40)
CLASS ES	2.340	-2.551	2.077	0.131	0.026
	(3.90)	(3.68)	(3.10)	(0.22)	(1.84)
CLASS CO	5.032	3.063	1.118	0.246	0.811
	(3.04)	(2.86)	(2.44)	(0.17)	(1.45)
CLASS IS	-4.427	1.192	-0.770	**-0.346***	-0.948
	(3.04)	(2.84)	(2.42)	(0.17)	(1.43)
N	189	191	193	193	193

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are Males, White, English, FPL 300+, and Agency 1. Other controls are age in months, test type for children tested only in Spanish (for the letter word test, and an interaction between test type for age, and an indicator for attendance, language or FPL information missing. Standardized scores are used for PPVT, and WJ or WM.

We were not able to include curriculum as it correlated too closely with agency. However, alternative analyses (not shown) to examine curriculum that excluded agency found no significant associations of curriculum with any of the outcomes measures.

Sensitivity Analyses

Three types of sensitivity analyses were conducted to assess the robustness of our findings. First, we repeated the analyses with raw scores because imperfections in the standardization could affect results. Second, we conducted analyses to see if choice of English or Spanish assessment for children tested in both languages affected the results. Third, we investigated whether there might be a quality threshold that made a difference.

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

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

(2) Results using tests in English only are reported in Appendix Tables B.5 using ECERS and B.6 using CLASS. Findings are consistent with those reported earlier, with even less gains documented for bilingual children.

(3) Analyses investigating thresholds of quality are reported in Appendix Tables B.3 for ECERS and B.4 for CLASS.[5] We find that an ECERS-3 score greater than 3 has a positive association with letter-word recognition of almost 6 standard points. We observe a positive association for CLASS CO levels above 5.5 with the DCCS.

3. What are SPP child attendance rates for 2015-16 and how do these compare to national averages?

Attendance rates were available (as provided by DEEL) for 88 percent of the analytic sample (Table 15). Consequently, we also looked at how attendance rates varied by child and center characteristics. Attendance is defined as the average percentage attendance across months. For the children for whom DEEL had attendance data (88.08% of the children) average attendance percentage across months was of 81%. Attendance varied only slightly across groups, ranging between of 79% for Black children or children of unknown ethnic/racial background to 84% for Asian children, and between 77% for children from homes under 100% FPL to 84% for children from homes over 300% FPL. For comparison, Head Start has an 85% daily attendance rate rule under which the program is required to investigate the causes of absenteeism[6] and San Antonio's Pre-K SA maintained a 92% average attendance in their 2013-2014 AY (Edvance, 2014).

[5] Burchinal et al. (2010) found evidence of CLASS IS thresholds at 3.25, and CLASS ES in the 5-7 range, and Hatfield et al. (2016) found evidence of CLASS IS threshold at 3 and CLASS ES and CO at 6. Given the distributions of quality in the sample, we chose to use a level of 3 for the ECERS (which was quite low to start with) and levels of 5.5 for CLASS emotional support and classroom organization scales, and a level of 3 for CLASS instructional supports.

[6] "If a program's monthly average daily attendance rate falls below 85 percent, the program must analyze the causes of absenteeism to identify any systematic issues that contribute to the program's absentee rate. The program must use this data to make necessary changes… and inform its continuous improvement efforts as described in §1302.102(c)." Head Start Program Performance Standards. 45 CFR Chapter XIII. RIN 0970-AC63. Available at https://eclkc.ohs.acf.hhs.gov/hslc/hs/docs/hspss-final.pdf

Table 15. Average attendance percentage across month by child characteristics

		Pre-Post N	N with attendance data	% with attendance data	Attendance		% without attendance data
					Mean	SD	
Total		193	170	88.08%	81.36	11.98	11.92%
Gender	Male	95	85	89.47%	81.40	12.14	10.53%
	Female	98	85	86.73%	81.33	11.88	13.27%
Age	Three Year Cohort	38	32	84.21%	81.22	12.02	15.79%
	Four Year Cohort	155	138	89.03%	81.40	12.01	10.97%
Ethnicity	White	56	53	94.64%	82.85	9.40	5.36%
	Black	47	46	97.87%	78.80	14.16	2.13%
	Asian	25	23	92.00%	84.30	10.01	8.00%
	Hispanic	24	21	87.50%	81.00	13.70	12.50%
	Other	13	12	92.31%	83.33	10.65	7.69%
	Unknown	28	15	53.57%	78.40	13.66	46.43%
Language	English	129	124	96.12%	81.51	11.52	3.88%
	Spanish	15	11	73.33%	80.27	13.89	26.67%
	Vietnamese	5	4	80.00%	82.00	5.72	20.00%
	Other	19	18	94.74%	84.00	13.26	5.26%
	Unknown	25	13	52.00%	77.08	14.50	48.00%
FPL	<100	58	50	86.21%	77.12	13.87	13.79%
	100-300	80	80	100.00%	82.73	11.72	0.00%
	>300	41	40	97.56%	83.95	8.22	2.44%
	Unknown	14	0	0.00%	-	-	100.00%

In relation to classroom characteristics, attendance rates varied between 71% and 90% across agencies (Table 16). There is a 7 percentage point difference in attendance rates between classrooms using Creative and classrooms using HighScope. Attendance rates were also higher for children in lower quality classrooms.

Table 16. Attendance percentage by center characteristics

		Pre-Post N	N with attendance data	% with attendance data	Attendance		% without attendance data
					Mean	SD	
Total		193	170	88.08%	81.36	11.98	11.92%
Agency	Agency 1	41	32	78.05%	71.13	11.83	21.95%
	Agency 2	76	68	89.47%	82.40	10.11	10.53%
	Agency 3	16	15	93.75%	75.07	6.04	6.25%
	Agency 4	23	22	95.65%	83.95	12.70	4.35%
	Agency 5	37	33	89.19%	90.30	8.57	10.81%
Class Size	18 or Less	87	78	89.66%	79.78	12.03	10.34%
	More than 18	106	92	86.79%	82.71	11.83	13.21%
Curriculum	Creative Curriculum	64	54	84.38%	76.35	13.65	15.62%
	HighScope	129	116	89.92%	83.70	10.37	10.08%
ECERS	Low	16	15	93.75%	84.33	8.10	6.25%
	Medium	177	155	87.57%	81.08	12.27	12.43%
CLASS ES	Less than 5.5	29	27	93.10%	85.48	9.79	6.90%
	5.5 or More	164	143	87.20%	80.59	12.22	12.80%
CLASS CO	Less than 5.5	52	47	90.38%	85.40	9.92	9.62%
	5.5 or More	141	123	87.23%	79.82	12.36	12.77%
CLASS IS	Less or Equal to 3	118	105	88.98%	83.01	11.41	11.02%
	More than 3	75	65	86.67%	78.71	12.48	13.33%

4. What is the overall observed quality of children's interactions with teachers, each other, and the physical environment in 2015-16?

Average ECERS-3 Results Spring 2016

Scores for the 14 classrooms that were observed using the ECERS-3 are presented in Table 17 below. The minimum, maximum, and mean item scores for all 35 ECERS-3 items, six subscales and overall scores are shown.

Table 17. ECERS-3 Item, Subscale, and Overall Means and Ranges, N = 14

ECERS-R Item and Subscales	Mean	SD	Minimum	Maximum
Overall	**3.57**	**0.46**	**2.94**	**4.50**
Space and Furnishings	3.88	0.55	2.86	4.57
Personal Care Routines	3.14	0.65	1.75	4.25
Language and Literacy	3.47	0.83	2.40	5.20
Learning Activities	2.87	0.56	2.10	4.00
Interaction	4.49	0.90	3.20	5.80
Program Structure	4.43	0.97	2.67	6.00

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

Average CLASS Scores Spring 2016

The scores presented here reflect overall means for the 14 pre-K classrooms that were observed using the CLASS instrument. Table 18 presents the minimum, maximum, and means item scores for all ten CLASS dimensions and three domains.

Table 18. CLASS Dimension and Domain Means and Ranges, N = 14

CLASS Dimensions and Domains	Mean	SD	Minimum	Maximum
Emotional Support Domain	6.14	0.53	4.88	6.81
Classroom Organization Domain	5.67	0.74	4.17	6.58
Instructional Support Domain	2.65	0.71	1.50	4.25

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

Distribution of Classroom Quality across Classrooms

The distribution of scores for ECERS-3 and each CLASS domain are depicted in Figure 3.1, below. The figure shows how all ECERS-3 scores are below the good quality threshold of 5. On the other hand, classrooms score quite high on Emotional Supports, with most classrooms heavily concentrated around the mean score of 5.67. Classroom organization also shows good scores, with a good portion of scores about the level of 5. Instructional support scores are lower, with most scores concentrated between 2.2 and slightly above 3. These patterns are overall consistent with the field.

Figure 3.1. ECERS-3 and CLASS Domain distributions of scores as box plot

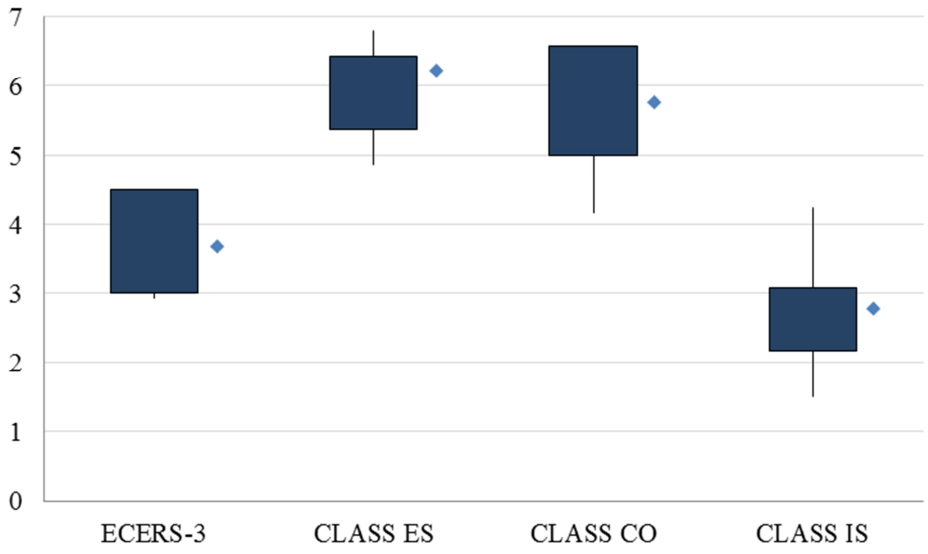
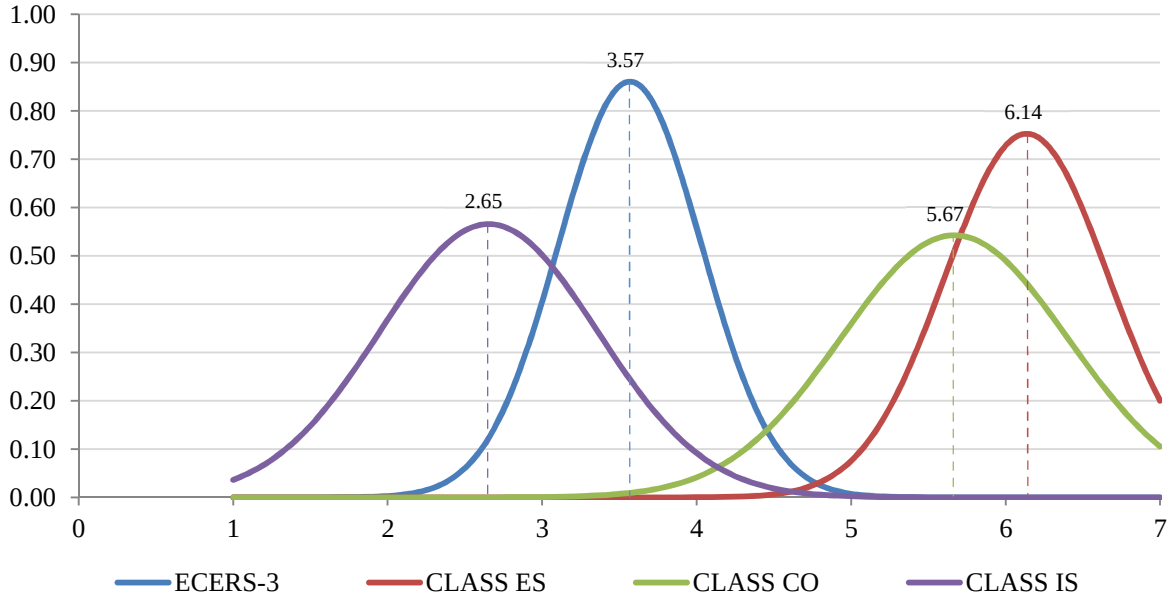


Figure 3.2. ECERS-3 and CLASS Domain distributions of normalized scores.



ECERS-3 is a newer version of the widely used ECERS-R measure. Like the ECERS-R, quality in the ECERS-3 is considered minimal when the average or subscale scores is between 3 and 5, as is the case of SPP classrooms. Higher quality classrooms are expected to average a score between 5 and 7. Table 19 provides for context average ECERS-3 scores for 3 studies: in GA, UW state pre-K and childcare centers and NJ Abbott districts for this same year. In addition, for comparison, we have included ECERS-R data (which allows seeing growth over time for the) two ECERS-R years provided) for two previous years for Abbott NJ districts.

Table 19. Studies with reported ECERS-3 scores

Study	1. Space/ Furnishing	2. Personal Care Routines	3. Language & Literacy	4. Learning Activities	5. Interaction	6. Program Structure	Average Total
SPP 2015-2016	**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)**
GA[1]	3.49	3.14	3.36	3.14	4.31	3.64	**3.46**
UW state pre-K & childcare study (n=299)[2]	3.45	2.89	3.40	2.68	3.88	3.63	**3.23**
NJ Abbott: 2015-16 (n=293)[3]	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)**
2007-08 (n=317)[4]	5.03	4.29	5.46	4.85	6.44	5.41	**5.20**
2002-03 (n=310)[4]	3.76 (1.00)	3.69 (1.35)	4.27 (1.30)	3.37 (0.94)	4.92 (1.60)	4.04 (1.57)	**3.96 (0.94)**

[1] Jenson (2015); [2] CQEL (Unpublished); [3] NIEER (2016); [4]ECERS-R was used in these evaluations. Available at http://www.state.nj.us/education/ece/research/elichome.htm

The National Overview of CLASS in Pre-K classrooms in 2015 (OHS, 2015) found that the highest scores were prevalent in the domain of Emotional Supports, with a national mean of 6.03, mid-high scores in the Classroom Organization section with a national mean of 5.80, and low scores in the Instructional Support domain with a national mean of 2.88. SPP classrooms show higher overall means on Emotional Supports and Classroom Organization but a lower average mean on the Instructional Support Domain. We report CLASS scores for this and other studies in comparison to SPP scores in Table 20 below. The SPP classroom average on emotional supports is higher than the averages reported in any of these studies. The SPP classroom average for classroom organization also show higher quality than several of these, while slightly lower than the average reported in NYC and the national average. However, on Instructional Supports, SPP classrooms scored on average lower than all minus one of the studies summarized. While the threshold suggested in the literature for quality is lower in instructional supports than for other CLASS dimensions, the SPP score is below thresholds found in the literature (see page 24).

Table 20. Classroom quality across the nation, and for selected programs

Study	Emotional Support	Classroom Organization	Instructional Support
SPP classrooms 2015-2016 (n=14)	**6.14 (0.53)**	**5.67 (0.74)**	**2.65 (0.71)**
Tulsa[1]			
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)
Boston[2] (n=83) (2009-2010)	5.63 (0.60)	5.10 (0.68)	4.30 (0.84)
NYC (n=555) (2012-13 to 2014-15)[3]	6.00	5.80	3.60
National Head Start Overview 2015[4]	6.03(0.28)	5.80(0.36)	2.88(0.54)
Head Start FACES 2009[5]	5.30	4.70	2.30
EA Validation study (n=75)[6]	5.96(0.66)	5.26 (0.77)	2.34(0.71)
NJ Abbott 2013-2014 (n=163)[7]	5.97 (0.63)	5.32 (0.89)	3.15 (0.96)

[1]Phillips et. al (2009); [2]Weiland et. al (2013); [3]NYC Department of Education (2015a,b); [4]Office of Head Start. (2015); [5] Aikens, et. al (2013); [6]CQEL (Unpublished); [7]NIEER (2014).

Interpreting ECERS-3 and CLASS scores

ECERS-3

The *Space and Furnishings* subscale examines the physical space of a classroom. Included are whether children have enough space and furniture, whether the arrangement of the furniture allows for learning and exploration and whether displays are meaningful and representative of the children in the class. Additionally, this subscale includes two items that assess the space and equipment used by children for gross motor, which mainly considers outside spaces. The range in scores on the items of this subscale demonstrates that some items had strong evidence, while other lacked considerably. The highest scoring item "Indoor Space" at a 6.43 indicates that the quality of classroom space reaches near excellence. The minimum of a 4.00 in the range indicates that the lowest score any classroom received was that of a 4.00, which means that classroom scoring a 4.00 are only missing one feature of the indicators under the "good" anchor point of the item. Specifically this could be due to the absence of direct natural lighting or ability to control ventilation in the classroom, or due to limited indoor space for children and adults to circulate easily.

The lowest scoring item in this subscale (aside from the gross motor items, which will be discussed subsequently) is that of "Child Related Display" where the average was of 3.36 and the range indicates that some classrooms scored a 1.00 (inadequate) rating. Generally this item seeks to find that in addition to the presence of appropriate classrooms displays, that these are meaningful to the children. To this end, the item rates not only the presence of displays that relate to children (made by children), but that they are individualized (e.g. process art as opposed to product art) and that it is used by teachers to spark conversation and motivate interactions that can build vocabulary. Scores of a 1.00 are likely the result of a complete lack of discussion of any wall display during the course of an observation. As with other items in the ECERS-3 despite that the wall display may be adequate for a high score, without the interaction piece, the item can not score high.

A second low score item is "room arrangement for play and learning" which examines that play areas have enough space for children to play and adults to move in. The item also calls

for the presence of a minimum of 3 interest centers for a score of 3.00 or "minimal" and five for a score of a 5.00 or "good." Additionally, the tool is extremely specific about how it defines an interest center. The final two low scoring items include "space for gross motor" and "gross motor equipment." Each of these has a time requirement of 15 minutes to receive credit in the "minimal" category of scoring and 30 minutes for "good." In addition, while these items are generally geared at assessing outside spaces, inside spaces and activities aimed at gross motor development can be considered for scoring, but have to still satisfy the time and safety requirements.

The second subscale *Personal Care Routines*, examines the health, hygiene and safety practices of the classroom. A large emphasis of both the "Meals/Snacks" item and the "Diapering/Toileting" item require that all children wash their hands with soap and water for a total of 20 seconds before and after each meal, as well as after each use of the bathroom. Other requirements for hand washing are outlined by the tool in the "Health Practices" item which requires five specific times for hand washing including before and after using wet or shared sensory materials, and upon arriving in the classroom. The low scores on these two items reveal that hand washing procedures likely need more attention. The "Safety Practices" item accounts for all safety hazards inside the classroom and in the outdoor gross motor area. This item categorizes safety hazards into two groups, major hazards and minor hazards. The difference between a major and a minor hazard is the degree to which a child can potentially be injured; major hazards could result in serious injury and potentially death, while minor hazards could result in only very minor injury. To earn a score of 5.00 or "good," there can be no more that two major safety hazards, and for a score of 7.00 or "excellent," no major safety hazards can be present, despite that a few minor hazards are acceptable. Other indicators of this item also include issues of supervision by staff.

The *Language and Literacy* subscale examines the ways that staff direct activities and materials towards developing children's language and literacy skills. The "Staff Use of Books" item received the lowest score of all the items in this subscale averaged at 3.07. This average score indicates that requirements of the 5.00 or "good" were not met. Even for a rating of 3.00 or "minimal" the requirements of the item include that staff read books to children at least once during the observation. To receive a score in the good to excellent range on this item all children must be actively involved during all story times. The other low scoring item of this subscale was the "Becoming Familiar with Print" item which averaged a 2.21. This item requires that most visible print is combined with pictures and staff take dictation of what children say in a way that is interesting and engaging to children for the purpose of showing that print is a useful tool.

The goal of the *Learning Activities* subscale is two-fold. First it seeks to assess the presence, variety, and accessibility of learning materials in the classroom for children. Second, it seeks to capture the extent to which teachers actively engage children with the materials assessed in each item. The items that comprise the Learning Activities subscale can not reach the "good" rating (5.00) without evidence during an observation of children interacting with teachers using the respective materials for that item. In short, this means that while the quantity and quality of materials needed to score high may be present, without evidence of interactions during the 3-hour observation period, the item can not score higher than a 2.00 or a 3.00, which is considered "minimal." Further, items in the Learning Activities subscale require careful attention to the way that the items define interest centers so that they can adequately arrange classrooms to satisfy the material requirements. Generally, the items within this subscale are designed to capture the extent to which teachers circulate around the classroom during a 3-hour period and utilize the

materials in the classroom to generate meaningful learning exchanges. While this can not necessarily be planned due to the child-directed nature of center-based learning, it can be intentional on the part of the teacher. In addition, the use of formative assessment of children can also help to meaningfully guide these interactions and make them part of everyday practice.

The *Interaction* subscale of the tool assesses the degree to which teachers supervise children during gross motor time, how they individualize teaching and learning and how children and teachers interact among each other. The "supervision of gross motor" item was the lowest scoring item of this subscale at 3.71. This means that in many cases some of the indicators in the "good" category were not observed. This category of the item requires that staff not only supervise children to ensure that they are safe, but also that staff interactions are all (almost all) positive and that they are highly interested in participating with children as they specifically engage in gross motor play activities.

The final subscale is that of *Program Structure* examines the general formats of the classroom and how the children spend their time. The lowest scoring item of this subscale was that of "whole-group activities for play and learning" which averaged 3.71. Generally this item takes into account the flexibility of the staff with children and how they respond to children's individual needs for different pacing in efforts to keep children engaged in group times. Ultimately, the indicators of this item seek to assess very specifically whether group times are meaningful and engaging for all the children in the class and how staff are being intentional about these times of the day. To achieve a score of "good" or a 5.00, staff would have to be seen being responsive to children's needs during group times (e.g. moving from a story to an interactive song when children become restless).

Table 20. ECERS-3 Item, Subscale, and Overall Means and Ranges by Item, N = 14

ECERS-R Item and Subscales	Mean	Minimum	Maximum
Space and Furnishings			
1. Indoor space	6.43	4.00	7.00
2. Furnishings for care, play and learning	4.36	4.00	7.00
3. Room arrangement for play and learning	3.64	2.00	7.00
4. Space for privacy	4.14	2.00	6.00
5. Child-related display	3.36	1.00	5.00
6. Space for gross motor play	3.14	1.00	4.00
7. Gross motor equipment	2.07	1.00	4.00
Personal Care Routines			
8. Meals/ snacks	3.07	1.00	4.00
9. Toileting/diapering	2.21	1.00	3.00
10. Health practices	2.93	2.00	4.00
11. Safety practices	4.36	2.00	7.00
Language and Literacy			
12. Helping children expand vocabulary	3.50	3.00	5.00
13. Encouraging children to use language	4.36	3.00	7.00
14. Staff use of books with children	3.07	1.00	6.00
15. Encouraging children's use of books	4.21	1.00	7.00
16. Becoming familiar with print	2.21	1.00	4.00

Learning Activities			
17. Fine motor	4.36	2.00	5.00
18. Art	3.71	2.00	6.00
19. Music and movement	3.50	2.00	5.00
20. Blocks	2.00	1.00	4.00
21. Dramatic Play	2.79	1.00	6.00
22. Nature/science	2.50	1.00	4.00
23. Math materials and activities	1.71	1.00	3.00
24. Math in daily events	2.86	1.00	5.00
25. Understanding written numbers	1.29	1.00	2.00
26. Promoting acceptance of diversity	4.21	3.00	6.00
Interaction			
27. Appropriate use of technology	N/A	1.00	1.00*
28. Supervision of gross motor	3.71	1.00	7.00
29. Individualized teaching and learning	4.21	3.00	7.00
30. Staff-child interaction	4.93	3.00	7.00
31. Peer interaction	5.00	3.00	7.00
32. Discipline	4.57	2.00	7.00
Program Structure			
33. Transitions and waiting times	4.86	3.00	7.00
34. Free play	4.50	3.00	6.00
35. Whole - group activities for play and learning	3.93	2.00	5.00

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

CLASS: Emotional Support Domain

The Emotional Support domain documents how the teacher fosters a nurturing and safe environment for children to learn. The "Positive Climate" and "Negative Climate" dimensions examine the emotional connection between teachers and students. Specifically, the Positive Climate dimension "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). The Negative Climate dimension "reflects the overall level of expressed negativity in the classroom" (p. 28). Mid to high level scores in the "Negative Climate" dimension indicate evidence of instances of observed harsh threats, yelling, a lack of eye contact, or sarcasm from the teachers while low level scores would indicate the opposite. Throughout this report, the Negative Climate scores have been inverted, meaning that high level scores indicate a lack of expressed negativity. The highest scoring dimension within the Emotional Support Domain is Negative Climate, with a mean of 6.86, meaning that teachers exhibited almost no negativity towards the children, and children exhibited very little negativity toward each other. Though "Positive Climate" had the lowest mean in this domain (5.80), it should not be thought of as contradictory with the results of the "Negative Climate" dimension. In short, it just means that while teachers were not negative towards students, there is still room for growth on supports that convey warmth, respect and enjoyment in the classroom.

The "Teacher Sensitivity" dimension considers the extent to which teachers are able to anticipate problems and provide support for children. One important feature of this domain is that teachers are not dismissive, as this too would affect scores negatively. An average score of 5.91 on this dimension reflects that teachers were mostly aware of children and responded to

children's emotions and needs for individualized support. In addition it seems that teachers were helpful in addressing problems and with conflict resolution as well as comforting children. To move from the mid range to the high range (6.00-7.00) on this item, would entail these same practices just more consistently.

Finally, the "Regard for Student Perspectives" dimension of this domain looks to see how comfortable students seem to be in their environment. Evidence for this is judged on how children participate, seek help and take risks, and whether the teachers foster an environment where children feel safe to behave in this way. The dimension also documents the degree to which interactions are based on children's interests and perspectives, and how well teachers encourage children to be autonomous. In this item, teachers are assessed on their flexibility and the amount of opportunities that they provide for children to share ideas. Again, as with the previous dimension, a mean score of 5.96 indicates that classrooms are in the upper range and higher levels of quality would be attained with the same practices on a more consistent level. That is, many opportunities for children to have time to express themselves and move about freely in the classroom as opposed to these happening "sometimes" or occasionally.

CLASS: Classroom Organization Domain

The Classroom Organization domain examines the supports through which the teachers manage behavior, time, and activities. The "Behavior Management" dimension examines not only whether behavior expectations are clear, but also whether they are consistent. This dimension also documents how proactive teachers are in preventing misbehavior. The "Productivity" dimension assesses the degree to which teachers manage time, pacing and transitions throughout the day and across activities. Finally, this domain also includes "Instructional Learning Formats" which measures how teachers maximize their facilitation of student learning during activities. This includes how effective questions are, how clear learning objectives are, and whether there is a range of opportunities for children to learn. Student interest is also taken into consideration in this dimension.

The highest scoring dimension within this domain was "Productivity" with a mean score of 6.05, also in the mid-high range. A score this high demonstrates that during the majority of the observation, the children had something to do and there were no periods of time where there were no activities offered. The lowest scoring dimension is "Instructional Learning Formats" with a mean score of 5.21, which constitutes a mid-range score. To score in the mid-high level on this dimension the teacher has to actively facilitate engagement from students through use of varied materials and modalities. In addition the teacher has to be seen focusing students on learning objectives and students should be seen consistently interested and engaged in activities with no periods of time where students are uninvolved.

CLASS: Instructional Supports Domain

The Instructional Supports Domain assesses the interactions through which teachers deliver and facilitate high-order thinking skills, and develop language. As mentioned previously, this domain is the most difficult, yet most important, domain when considering teacher practices that bare impacts on student growth, this is why this is at the same time, the one that consistently scores the lowest. The first dimension, "Concept Development," measures teachers' use of discussions to stimulate reasoning and analysis. It also looks to assess the extent to which teachers encourage

creativity as well as how they integrate concepts into children's lives. High scoring classrooms in this dimension are those where teachers are consistent and intentional about how they present questions and promote problem solving. A key element of this item is not that teachers do these things in isolation once or twice, but that they are consistently happening throughout the day. Concept Development was the lowest scoring dimension of this evaluation with a mean of 2.07. Increases in this dimension would entail the use of discussions and activities that foster reasoning and analysis by students as well as the opportunities for children to create and generate products from their own ideas. In addition the dimension also seeks to capture the presence of the teacher making associations for children and relating new concepts to those previously learned and students' lives.

Similarly, the "Quality of Feedback" dimension measures the quality of teacher responses to children's talk. It seeks to see whether teachers provide hints, are persistent, ask for explanations of thinking, and how specific they are in responses to children. High level classrooms in this dimension are those that find teachers scaffolding, helping children to solve a problem by providing resources or added questions, and doing so for as long as it takes the child to come to a resolution. This dimension scored a mean of 2.71 indicating work is needed to increase scores to a mid-range score. Included in this dimension is that teachers scaffold children through their process of problem solving or understanding a concept, as well as engage in feedback loops with children regularly. In addition teachers would need to expand on what students say and prompt students to explain their own thinking when they do provide a response. Finally teachers need to encourage students' efforts.

The final dimension under this domain is that of "Language Modeling" which measures both the quality and amount of teacher's language used for the purpose of developing language in children. This item averaged a 3.29, in the low-mid range. Mid-high range classrooms on this dimension would exhibit frequent conversations between teachers and children, many open-ended questions, and the use of self- and parallel talk when working with children in play areas. Finally the use of advanced language with students including the use of varied words.

Table 21. CLASS Dimension and Domain Means and Range by Item, N = 14

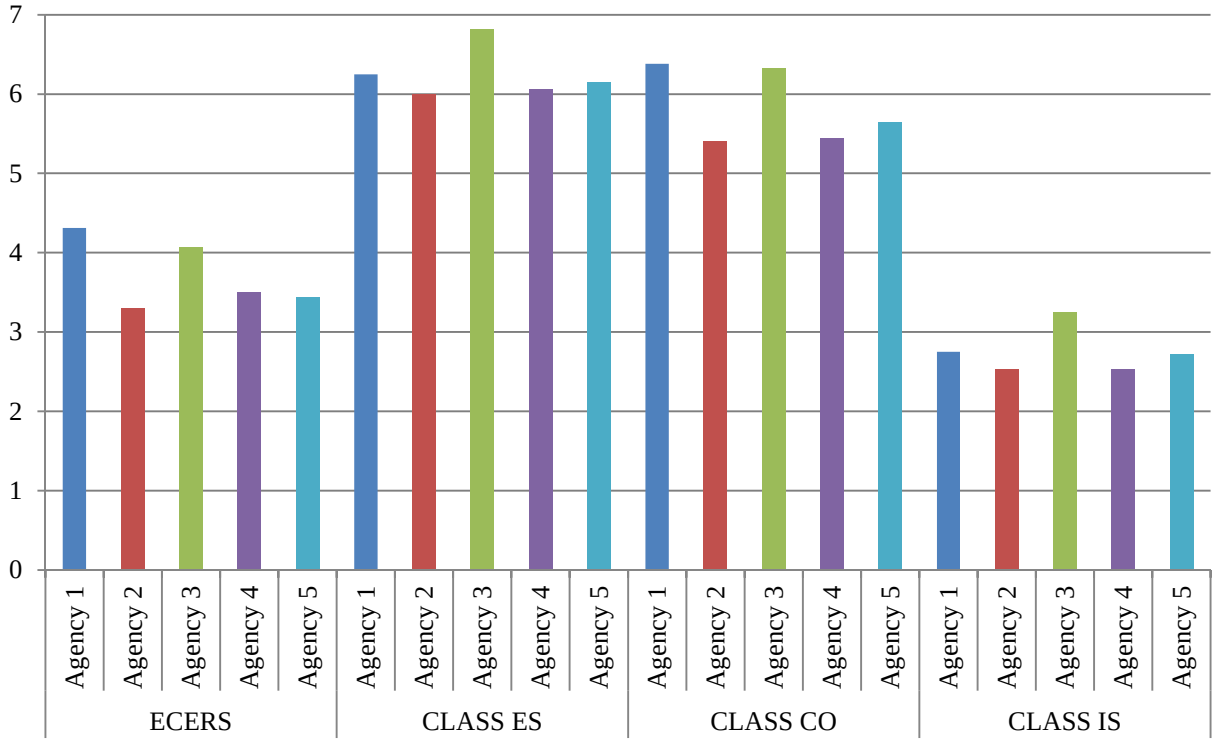
CLASS Dimensions and Domains	Mean	Minimum	Maximum
Emotional Support Domain	*6.14*	*4.88*	*6.81*
1. Positive Climate	5.80	4.25	7.00
2. Negative Climate*	6.86	5.75	7.00
3. Teacher Sensitivity	5.91	4.25	6.75
4. Regard for Student Perspectives	5.96	4.25	7.00
Classroom Organization Domain	*5.67*	*4.17*	*6.58*
5. Behavior Management	5.73	3.75	7.00
6. Productivity	6.05	4.50	7.00
7. Instructional Learning Formats	5.21	3.50	6.50
Instructional Support Domain	*2.65*	*1.50*	*4.25*
8. Concept Development	2.07	1.25	3.50
9. Quality of Feedback	2.61	1.50	4.25
10. Language Modeling	3.29	1.75	5.00

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

5. How does quality vary within SPP across children and providers?

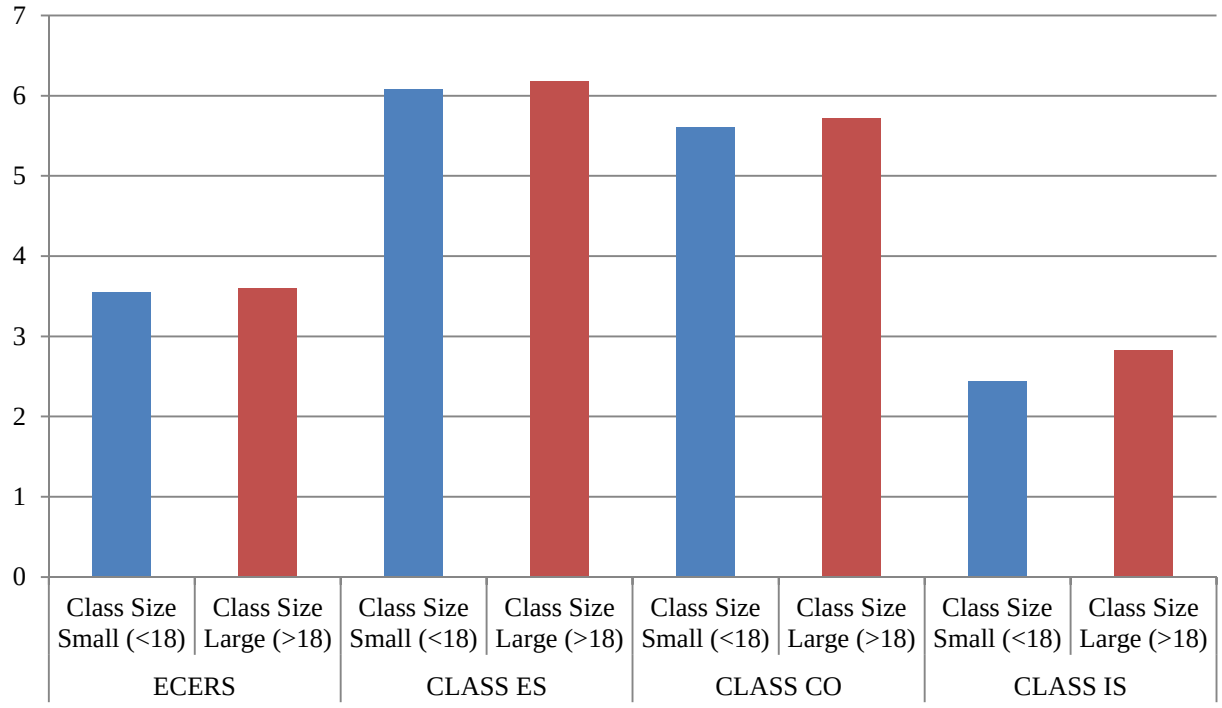
Figure 4 illustrates average classroom quality scores for ECERS and all three CLASS subscale scores across agencies. For the most part, score patterns are quite similar, with ECERS scores in the mid range for all agencies, CLASS ES & CO scores in the 5-7 range across all agencies, and CLASS IS scores in the 2-3 range with only a slight advantage in the later for Agency 3. Scores by Agency at the item level are reported in Appendix C, Tables C.1 and C.2.

Figure 4. ECERS and CLASS Domain scores by Agency



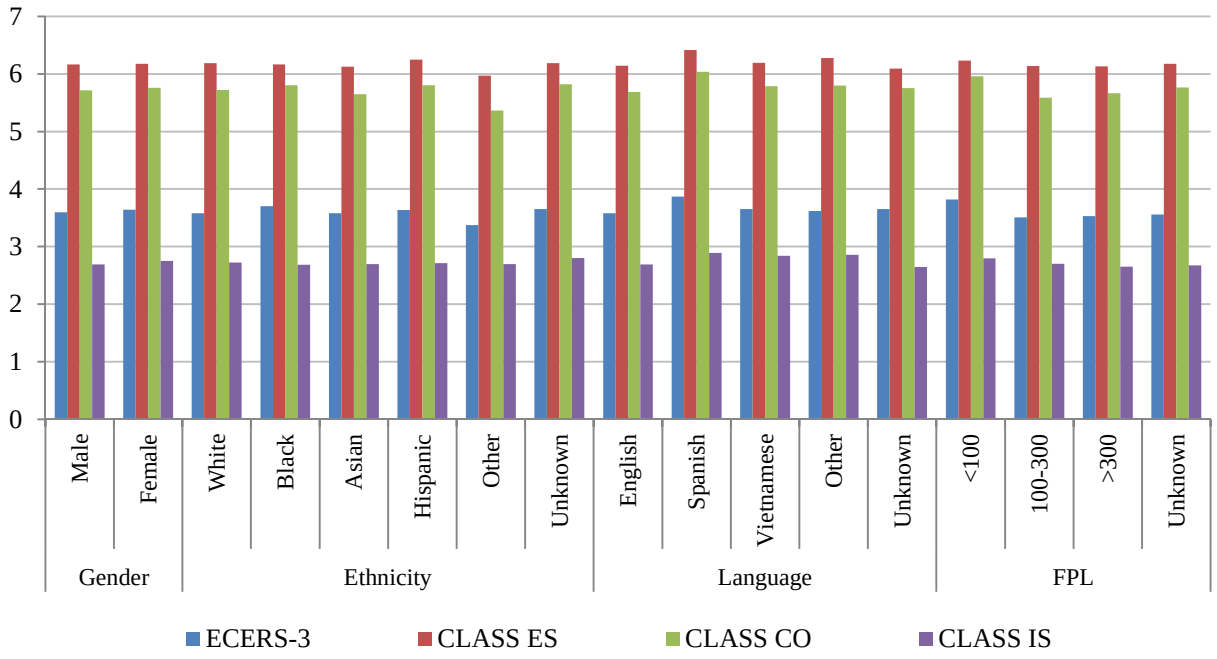
Similarly, Figure 5 illustrates ECERS-3 and CLASS domain scores for smaller (classrooms with 18 or less children) and larger (with more than 18 children) classrooms in the sample. Overall, classroom quality patterns are very close together regardless of class size, being these between 3 and 4 for ECERS, about 6 for CLASS ES, between 5 and 6 for CLASS CO and between 2 and 3 for CLASS IS. Test of statistical significances between groups showed no differences in scores between smaller and larger classrooms.

Figure 5. ECERS and CLASS Domain scores by Class Size



Using DEEL information on children's gender, ethnicity/race, language background and FPL, Figure 6 illustrates the quality of care all 234 enrolled children in SPP classrooms experience, averaged by their individual characteristics. No distinguishable patterns were observed of one group receiving better/lower quality than peers of any other particular group, with subpopulation groups aligned at the same levels of quality reported in the previous two figures.

Figure 6. ECERS and CLASS Domain scores by Child Characteristics



6. What activities do children engage in, and is there scope for their interests and active participation?

In part, this question was somewhat address when looking at how to interpret CLASS scores in pages 35-36. However, in addition to inquire into whether classrooms offered scope for children's interests and active participation, this section extracts specific indicators across ECERS-3 that expressly address interactions and the ways in which staff actively engage children. What is presented below is the frequency with which classrooms met these specific indicators. Indicators are graphed grouped by item.

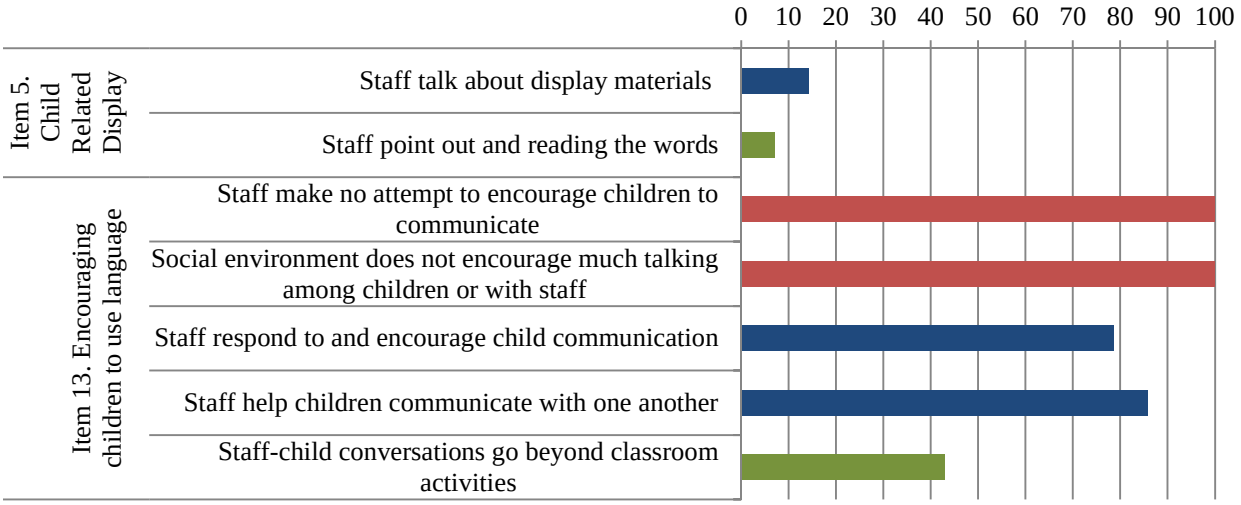
Indicators in the ECERS-3 are organized in 4 levels: inadequate, minimal, good and excellent. Taking this into account, indicators in red are scored positively even if the indicator is phrased negatively and represent the lowest level of quality on the ECERS-3 which falls under the anchor of "inadequate." Despite that these show that classrooms did not do these negative things, meeting these indicators means that classrooms are not "inadequate" or represent a score of higher than 1.00. Indicators met in the "minimal" category or scores of a 3.00 on the ECERS-3 are denoted by yellow bars, indicating that classrooms met indicators at a level of "minimal" on the tool. To represent indicators that constitute a score of a 5.00 or "good" rating, blue bars are used, and for a 7.00 or "excellent" green is used. To provide an initial look at problem areas that need improvement and could be targeted through the continuous quality improvement cycle, this system will help visually to see the percentage of classrooms meeting indicators relative to the observed engagement of children by staff within each of the items. While it is true that at the basic or "inadequate" level it is promising that 100 percent of classrooms are meeting requirements, special attention should be given to the blue and green bars which show that interactions needed to achieve these are occurring in classrooms less frequently. In some cases

not all items have indicators about engagement that range the full course (inadequate, minimal, good and excellent) so what is presented here is that which is captured by the tool on engagement regardless of the level.

Item 5 on the ECERS-3 relates to the child-related displays. Two indicators under this item capture whether staff talk about displayed materials and whether they incorporate them into their conversations as the intent is that displays are purposeful and relevant to children for use in instruction. This can be done by pointing out and naming objects depicted on a wall or asking children a question about pictures displayed about a recent class visitor and what they might remember. These types of activities are almost not present in SPP classrooms as can be seen by the blue and green bars in Figure 7.

Item 13, "encouraging children to use language," captures ways in which staff encourage children to use language throughout the day. Different indicators capture how much of this occurs and what types of conversations are encouraged. For the most part, these tend to be met on over 80 percent of the classrooms on the most basic indicators (denoted by red bars) showing that teachers do not ignore children and that generally there are attempts to communicate with children. Even at the "good" level or 5.00 score category (seen in blue), classrooms largely show that staff meet indicators that capture the ways that teachers engage children in conversation and encourage them to say more. At the 7.00 or "excellent" level (seen in green), practices that entail encouraging conversations that go beyond classroom activities (e.g. conversations about home and family life, about activities in the community, about feelings) occur less frequently. This indicator could potentially be incorporated into CQI activities with specific awareness around focusing on interactions that connect discussions across all routines of the day with children's home activities e.g. when children are eating, conversations about dinner times and practices with their families could be very natural.

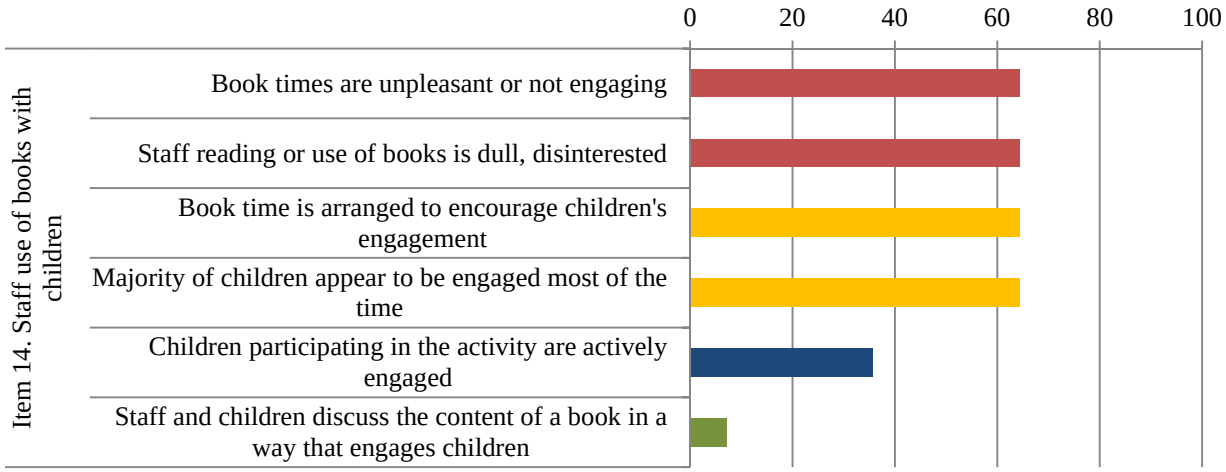
Figure 7. Indicators met on display and use of language



Item 14, "staff use of books with children," on the ECERS (illustrated in Figure 8) focuses on how books are incorporated into classroom activities, how engaging these activities are, and how books are discussed with children. Book reading, while theoretically a simple task, seems to be quite often a challenging one in that staff sometimes struggle to keep children interested and engaged. This seems to be the case for a portion of the SPP classrooms as what the indicator

analysis of this item shows is that even for the most basic indicators (coded by red and yellow bars) close to 20 percent of classrooms did not meet the requirements set forth. The higher levels of quality (blue and green bars) on this item were even less frequently met. While books are present, and children are generally engaged, this item seeks to capture that children are engaged and appear to be enjoying the book being read with few to no children looking away or showing behaviors that indicate that they are disinterested. Finally, in addition to attention to the quality of reading itself by the staff, the item also assesses at the highest level (excellent or green bars) whether conversations subsequent to reading about the book are engaging. Less than 20 percent of classrooms met this indicator. This is important as these discussions are the basis to gauge comprehension and to foster conversations to learn more about the topic presented in the book, and make connections to other reading. Given that discussions do not have pictures as does a book itself, careful attention must be given to the quality of these conversations to ensure that all children are benefiting.

Figure 8. Use of books

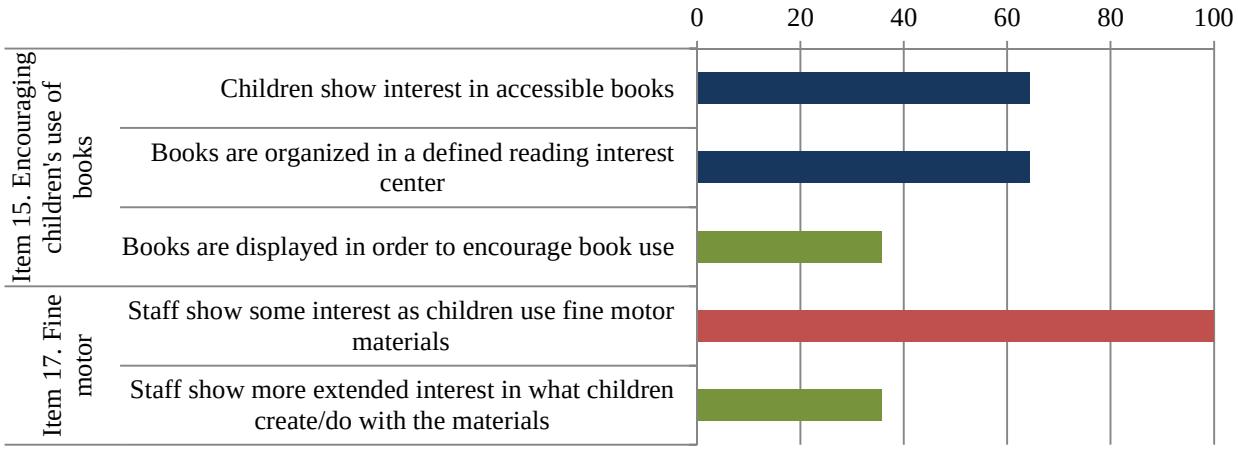


Item 15, “encouraging children’s use of books” (illustrated in Figure 9), is centered on inspiring the use of books by children themselves. Three indicators specifically attempt to capture children’s interest in books through the way that they are made accessible. In large part the item assesses the presence of a defined reading center, as well as whether books are displayed in meaningful ways that encourage their use. In addition, the item generally assesses both the condition/quality of books as well as the range of types of books for children to choose from. These indicators capture decisions on accessibility that encourage children’s use of book and are easy to implement by creating awareness on the part of teachers to be intentional about what books they give to students and how they present them. However, they were present in less than 2/3rd of classrooms and only 5 out of the 14 classrooms displayed books appropriately as can be seen by the blue and green bars in Figure 9. These include requirements that books are not crowded, with clear view of the covers, are in good condition, and that there is a wide selection of topics/genres.

Item 17, fine motor, assesses general quality of the presence of fine motor materials and also about how staff across engage children in using the materials. The data shows that although the staff did not ever ignore children engaging with fine motor materials, that only about a 1/3rd of the classrooms met indicators that observed staff having more extended conversations with

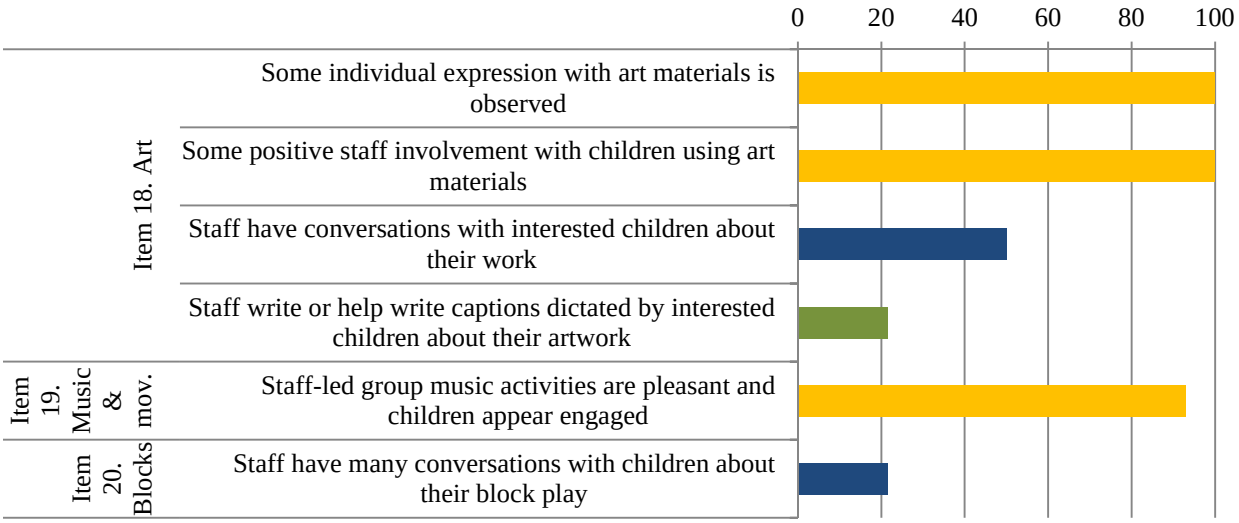
children about their use of the materials. These would include conversations about what children create y create with these materials, provide examples on how to use materials, or give children opportunities to select materials according to differing interests & difficulty based on their assessment of children's level of mastery. Meeting indicators of engagement in this category would simply mean more frequent and sustained use of language (e.g. question asking, object naming and participation) as children play and work with toys that are considered under the "fine motor" item.

Figure 9. Books and fine motor



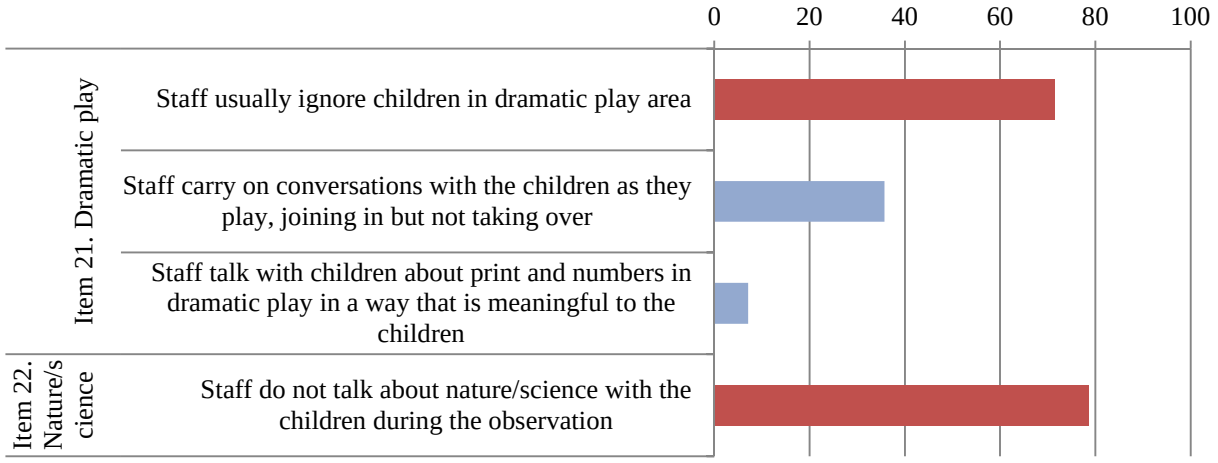
The following three set of items focused on Art, Music and Blocks (Figure 10). Across all three, the ECERS captures whether staff engages in conversations about the activities and/or whether children's engagement is evident. For item 18, Art, all classrooms showed evidence of allowing children to engage individually in art expression, and positive staff involvement using art materials. Most classrooms evidenced staff-led group activities were children appeared engaged. On the other hand, classrooms where staff engaged in conversations with children about their artwork were more infrequent. Even less evident were classrooms in which staff wrote or helped write captions dictated by children about their artwork. Similarly, for item 20, Blocks, conversations relative to block play (e.g. questions about what they will build/ are building, about shapes using, about structures, among others) were observed very infrequently despite that the indicator analyzed is at the level of "good." Due to the very open-ended nature of blocks, teachers may need resources, and coaching on how to engage meaningfully in block play by providing language and moving children along the developmental continuum of the stages of block play.

Figure 10. Art, music and blocks



In Figure 11, Item 21., Dramatic play includes indicators relative to how the teacher supports dramatic play by way of participating with children. At a high level, teachers would not only participate in conversations but also incorporate print and number concepts to add to potential learning experiences to the play. The data shows that while staff do not necessarily ignore children in dramatic play scenarios, there is little evidence that sustained conversations to encourage deeper thinking was only in evidence for a very few number of SPP classrooms. Item 22, Science/Nature, also shows that few classrooms (3 out of 14) evidenced no conversations on this topic on the day of the observation.

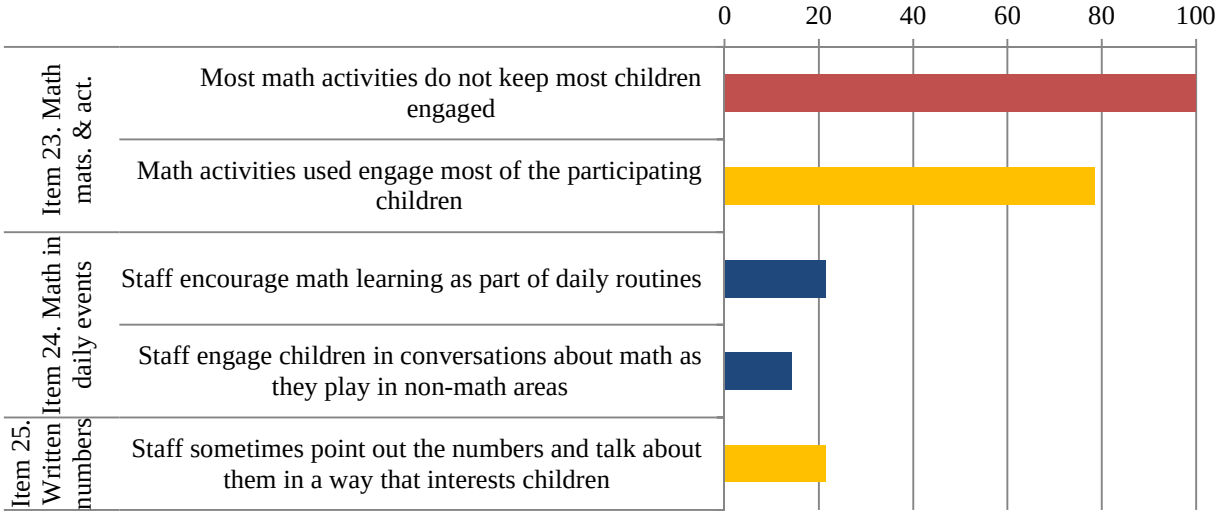
Figure 11. Dramatic play and nature/science



Items 23-25 focus on math and numbers, shown in Figure 12. A series of indicators in these capture whether staff make efforts to engage children in various ways. While it seems that activities are engaging in all classrooms, most classrooms did not show evidence of encouraging math learning in daily routines (examples provided in the ECERS include: explaining table-setting, naming rectangular and round tables when saying where to put plates and cups, and

counting to 20 while washing hands), engaging children in conversations about math in non-math areas and free play (examples included in the ECERS include: discussing using measuring cups to water plant, counting how many teacups are needed for dolls, and talking about how to measure feet in play shoe store). This is also the case for pointing out the numbers on materials and talking about these (e.g. on play money, a play cash register, or a play phone).

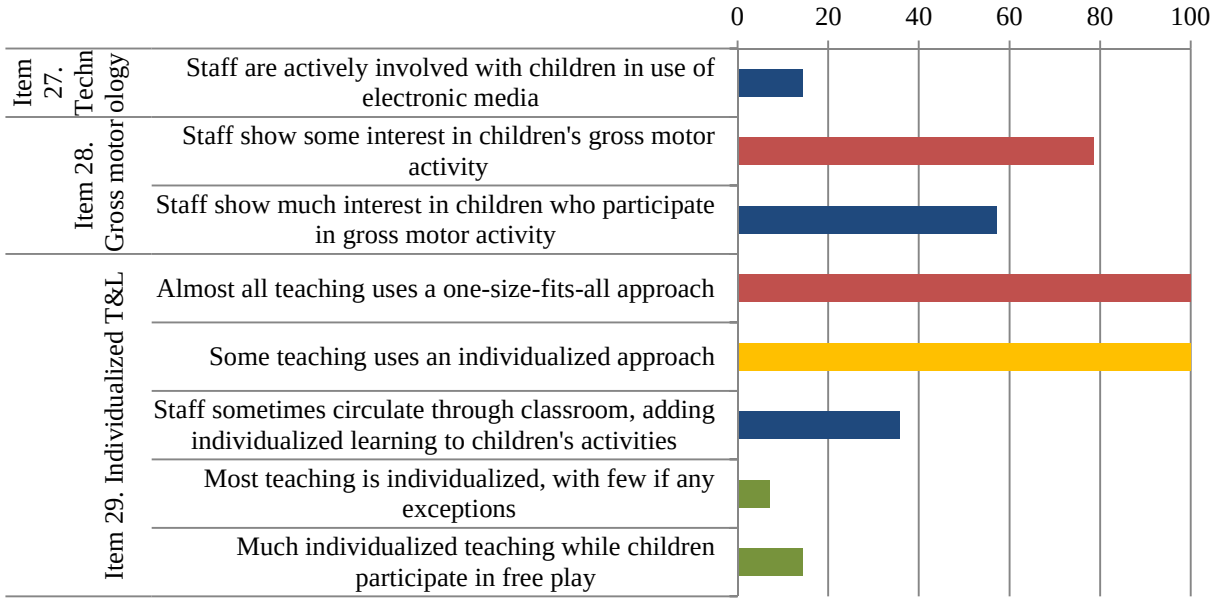
Figure 12. Math and numbers



The next sets of items is on technology, gross motor, and individualized teaching and learning, and are shown in Figure 13. Like before, the indicators highlighted are those that touch upon whether staff actively engages children in various activities. For technology, only in 2 classrooms were staff observed actively involving children in the use of electronic media. In terms of gross motor, in most classrooms staff show interest in children's gross motor activities and encouraged activities that got children moving. Item 29, "individualized teaching and learning," includes indicators relative to how teachers respond to the varied abilities, interests and needs of the children. To this end, the item seeks to capture the ways that staff interact with children informally with little use of directive or "one-size-fits-all" teaching styles that aim to group children into large groups. While generally indicators identified under this item are met across classrooms at a basic level (denoted by red and yellow bars), there are fewer instances where classrooms were observed fulfilling the requirements to receive scores of 5.00-7.00 ("good" and "excellent"). In particular, this is true of the final few indicators which can be seen in blue and green in Figure 13 where less than 40 percent met on the blue or "good" indicator and less than 20 percent of classrooms met on the green or "excellent." To meet these indicators, teachers must utilize the informal (unplanned) instructional times (free play) to circulate through the classroom to respond to children's interests as they play and work in the interest areas set up by the teacher in the classroom. For the highest scores which would be the result of meeting the final two indicators (denoted by green bars), this approach should be the primary vehicle of instruction, occurring most of the day and leaving whole-group, teacher directed times to a minimum. In addition, teaching in this way during free play times would mean that teachers are not just wandering from group to group, or child to child, but engaging deeply in response to children with observably intentional interactions. These interactions would require more than one

back and forth exchange between teachers and children and would incorporate language development with the development of conceptual knowledge about the topic, material or activity of focus. Boosts in these scores would result in higher scores on all of the items in the Language and Literacy subscale as well as the Activities subscale, which all examine engagement and language building.

Figure 13. Technology, gross motor and individualized teaching and learning



Summary

The first year of the SPP successfully enrolled a diverse group of children across ethnic/racial groups and from the target population of families of under 300 FPL. Children served were more ethnically diverse, although less linguistically diverse, than children in Seattle public schools. Children enrolled in SPP classrooms made modest gains in vocabulary, literacy, math and executive functions. Relative to gains expected just because they were older at post-test, children made larger than expected gains in language and letter knowledge and smaller than expected in math. Looking at gains while controlling for children and classroom characteristics simultaneously, attendance was consistently associated with higher outcomes. In addition, Bilingual children and those below the FPL had smaller gains than others. Attendance rates appear to be overall lower than should be expected. However, incomplete attendance data could be an important driver of this rate.

Quality was lower than desired for the ECERS-3 and for the Instructional Support domain of the CLASS. Quality was relatively good for the Emotional Support and Classroom organization domains of the CLASS. The report provides detailed information at the item level to allow SPP to identify indicators associated with higher and lower scores. In particular, programs are scoring fairly poorly with respect to engaging children in conversations openly, or in response to an activity they are involved in, or bringing concepts of math, science, or another topic into any free play activity.

Recommendations

Given the small numbers of classrooms and children and the lack of any no-treatment control group, results from the first year should be interpreted cautiously. They do provide a baseline against which to judge future results, which should be expected to improve over time, especially for programs that stay in SPP. In addition, it is possible to make several clear recommendations.

First, SPP needs to continue to strengthen the collection of information on children, whether through teachers, or observations. Children's background characteristics are associated with differences in gains and this needs to be carefully monitored. Children for whom data was not obtained also seemed to differ in some important respects in their gains, which means that it is important to learn more about them.

Second, attendance is associated with gains and is lower than it should be. Again, we caution against interpreting this as causal—perhaps children who have greater barriers to attendance suffer from more difficulties at home and in the community or have poorer health which reduces their gains. However, it is clear that children cannot benefit if they do not attend. In addition, there was incomplete information on attendance and given its importance this should be rectified.

Third, program quality needs to be raised by the SPP continuous improvement process. These data provide a baseline and offer a first set of indications of where to focus on guidance and professional development including the efforts of coaches and others generally, though the coaching to be effective must be individualized. Clearly, results were more disappointing for math than for other domains indicating the need for an overall increase in focus on improving math teaching. In addition, we suggest that teachers would benefit from a structured approach to supporting bilingual children that could scaffold for them this difficult activity. In addition, we make three more fine gained suggestions based on our detailed look at quality assessments: (1) intentional integration of math across all areas and centers and through the day, (2) more intentional engagement with children when in free play and across activities, intentionally integrating concepts and providing feedback, and (3) a focus in language modeling which can take place during free play, but should be individualized.

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.

Barnett, W.S. (2013) Expanding Access to Quality Pre-K is Sound Public Policy. New Brunswick, NJ: National Institute for Early Education Research.

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.

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.

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. http://www.qrisnetwork.org/sites/all/files/conference-session/resources/703ECERS3_0.pdf

Lamy, C.E., Frede, E. Seplocha, H., Strasser, J., Jambunathan, S., Juncker, J. A., Ferrar, H. Wiley, L., & Wolock, E. (2004) Inch by Inch, Row by Row Gonna Make This Garden Grow. Classroom Quality and Language Skills in the Abbott Preschool Program. Year One Report, 2002-2003 Early Learning Improvement Consortium. New Jersey: Early

Learning Improvement Consortium Available at http://www.state.nj.us/education/ece/research/inch.pdf

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

NIEER (2014) New Jersey Abbott Preschool Quality Evaluation Study. Summary Report. New Brunswick: Author.

NIEER (2016) New Jersey Abbott Preschool Quality Evaluation Study. Summary Report. New Brunswick: Author.

NYC Department of Education. (2015) Pre-K Program Assessments Classroom Assessment Scoring System (CLASS) and Early Childhood Environmental Rating Scale – Revised (ECERS-R) Release. New York: Author. Available at http://schools.nyc.gov/NR/rdonlyres/A8A27BFE-7C58-4F03-8EB7-B90E01BA3D0D/0/CLASSandECERSRReleaseDeckFinal.pdf

NYC Department of Education. (December 18, 2015) Mayor de Blasio Announces Over 68,500 Students Enrolled in Pre-K for All. Press office. New York; Author. Avaible at http://www1.nyc.gov/office-of-the-mayor/news/954-15/mayor-de-blasio-over-68-500-students-enrolled-pre-k-all

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

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.

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. Raw Score Analyses.
Appendix B. Sensitivity Analyses.
Appendix C. ECERS-3 and CLASS scores by Agency and Class Size. Item level.
Appendix D. Family Survey.

Appendix A. Raw Score Tables.

Table A.1. Receptive vocabulary raw score means and gains by child characteristics

		PPVT Raw 2015 Fall		PPVT Raw 2016 Spring		PPVT Raw Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=189)		71.45	24.59	83.68	25.53	12.23	15.00
Gender	Male (N=92)	70.08	24.71	82.59	23.62	12.51	14.26
	Female (N=97)	72.75	24.54	84.71	27.29	11.96	15.75
Age	Three Year Cohort (N=37)	51.65	16.92	65.30	23.79	13.65	14.42
	Four Year Cohort (N=152)	76.27	23.77	88.15	23.94	11.88	15.17
Ethnicity	White (N=55)	85.27	20.87	97.55	19.18	12.27	17.21
	Black (N=47)	65.60	21.76	76.83	26.35	11.23	13.89
	Asian (N=24)	62.21	24.15	77.17	26.20	14.96	16.79
	Hispanic (N=22)	66.05	23.38	78.68	21.94	12.64	12.94
	Other (N=13)	70.46	28.03	80.46	26.90	10.00	18.17
	Unknown (N=28)	66.75	26.27	78.93	27.90	12.18	10.89
Language	English (N=127)	77.36	23.35	89.18	23.44	11.82	16.51
	Spanish (N=13)	55.46	19.05	70.62	22.66	15.15	12.91
	Vietnamese (N=5)	46.80	10.87	58.20	18.77	11.40	8.62
	Other (N=19)	55.42	20.13	67.68	26.16	12.26	13.94
	Unknown (N=25)	66.84	26.36	79.76	28.00	12.92	9.16
FPL	<100 (N=55)	62.87	20.64	76.98	25.25	14.11	16.44
	100-300 (N=80)	70.86	24.77	79.94	24.64	9.08	14.18
	>300 (N=40)	85.50	22.10	99.70	18.96	14.20	15.78
	Unknown (N=14)	68.36	29.14	85.57	30.59	17.21	7.04

Table A.2. Receptive vocabulary raw score means and gains by center characteristics

		PPVT Raw 2015 Fall		PPVT Raw 2016 Spring		PPVT Raw Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=189)		71.45	24.59	83.68	25.53	12.23	15.00
Agency	Agency 1 (N=41)	64.27	22.72	77.29	29.58	13.02	15.69
	Agency 2 (N=74)	74.12	24.03	88.00	23.31	13.88	13.55
	Agency 3 (N=15)	82.60	24.40	87.47	27.71	4.87	19.20
	Agency 4 (N=23)	84.22	24.82	95.22	19.17	11.00	13.19
	Agency 5 (N=36)	61.33	22.09	73.11	23.08	11.78	16.04
Class Size	18 or Less (N=87)	72.30	22.94	84.70	23.06	12.40	12.93
	More than 18 (N=102)	70.73	26.01	82.80	27.54	12.08	16.63
Curriculum	Creative Curriculum (N=64)	71.44	25.22	83.73	27.55	12.30	14.76
	HighScope (N=125)	71.46	24.37	83.65	24.54	12.19	15.18
ECERS	Less then 3 (N=15)	77.67	29.26	87.07	26.46	9.40	14.65
	3 or More (N=174)	70.91	24.17	83.39	25.50	12.47	15.05
CLASS ES	Less than 5,5 (N=28)	73.43	26.18	80.64	25.13	7.21	16.36
	5.5 or More (N=161)	71.11	24.37	84.20	25.63	13.10	14.63
CLASS CO	Less than 5.5 (N=51)	74.04	25.71	81.73	22.75	7.69	15.42
	5.5 or More (N=138)	70.49	24.19	84.40	26.52	13.91	14.55
CLASS IS	Less than 3 (N=116)	70.97	24.38	82.06	24.12	11.09	13.57
	3 or More (N=73)	72.22	25.08	86.25	27.59	14.03	16.97

Table A.3. Literacy raw score means and gains by child characteristics

		WJ-LW Raw 2015 Fall		WJ-LW Raw 2016 Spring		WJ-LW Raw Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=186)		8.39	5.55	11.15	6.62	2.76	3.73
Gender	Male (N=92)	7.78	4.09	10.55	5.32	2.77	3.91
	Female (N=94)	8.99	6.64	11.73	7.67	2.74	3.56
Age	Three Year Cohort (N=36)	4.33	3.02	6.78	4.08	2.44	3.93
	Four Year Cohort (N=150)	9.37	5.58	12.20	6.69	2.83	3.69
Ethnicity	White (N=56)	8.32	4.46	11.89	5.83	3.57	4.28
	Black (N=47)	8.34	6.67	11.00	8.38	2.66	3.70
	Asian (N=25)	9.44	4.10	11.52	4.02	2.08	2.66
	Hispanic (N=17)	5.29	3.48	7.18	4.57	1.88	2.83
	Other (N=13)	8.92	3.45	11.38	5.06	2.46	2.79
	Unknown (N=28)	9.32	7.67	11.89	7.84	2.57	4.21
Language	English (N=129)	7.95	4.96	10.90	6.21	2.95	3.70
	Spanish (N=8)	6.50	3.07	8.63	4.87	2.13	3.44
	Vietnamese (N=5)	8.80	4.32	9.40	4.16	0.60	1.82
	Other (N=19)	10.84	5.96	13.58	7.78	2.74	3.51
	Unknown (N=25)	9.36	8.07	11.76	8.29	2.40	4.43
FPL	<100 (N=55)	7.45	4.73	10.20	6.46	2.75	3.45
	100-300 (N=76)	8.49	5.47	11.05	6.71	2.57	3.48
	>300 (N=41)	9.17	4.63	12.49	5.74	3.32	4.35
	Unknown (N=14)	9.29	9.96	11.50	8.93	2.21	4.32

Table A.4. Literacy raw score means and gains by center characteristics

		WJLW Raw 2015 Fall		WJLW Raw 2016 Spring		WJLW Raw Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=186)		8.39	5.55	11.15	6.62	2.76	3.73
Agency	Agency 1 (N=41)	7.37	4.24	9.24	5.85	1.88	3.49
	Agency 2 (N=75)	8.35	5.67	11.73	7.40	3.39	4.05
	Agency 3 (N=14)	9.36	3.86	13.21	4.81	3.86	4.52
	Agency 4 (N=23)	11.22	8.44	13.43	7.74	2.22	2.63
	Agency 5 (N=33)	7.39	4.19	9.73	4.52	2.33	3.36
Class Size	18 or Less (N=87)	8.49	6.30	10.87	6.97	2.38	3.37
	More than 18 (N=99)	8.30	4.82	11.39	6.32	3.09	4.00
Curriculum	Creative Curriculum (N=64)	8.75	6.30	10.75	6.84	2.00	3.19
	HighScope (N=122)	8.20	5.12	11.36	6.52	3.16	3.94
ECERS	Less then 3 (N=15)	7.00	4.34	9.00	4.47	2.00	3.07
	3 or More (N=171)	8.51	5.63	11.34	6.75	2.82	3.78
CLASS ES	Less than 5,5 (N=28)	6.04	4.36	8.32	4.12	2.29	3.15
	5.5 or More (N=158)	8.81	5.64	11.65	6.86	2.84	3.82
CLASS CO	Less than 5.5 (N=51)	7.49	4.90	9.80	4.90	2.31	2.85
	5.5 or More (N=135)	8.73	5.75	11.66	7.11	2.93	4.00
CLASS IS	Less than 3 (N=114)	7.84	5.87	10.04	6.50	2.19	3.41
	3 or More (N=72)	9.26	4.90	12.92	6.47	3.65	4.04

Table A.5. Math raw score means and gains by child characteristics

		WJ-AP Raw 2015 Fall		WJ-AP Raw 2016 Spring		WJ-AP Raw Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=186)		11.93	5.04	13.64	4.71	1.71	3.29
Gender	Male (N=92)	11.37	5.20	13.42	4.76	2.05	3.65
	Female (N=94)	12.48	4.85	13.85	4.68	1.37	2.88
Age	Three Year Cohort (N=36)	7.42	3.56	9.42	4.27	2.00	2.82
	Four Year Cohort (N=150)	13.01	4.74	14.65	4.24	1.64	3.40
Ethnicity	White (N=56)	13.95	4.63	15.46	4.38	1.52	3.12
	Black (N=47)	10.15	5.51	12.15	4.57	2.00	3.34
	Asian (N=25)	11.16	4.43	14.00	4.77	2.84	4.02
	Hispanic (N=17)	12.29	3.62	15.24	2.99	2.94	3.54
	Other (N=13)	12.31	5.50	12.62	4.23	0.31	2.02
	Unknown (N=28)	11.18	5.00	11.68	5.21	0.50	2.65
Language	English (N=129)	12.42	5.01	14.12	4.38	1.71	3.16
	Spanish (N=8)	12.25	4.65	15.75	3.20	3.50	3.42
	Vietnamese (N=5)	8.20	3.56	10.20	4.66	2.00	3.94
	Other (N=19)	10.89	5.73	13.00	6.09	2.11	4.54
	Unknown (N=25)	10.84	4.77	11.64	4.99	0.80	2.57
FPL	<100 (N=55)	10.35	4.77	12.62	4.33	2.27	3.29
	100-300 (N=76)	11.74	5.26	13.20	5.14	1.46	3.55
	>300 (N=41)	14.49	4.04	16.00	3.47	1.51	2.83
	Unknown (N=14)	11.71	5.08	13.14	5.05	1.43	3.13

Table A.6. Math raw score means and gains by center characteristics

		WJAP Raw 2015 Fall		WJAP Raw 2016 Spring		WJAP Raw Gains	
		Mean	SD	Mean	SD	Mean	SD
Total (N=186)		11.93	5.04	13.64	4.71	1.71	3.29
Agency	Agency 1 (N=41)	10.37	4.67	11.80	4.72	1.44	2.64
	Agency 2 (N=75)	12.19	5.00	13.80	4.67	1.61	3.53
	Agency 3 (N=14)	13.93	4.73	15.50	4.54	1.57	4.20
	Agency 4 (N=23)	14.61	4.43	15.96	4.25	1.35	2.71
	Agency 5 (N=33)	10.58	5.23	13.15	4.40	2.58	3.44
Class Size	18 or Less (N=87)	11.76	5.00	13.49	4.59	1.74	3.33
	More than 18 (N=99)	12.08	5.10	13.77	4.84	1.69	3.27
Curriculum	Creative Curriculum (N=64)	11.89	4.99	13.30	4.95	1.41	2.64
	HighScope (N=122)	11.95	5.09	13.82	4.60	1.87	3.58
ECERS	Less than 3 (N=15)	12.53	5.53	13.33	5.22	0.80	2.78
	3 or More (N=171)	11.88	5.01	13.67	4.68	1.79	3.33
CLASS ES	Less than 5,5 (N=28)	11.11	5.51	12.18	4.79	1.07	2.88
	5.5 or More (N=158)	12.08	4.96	13.90	4.67	1.82	3.35
CLASS CO	Less than 5.5 (N=51)	11.45	5.16	12.55	4.64	1.10	2.97
	5.5 or More (N=135)	12.11	5.01	14.05	4.69	1.94	3.39
CLASS IS	Less than 3 (N=114)	11.46	5.27	13.18	4.62	1.72	3.36
	3 or More (N=72)	12.68	4.59	14.38	4.80	1.69	3.19

Appendix B. Sensitivity Analyses.

Table B.1. Multivariate analyses of children's 2015-16 gains in Raw and Standard scores in relation to child and site or classroom characteristics (if tested in Spanish and English, Spanish score used) with ECERS-3

	Rec. Vocabulary Raw (PPVT/TVIP)	Rec. Vocabulary Standard (PPVT/TVIP)	Literacy Raw (WJ/WM-LW)	Literacy Standard (WJ/WM-LW)	Math Raw (WJ/WM-AP)	Math Standard (WJ/WM-AP)	Executive Function	
							DCCS	PT
Pre Test	**0.694*****	**0.620*****	**0.951*****	**0.703*****	**0.640*****	**0.598*****	**0.015***	**0.469*****
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.01)	(0.06)
Attendance	**0.255***	**0.173***	0.017	0.066	**0.062****	**0.157****	0.004	0.046
	(0.11)	(0.08)	(0.03)	(0.07)	(0.02)	(0.06)	(0.00)	(0.04)
Missing Attendance	**20.110***	**14.061***	3.541	11.006	**4.930***	**12.777***	0.528	**6.352***
	(9.75)	(6.74)	(2.49)	(6.16)	(1.97)	(5.29)	(0.38)	(3.05)
Days Between Tests	-0.143	-0.134	-0.053	**-0.192****	-0.009	-0.080	-0.001	0.013
	(0.10)	(0.07)	(0.03)	(0.07)	(0.02)	(0.06)	(0.00)	(0.03)
Female	0.102	-0.214	-0.035	0.286	-0.188	-0.856	0.057	-0.434
	(1.97)	(1.36)	(0.51)	(1.27)	(0.40)	(1.08)	(0.08)	(0.63)
Age in Months	**6.862****	0.966	0.548	-2.019	0.541	**-4.565*****	**0.418*****	**1.556***
	(2.27)	(1.32)	(0.56)	(1.26)	(0.47)	(1.08)	(0.07)	(0.64)
Black	-4.730	-3.663	0.251	0.574	-0.319	-1.524	**-0.274***	-1.803
	(3.21)	(2.22)	(0.83)	(2.04)	(0.65)	(1.76)	(0.12)	(1.02)
Asian	-4.015	-3.097	-1.173	-2.903	0.586	1.741	-0.170	0.667
	(3.74)	(2.58)	(0.93)	(2.30)	(0.74)	(1.99)	(0.14)	(1.16)
Hispanic	0.120	0.006	-0.765	-3.605	1.596	3.468	0.062	1.005
	(4.02)	(2.78)	(1.12)	(2.77)	(0.86)	(2.31)	(0.16)	(1.26)
Other Race	-7.775	-4.489	-1.117	-3.285	-1.503	-4.164	-0.065	-0.528
	(4.31)	(2.98)	(1.10)	(2.72)	(0.87)	(2.33)	(0.16)	(1.36)
Missing Race	-10.569	-7.742	-0.166	-0.984	**-2.727***	**-7.537***	-0.383	-1.889
	(6.54)	(4.51)	(1.71)	(4.21)	(1.33)	(3.59)	(0.25)	(2.13)
Bilingual	**-7.079***	**-4.911***	-0.618	-1.313	-0.428	-1.278	-0.144	-1.652
	(3.07)	(2.12)	(0.77)	(1.91)	(0.61)	(1.63)	(0.11)	(0.92)
Missing Language	3.576	2.909	0.225	1.586	1.237	3.511	0.296	1.641
	(6.53)	(4.51)	(1.72)	(4.25)	(1.34)	(3.62)	(0.26)	(2.12)
FPL $\leq$100	-1.992	-1.682	-0.214	-1.300	0.078	0.100	-0.181	-0.594
	(3.32)	(2.29)	(0.84)	(2.09)	(0.67)	(1.80)	(0.13)	(1.04)
FPL 100-300	**-6.578***	**-4.523***	-0.602	-1.811	-0.800	-1.657	**-0.224***	**-1.368**
	(2.82)	(1.95)	(0.72)	(1.79)	(0.57)	(1.54)	(0.11)	(0.90)
Missing FPL	4.611	2.081	-2.911	-9.298	0.476	0.144	-0.398	**-5.404***

	(7.51)	(5.18)	(1.93)	(4.77)	(1.51)	(4.07)	(0.29)	(2.35)
Agency 2	7.240	4.901	**2.723***	4.672	1.827	4.091	-0.021	-1.711
	(5.13)	(3.54)	(1.37)	(3.39)	(1.05)	(2.82)	(0.21)	(1.63)
Agency 3	-3.918	-3.486	2.479	5.402	1.600	3.563	0.112	-0.083
	(4.96)	(3.42)	(1.29)	(3.19)	(1.00)	(2.68)	(0.20)	(1.56)
Agency 4	8.647	6.462	3.149	**9.290***	1.482	3.290	**0.528***	-0.098
	(6.65)	(4.59)	(1.76)	(4.32)	(1.35)	(3.64)	(0.27)	(2.10)
Agency 5	-2.585	-1.638	0.812	0.173	1.245	2.560	-0.223	-1.715
	(5.97)	(4.12)	(1.60)	(3.95)	(1.21)	(3.25)	(0.24)	(1.89)
Class Size	1.070	0.847	**0.403***	**1.342****	0.014	0.119	**0.063***	0.102
	(0.76)	(0.53)	(0.20)	(0.48)	(0.15)	(0.41)	(0.03)	(0.24)
ECERS	6.098	4.089	0.729	-0.228	1.379	2.808	-0.114	-1.123
	(4.36)	(3.01)	(1.16)	(2.87)	(0.88)	(2.37)	(0.18)	(1.38)
N	189	189	191	191	193	193	193	193

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are Males, White, English, FPL 300+, and Agency 1.Other controls are test type for children tested in Spanish and an interaction between test type and age to align the English and Spanish tests between them.

Table B.2. Multivariate analyses of children's 2015-16 gains in Raw and Standard scores in relation to child and site or classroom characteristics (if tested in Spanish and English, Spanish score used) with CLASS domains

	Rec. Vocabulary Raw **(PPVT/TVIP)**	Rec. Vocabulary Standard **(PPVT/TVIP)**	Literacy Raw **(WJ/WM-LW)**	Literacy Standard **(WJ/WM-LW)**	Math Raw **(WJ/WM-AP)**	Math Standard **(WJ/WM-AP)**	Executive Function	
							DCCS	PT
Pre Test	**0.694*****	**0.617*****	**0.941*****	**0.695*****	**0.639*****	**0.598*****	**0.013***	**0.467*****
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.01)	(0.06)
Attendance	**0.232***	**0.159***	0.019	0.078	**0.059***	**0.152***	0.003	0.043
	(0.11)	(0.08)	(0.03)	(0.07)	(0.02)	(0.06)	(0.00)	(0.04)
Missing Attendance	17.511	12.580	3.779	12.188	**4.613***	**12.362***	0.363	5.990
	(9.89)	(6.83)	(2.53)	(6.27)	(2.01)	(5.38)	(0.39)	(3.15)
Days Between Tests	-0.193	**-0.163***	-0.045	**-0.152***	-0.015	-0.084	-0.002	0.015
	(0.10)	(0.07)	(0.03)	(0.07)	(0.02)	(0.06)	(0.00)	(0.03)
Female	0.097	-0.243	-0.106	0.066	-0.210	-0.954	0.062	-0.436
	(1.94)	(1.34)	(0.51)	(1.26)	(0.40)	(1.08)	(0.08)	(0.63)
Age in Months	**6.538****	0.726	0.462	-2.362	0.472	**-4.796*****	**0.419*****	**1.571***
	(2.26)	(1.31)	(0.56)	(1.26)	(0.47)	(1.07)	(0.07)	(0.65)
Black	-4.686	-3.589	0.341	0.772	-0.298	-1.466	**-0.299***	-1.897
	(3.16)	(2.18)	(0.82)	(2.02)	(0.65)	(1.75)	(0.12)	(1.02)

Asian	-4.315	-3.271	-1.076	-2.556	0.568	1.752	-0.188	0.640
	(3.68)	(2.55)	(0.92)	(2.29)	(0.74)	(1.98)	(0.14)	(1.16)
Hispanic	-0.074	-0.088	-0.773	-3.535	1.562	3.425	0.025	0.897
	(3.96)	(2.73)	(1.11)	(2.74)	(0.86)	(2.30)	(0.16)	(1.26)
Other Race	-6.993	-3.907	-0.891	-2.541	-1.445	-3.921	-0.044	-0.393
	(4.25)	(2.93)	(1.09)	(2.71)	(0.86)	(2.32)	(0.16)	(1.37)
Missing Race	-11.331	-8.273	-0.305	-1.355	**-2.828***	**-7.846***	-0.426	-2.032
	(6.42)	(4.43)	(1.69)	(4.17)	(1.32)	(3.56)	(0.25)	(2.13)
Bilingual	**-7.422***	**-5.168***	-0.654	-1.402	-0.487	-1.429	-0.170	-1.758
	(3.02)	(2.09)	(0.76)	(1.89)	(0.60)	(1.62)	(0.11)	(0.92)
Missing Language	5.467	4.215	0.416	2.003	1.407	3.910	0.363	1.870
	(6.46)	(4.46)	(1.72)	(4.24)	(1.34)	(3.61)	(0.26)	(2.14)
FPL <100	-1.272	-1.210	-0.197	-1.278	0.127	0.184	-0.152	-0.513
	(3.29)	(2.27)	(0.84)	(2.08)	(0.67)	(1.80)	(0.13)	(1.05)
FPL 100-300	**-5.699***	**-3.916***	-0.498	-1.491	-0.745	-1.488	-0.177	-1.198
	(2.79)	(1.92)	(0.72)	(1.79)	(0.57)	(1.54)	(0.11)	(0.90)
Missing FPL	5.137	2.380	-2.861	**-9.243***	0.555	0.342	-0.360	**-5.370***
	(7.37)	(5.09)	(1.91)	(4.72)	(1.50)	(4.04)	(0.29)	(2.35)
Agency 2	8.168	5.932	**2.972***	**7.323***	1.093	2.877	0.321	0.061
	(4.62)	(3.19)	(1.21)	(2.99)	(0.95)	(2.55)	(0.18)	(1.52)
Agency 3	-6.101	-4.574	2.396	6.490	0.796	1.876	0.152	0.445
	(5.61)	(3.88)	(1.44)	(3.57)	(1.13)	(3.04)	(0.22)	(1.79)
Agency 4	**21.724***	**15.758***	3.854	11.552	1.980	4.223	**1.266****	2.464
	(10.87)	(7.51)	(2.84)	(6.99)	(2.24)	(6.03)	(0.42)	(3.52)
Agency 5	-5.215	-3.079	0.845	2.586	0.167	0.668	-0.009	-0.271
	(4.43)	(3.05)	(1.17)	(2.89)	(0.91)	(2.45)	(0.17)	(1.48)
Class Size	**2.963***	**2.114***	0.418	1.136	0.192	0.399	**0.133****	0.260
	(1.21)	(0.84)	(0.32)	(0.78)	(0.25)	(0.68)	(0.05)	(0.40)
CLASS_ES	4.614	2.340	-0.312	-2.551	1.087	2.077	0.131	0.026
	(5.66)	(3.90)	(1.49)	(3.68)	(1.16)	(3.10)	(0.22)	(1.84)
CLASS_CO	6.633	5.032	1.046	3.063	0.398	1.118	0.246	0.811
	(4.40)	(3.04)	(1.16)	(2.86)	(0.91)	(2.44)	(0.17)	(1.45)
CLASS_IS	-6.855	-4.427	0.196	1.192	-0.540	-0.770	**-0.346***	-0.948
	(4.41)	(3.04)	(1.15)	(2.84)	(0.90)	(2.42)	(0.17)	(1.43)
N	189	189	191	191	193	193	193	193

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are Males, White, English, FPL 300+, and Agency 1. Other controls are test type for children tested in Spanish and an interaction between test type and age to align the English and Spanish tests between them.

Table B.3. Multivariate analyses of children's 2015-16 gains in Raw and Standard scores in relation to child and site or classroom characteristics (if tested in Spanish and English, Spanish score used, with an ECERS-3 threshold

	Rec. Vocabulary Raw (PPVT/TVIP)	Rec. Vocabulary Standard (PPVT/TVIP)	Literacy Raw (WJ/WM-LW)	Literacy Standard (WJ/WM-LW)	Math Raw (WJ/WM-AP)	Math Standard (WJ/WM-AP)	Executive Function	
							DCCS	PT
Pre Test	**0.690*****	**0.616*****	**0.942*****	**0.693*****	**0.644*****	**0.604*****	**0.014***	**0.465*****
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.01)	(0.06)
Attendance	**0.268***	**0.181***	0.019	0.072	**0.063****	**0.159****	0.004	0.047
	(0.11)	(0.08)	(0.03)	(0.07)	(0.02)	(0.06)	(0.00)	(0.04)
Missing Attendance	**20.692***	**14.485***	3.498	10.961	**4.844***	**12.583***	0.530	**6.378***
	(9.74)	(6.72)	(2.46)	(6.07)	(1.97)	(5.27)	(0.38)	(3.06)
Days Between Tests	-0.190	**-0.166***	**-0.058***	**-0.187****	-0.019	-0.101	-0.000	0.021
	(0.10)	(0.07)	(0.03)	(0.06)	(0.02)	(0.05)	(0.00)	(0.03)
Female	-0.045	-0.322	-0.091	0.123	-0.220	-0.974	0.053	-0.458
	(1.96)	(1.36)	(0.51)	(1.25)	(0.40)	(1.08)	(0.08)	(0.63)
Age in Months	**6.947****	0.963	0.555	-2.180	0.545	**-4.550*****	**0.413*****	**1.538***
	(2.25)	(1.32)	(0.55)	(1.24)	(0.47)	(1.07)	(0.07)	(0.65)
Black	-4.281	-3.361	0.309	0.551	-0.202	-1.275	**-0.282***	-1.889
	(3.19)	(2.20)	(0.81)	(2.00)	(0.65)	(1.74)	(0.12)	(1.02)
Asian	-4.085	-3.151	-1.102	-2.701	0.637	1.887	-0.170	0.659
	(3.73)	(2.58)	(0.92)	(2.27)	(0.74)	(1.98)	(0.14)	(1.16)
Hispanic	0.430	0.218	-0.856	-3.863	1.582	3.411	0.056	0.947
	(4.01)	(2.76)	(1.11)	(2.73)	(0.86)	(2.31)	(0.16)	(1.26)
Other Race	-8.156	-4.740	-1.126	-3.121	-1.574	-4.281	-0.059	-0.448
	(4.28)	(2.95)	(1.08)	(2.68)	(0.87)	(2.31)	(0.16)	(1.36)
Missing Race	-11.093	-8.139	-0.428	-1.856	**-2.768***	**-7.804***	-0.406	-2.037
	(6.54)	(4.51)	(1.69)	(4.16)	(1.34)	(3.58)	(0.26)	(2.14)
Bilingual	**-6.917***	**-4.803***	-0.658	-1.492	-0.414	-1.284	-0.151	-1.724
	(3.06)	(2.11)	(0.76)	(1.88)	(0.61)	(1.62)	(0.11)	(0.92)
Missing Language	3.843	3.106	0.334	1.952	1.247	3.599	0.301	1.693
	(6.52)	(4.50)	(1.70)	(4.19)	(1.35)	(3.60)	(0.26)	(2.13)
FPL <100	-1.582	-1.393	-0.052	-0.867	0.178	0.401	-0.179	-0.608
	(3.32)	(2.29)	(0.84)	(2.07)	(0.67)	(1.80)	(0.13)	(1.04)
FPL 100-300	**-6.838***	**-4.691***	-0.601	-1.704	-0.845	-1.702	**-0.215***	-1.292
	(2.79)	(1.93)	(0.71)	(1.76)	(0.57)	(1.53)	(0.11)	(0.89)
Missing FPL	5.606	2.801	-2.462	-8.120	0.722	0.929	**-0.379**	**-5.351***
	(7.52)	(5.19)	(1.92)	(4.73)	(1.53)	(4.09)	(0.29)	(2.37)
Agency 2	2.861	2.020	**2.522****	**6.183****	0.771	2.194	0.101	-0.631
	(3.22)	(2.22)	(0.83)	(2.04)	(0.65)	(1.75)	(0.13)	(1.04)

Agency 3	-5.472	-4.539	2.251	5.354	1.233	2.755	0.133	0.173
	(4.86)	(3.36)	(1.24)	(3.06)	(0.98)	(2.62)	(0.20)	(1.53)
Agency 4	7.415	5.676	**3.345***	**10.684****	1.126	2.752	**0.584***	0.372
	(6.47)	(4.46)	(1.68)	(4.12)	(1.31)	(3.51)	(0.26)	(2.05)
Agency 5	**-9.560***	**-6.347***	-0.124	-0.032	-0.281	-0.658	-0.114	-0.599
	(3.95)	(2.73)	(1.02)	(2.52)	(0.81)	(2.16)	(0.16)	(1.29)
Class Size	**1.713***	**1.287****	**0.523****	**1.499*****	0.146	0.418	**0.057***	0.027
	(0.70)	(0.48)	(0.18)	(0.44)	(0.14)	(0.38)	(0.03)	(0.22)
ECERS>3	6.673	4.755	**2.244***	**5.899***	1.192	3.635	0.071	0.110
	(3.91)	(2.70)	(1.01)	(2.49)	(0.79)	(2.10)	(0.16)	(1.24)
N	189	189	191	191	193	193	193	193

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are Males, White, English, FPL 300+, and Agency 1. Other controls are test type for children tested in Spanish and an interaction between test type and age to align the English and Spanish tests between them.

Table B.4. Multivariate analyses of children's 2015-16 gains in Raw and Standard scores in relation to child and site or classroom characteristics (if tested in Spanish and English, Spanish score used) with CLASS domain thresholds

	Rec. Vocabulary Raw **(PPVT/TVIP)**	Rec. Vocabulary Standard **(PPVT/TVIP)**	Literacy Raw **(WJ/WM-LW)**	Literacy Standard **(WJ/WM-LW)**	Math Raw **(WJ/WM-AP)**	Math Standard **(WJ/WM-AP)**	Executive Function	
							DCCS	PT
Pre Test	**0.696*****	**0.621*****	**0.944*****	**0.701*****	**0.629*****	**0.590*****	**0.014***	**0.465*****
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.01)	(0.06)
Attendance	**0.250***	**0.169***	0.017	0.072	**0.059****	**0.151***	0.004	0.046
	(0.11)	(0.08)	(0.03)	(0.07)	(0.02)	(0.06)	(0.00)	(0.04)
Missing Attendance	**20.346***	**14.315***	3.591	11.479	**4.797***	**12.482***	0.496	**6.334***
	(9.66)	(6.67)	(2.48)	(6.14)	(1.95)	(5.23)	(0.37)	(3.07)
Days Between Tests	-0.149	-0.131	-0.037	-0.122	-0.025	-0.111	-0.004	0.019
	(0.12)	(0.08)	(0.03)	(0.08)	(0.02)	(0.07)	(0.00)	(0.04)
Female	-0.114	-0.376	-0.087	0.114	-0.194	-0.886	0.058	-0.455
	(1.94)	(1.34)	(0.51)	(1.26)	(0.40)	(1.07)	(0.08)	(0.63)
Age in Months	**6.467****	0.700	0.392	**-2.543***	0.573	**-4.656*****	**0.447*****	**1.544***
	(2.30)	(1.34)	(0.57)	(1.28)	(0.48)	(1.08)	(0.07)	(0.66)
Black	-4.555	-3.520	0.295	0.673	-0.377	-1.700	-0.298*	-1.911
	(3.17)	(2.19)	(0.82)	(2.03)	(0.65)	(1.74)	(0.12)	(1.02)
Asian	-3.906	-3.023	-1.238	-3.117	0.626	1.815	-0.143	0.665
	(3.69)	(2.55)	(0.92)	(2.29)	(0.73)	(1.96)	(0.14)	(1.16)
Hispanic	0.379	0.200	-0.771	-3.564	1.600	3.476	0.041	0.934

	(3.97)	(2.74)	(1.12)	(2.75)	(0.85)	(2.28)	(0.15)	(1.26)
Other Race	-6.994	-3.923	-1.053	-3.048	-1.411	-3.898	-0.031	-0.435
	(4.27)	(2.94)	(1.10)	(2.72)	(0.86)	(2.30)	(0.16)	(1.37)
Missing Race	-10.753	-7.866	-0.336	-1.489	**-2.765***	**-7.717***	-0.378	-2.038
	(6.45)	(4.45)	(1.69)	(4.18)	(1.31)	(3.54)	(0.25)	(2.14)
Bilingual	**-7.202***	**-5.003***	-0.631	-1.353	-0.506	-1.480	-0.167	-1.739
	(3.03)	(2.09)	(0.77)	(1.89)	(0.60)	(1.61)	(0.11)	(0.92)
Missing Language	4.697	3.716	0.330	1.818	1.407	3.942	0.328	1.701
	(6.47)	(4.47)	(1.72)	(4.23)	(1.33)	(3.58)	(0.25)	(2.13)
FPL <100	-2.076	-1.749	-0.299	-1.531	0.070	0.064	-0.169	-0.616
	(3.28)	(2.26)	(0.84)	(2.07)	(0.66)	(1.78)	(0.12)	(1.04)
FPL 100-300	**-6.333***	**-4.333***	-0.528	-1.511	-0.814	-1.650	**-0.212***	-1.290
	(2.78)	(1.92)	(0.72)	(1.78)	(0.56)	(1.52)	(0.11)	(0.90)
Missing FPL	3.658	1.378	-2.843	-9.063	0.283	-0.268	-0.442	**-5.380***
	(7.45)	(5.14)	(1.92)	(4.75)	(1.50)	(4.04)	(0.29)	(2.37)
Agency 2	3.903	2.767	**2.452****	**5.801****	0.920	2.437	0.101	-0.632
	(3.26)	(2.25)	(0.85)	(2.09)	(0.66)	(1.77)	(0.13)	(1.06)
Agency 3	-5.765	-4.808	1.742	3.717	1.481	3.269	0.271	0.222
	(5.05)	(3.49)	(1.32)	(3.26)	(1.01)	(2.73)	(0.20)	(1.61)
Agency 4	9.283	6.849	2.321	6.974	2.132	4.842	**0.922*****	0.509
	(7.13)	(4.92)	(1.89)	(4.65)	(1.45)	(3.91)	(0.27)	(2.28)
Agency 5	-6.223	-3.983	0.718	1.933	0.202	0.670	-0.152	-0.572
	(4.12)	(2.84)	(1.08)	(2.67)	(0.84)	(2.26)	(0.16)	(1.36)
Class Size	1.562	1.132	0.287	0.704	0.270	0.655	**0.109****	0.053
	(0.94)	(0.65)	(0.25)	(0.62)	(0.19)	(0.52)	(0.04)	(0.31)
CLASS ES>5.5	0.847	0.220	0.451	-0.010	0.827	2.076	-0.055	0.151
	(4.39)	(3.03)	(1.15)	(2.85)	(0.88)	(2.37)	(0.17)	(1.42)
CLASS CO>5.5	6.180	4.423	0.195	0.066	1.137	2.619	**0.291***	0.085
	(3.47)	(2.40)	(0.92)	(2.26)	(0.71)	(1.92)	(0.14)	(1.14)
CLASS IS>3	-0.138	0.126	1.099	3.706	-0.751	-1.537	**-0.343***	-0.144
	(3.63)	(2.50)	(0.97)	(2.40)	(0.74)	(2.00)	(0.14)	(1.19)
N	189	189	191	191	193	193	193	193

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are Males, White, English, FPL 300+, and Agency 1. Other controls are test type for children tested in Spanish and an interaction between test type and age to align the English and Spanish tests between them.

Table B.5. Multivariate analyses of children's 2015-16 gains in relation to child and site or classroom characteristics (if tested in Spanish and English, English score used) with ECERS-3

	Rec. Vocabulary Raw (PPVT/TVIP)	Rec. Vocabulary Standard (PPVT/TVIP)	Literacy Raw (WJ/WM-LW)	Literacy Standard (WJ/WM-LW)	Math Raw (WJ/WM-AP)	Math Standard (WJ/WM-AP)	Executive Function	
							DCCS	PT
Pre Test	**0.678*****	**0.611*****	**0.952*****	**0.705*****	**0.637*****	**0.595*****	**0.015***	**0.469*****
	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.01)	(0.06)
Attendance	**0.251***	**0.163***	0.018	0.072	**0.064****	**0.161****	0.004	0.046
	(0.11)	(0.08)	(0.03)	(0.07)	(0.02)	(0.06)	(0.00)	(0.04)
Missing Attendance	**20.081***	**13.589***	3.954	12.026	**5.281****	**13.848***	0.528	**6.352***
	(9.81)	(6.90)	(2.57)	(6.35)	(2.03)	(5.42)	(0.38)	(3.05)
Days Between Tests	-0.141	**-0.143***	**-0.056***	**-0.198****	-0.006	-0.069	-0.001	0.013
	(0.10)	(0.07)	(0.03)	(0.07)	(0.02)	(0.06)	(0.00)	(0.03)
Female	0.104	-0.384	-0.121	0.070	-0.185	-0.836	0.057	-0.434
	(1.96)	(1.38)	(0.53)	(1.31)	(0.41)	(1.11)	(0.08)	(0.63)
Age in Months	**7.345****	1.053	0.528	-2.058	0.527	**-4.649*****	**0.418*****	**1.556***
	(2.31)	(1.37)	(0.57)	(1.28)	(0.48)	(1.08)	(0.07)	(0.64)
Black	-4.680	-3.304	0.370	0.836	-0.318	-1.531	**-0.274***	-1.803
	(3.21)	(2.25)	(0.85)	(2.09)	(0.67)	(1.80)	(0.12)	(1.02)
Asian	-4.392	-3.013	-1.104	-2.771	0.626	1.875	-0.170	0.667
	(3.75)	(2.64)	(0.94)	(2.33)	(0.75)	(2.00)	(0.14)	(1.16)
Hispanic	2.222	1.518	-0.627	-3.283	1.654	3.652	0.062	1.005
	(4.21)	(2.96)	(1.15)	(2.82)	(0.88)	(2.35)	(0.16)	(1.26)
Other Race	-7.838	-4.281	-1.058	-3.156	-1.452	-3.990	-0.065	-0.528
	(4.31)	(3.03)	(1.12)	(2.76)	(0.88)	(2.34)	(0.16)	(1.36)
Missing Race	-10.939	-7.821	-0.124	-0.928	**-2.699***	**-7.441***	-0.383	-1.889
	(6.53)	(4.59)	(1.73)	(4.26)	(1.35)	(3.60)	(0.25)	(2.13)
Bilingual	**-6.397***	**-4.536***	-0.647	-1.380	-0.458	-1.381	-0.144	-1.652
	(3.11)	(2.19)	(0.78)	(1.93)	(0.61)	(1.63)	(0.11)	(0.92)
Missing Language	4.483	3.582	0.272	1.704	1.259	3.586	0.296	1.641
	(6.54)	(4.60)	(1.74)	(4.30)	(1.36)	(3.63)	(0.26)	(2.12)
FPL <100	-2.490	-1.922	-0.166	-1.171	0.144	0.306	-0.181	-0.594
	(3.34)	(2.34)	(0.86)	(2.11)	(0.68)	(1.81)	(0.13)	(1.04)
FPL 100-300	**-6.701***	**-4.674***	-0.643	-1.878	-0.852	-1.836	**-0.224***	-1.368
	(2.82)	(1.98)	(0.73)	(1.81)	(0.58)	(1.55)	(0.11)	(0.90)
Missing FPL	3.323	1.280	-3.158	**-9.732***	0.212	-0.718	-0.398	**-5.404***
	(7.52)	(5.28)	(1.99)	(4.90)	(1.56)	(4.17)	(0.29)	(2.35)
Agency 2	7.180	4.497	2.676	4.640	**2.192***	5.303	-0.021	-1.711
	(5.15)	(3.62)	(1.40)	(3.44)	(1.08)	(2.90)	(0.21)	(1.63)
Agency 3	-2.415	-1.779	**2.724***	5.901	1.870	4.481	0.112	-0.083
	(4.96)	(3.49)	(1.35)	(3.32)	(1.05)	(2.81)	(0.20)	(1.56)

Agency 4	9.023	6.599	3.186	**9.367***	1.716	4.078	**0.528***	-0.098
	(6.64)	(4.66)	(1.78)	(4.37)	(1.37)	(3.67)	(0.27)	(2.10)
Agency 5	-2.423	-2.123	0.598	-0.319	1.666	3.978	-0.223	-1.715
	(5.99)	(4.22)	(1.65)	(4.08)	(1.29)	(3.44)	(0.24)	(1.89)
Class Size	1.155	0.945	**0.418***	**1.373****	-0.013	0.030	**0.063***	0.102
	(0.76)	(0.54)	(0.20)	(0.49)	(0.16)	(0.42)	(0.03)	(0.24)
ECERS	5.976	3.525	0.591	-0.456	1.698	3.867	-0.114	-1.123
	(4.38)	(3.08)	(1.19)	(2.95)	(0.93)	(2.49)	(0.18)	(1.38)
N	189	189	186	186	186	186	193	193

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are Males, White, English, FPL 300+, and Agency 1. Other controls are test type for children tested only in Spanish and an interaction between test type and age to align the English and Spanish tests between them.

Table B.6. Multivariate analyses of children's 2015-16 gains in relation to child and site or classroom characteristics (if tested in Spanish and English, English score used) with CLASS dimensions

	Rec. Vocabulary Raw **(PPVT/TVIP)**	Rec. Vocabulary Standard **(PPVT/TVIP)**	Literacy Raw **(WJ/WM-LW)**	Literacy Standard **(WJ/WM-LW)**	Math Raw **(WJ/WM-AP)**	Math Standard **(WJ/WM-AP)**	Executive Function	
							DCCS	PT
Pre Test	**0.680*****	**0.612*****	**0.942*****	**0.696*****	**0.637*****	**0.595*****	**0.013***	**0.467*****
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.01)	(0.06)
Attendance	**0.226***	0.150	0.021	0.084	**0.059***	**0.152***	0.003	0.043
	(0.11)	(0.08)	(0.03)	(0.07)	(0.02)	(0.06)	(0.00)	(0.04)
Missing Attendance	17.390	12.178	4.241	**13.236***	**4.836***	**13.023***	0.363	5.990
	(9.91)	(6.96)	(2.63)	(6.50)	(2.08)	(5.54)	(0.39)	(3.15)
Days Between Tests	-0.187	**-0.165***	-0.045	**-0.157***	-0.015	-0.084	-0.002	0.015
	(0.10)	(0.07)	(0.03)	(0.07)	(0.02)	(0.06)	(0.00)	(0.03)
Female	0.067	-0.456	-0.188	-0.121	-0.200	-0.901	0.062	-0.436
	(1.94)	(1.36)	(0.53)	(1.30)	(0.41)	(1.11)	(0.08)	(0.63)
Age in Months	**6.969****	0.777	0.442	-2.377	0.472	**-4.838*****	**0.419*****	**1.571***
	(2.30)	(1.36)	(0.56)	(1.27)	(0.48)	(1.08)	(0.07)	(0.65)
Black	-4.631	-3.238	0.460	1.023	-0.282	-1.432	**-0.299***	-1.897
	(3.16)	(2.21)	(0.84)	(2.08)	(0.67)	(1.79)	(0.12)	(1.02)
Asian	-4.651	-3.139	-0.997	-2.422	0.603	1.863	-0.188	0.640
	(3.70)	(2.60)	(0.94)	(2.32)	(0.75)	(2.00)	(0.14)	(1.16)
Hispanic	1.853	1.312	-0.633	-3.228	1.613	3.577	0.025	0.897
	(4.14)	(2.91)	(1.14)	(2.80)	(0.87)	(2.33)	(0.16)	(1.26)
Other Race	-7.001	-3.602	-0.823	-2.421	-1.411	-3.810	-0.044	-0.393

	(4.24)	(2.97)	(1.11)	(2.75)	(0.88)	(2.34)	(0.16)	(1.37)
Missing Race	-11.730	-8.412	-0.256	-1.262	**-2.778***	**-7.684***	-0.426	-2.032
	(6.40)	(4.50)	(1.71)	(4.21)	(1.34)	(3.57)	(0.25)	(2.13)
Bilingual	**-6.807***	**-4.829***	-0.678	-1.452	-0.516	-1.529	-0.170	-1.758
	(3.05)	(2.15)	(0.77)	(1.91)	(0.61)	(1.63)	(0.11)	(0.92)
Missing Language	6.312	4.849	0.453	2.106	1.445	4.032	0.363	1.870
	(6.46)	(4.54)	(1.74)	(4.28)	(1.36)	(3.63)	(0.26)	(2.14)
FPL <100	-1.815	-1.493	-0.155	-1.144	0.201	0.421	-0.152	-0.513
	(3.30)	(2.32)	(0.85)	(2.11)	(0.68)	(1.81)	(0.13)	(1.05)
FPL 100-300	**-5.793***	**-4.020***	-0.540	-1.555	-0.805	-1.697	-0.177	-1.198
	(2.78)	(1.95)	(0.73)	(1.81)	(0.58)	(1.55)	(0.11)	(0.90)
Missing FPL	3.933	1.646	-3.137	**-9.748***	0.294	-0.524	-0.360	**-5.370***
	(7.38)	(5.18)	(1.97)	(4.85)	(1.55)	(4.14)	(0.29)	(2.35)
Agency 2	8.052	5.921	**3.061***	**7.581***	1.208	3.254	0.321	0.061
	(4.61)	(3.24)	(1.23)	(3.03)	(0.96)	(2.58)	(0.18)	(1.52)
Agency 3	-4.841	-2.885	2.752	7.226	0.973	2.439	0.152	0.445
	(5.60)	(3.94)	(1.53)	(3.79)	(1.20)	(3.21)	(0.22)	(1.79)
Agency 4	**21.532***	**15.444***	3.913	11.979	2.309	5.317	**1.266****	2.464
	(10.85)	(7.63)	(2.88)	(7.08)	(2.27)	(6.08)	(0.42)	(3.52)
Agency 5	-4.979	-2.986	0.813	2.426	0.268	1.039	-0.009	-0.271
	(4.41)	(3.10)	(1.19)	(2.93)	(0.93)	(2.49)	(0.17)	(1.48)
Class Size	**2.964***	**2.078***	0.411	1.174	0.219	0.489	**0.133****	0.260
	(1.21)	(0.85)	(0.32)	(0.80)	(0.26)	(0.68)	(0.05)	(0.40)
CLASS_ES	4.905	2.348	-0.488	-2.920	1.233	2.615	0.131	0.026
	(5.65)	(3.97)	(1.52)	(3.75)	(1.19)	(3.18)	(0.22)	(1.84)
CLASS_CO	6.410	4.890	1.086	3.232	0.430	1.212	0.246	0.811
	(4.39)	(3.08)	(1.17)	(2.90)	(0.92)	(2.46)	(0.17)	(1.45)
CLASS_IS	-6.638	-4.080	0.271	1.158	-0.692	-1.283	**-0.346***	-0.948
	(4.40)	(3.09)	(1.17)	(2.90)	(0.92)	(2.47)	(0.17)	(1.43)
N	189	189	186	186	186	186	193	193

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are Males, White, English, FPL 300+, and Agency 1. Other controls are test type for children tested only in Spanish and an interaction between test type and age to align the English and Spanish tests between them.

Appendix C. ECERS-3 and CLASS scores by Agency and Class Size. Item level.

Table C.1. ECERS-3 Item, Subscale, and Overall Means by Agency, N=14

ECERS-3 Item and Subscales	Agency 1	Agency 2	Agency 3	Agency 4	Agency 5
Overall	**4.31**	**3.30**	**4.06**	**3.50**	**3.43**
Space and Furnishings	4.43	3.66	4.29	3.86	3.76
1. Indoor space	6.50	6.80	7.00	5.67	6.33
2. Furnishings for care, play and learning	4.00	4.40	4.00	5.00	4.00
3. Room arrangement for play and learning	5.50	3.20	3.00	3.00	4.00
4. Space for privacy	5.50	3.60	5.00	5.00	3.00
5. Child-related display	4.00	2.80	4.00	3.33	3.67
6. Space for gross motor play	3.50	2.80	4.00	3.33	3.00
7. Gross motor equipment	2.00	2.00	3.00	1.67	2.33
Personal Care Routines	3.25	2.75	2.50	3.67	3.42
8. Meals/ snacks	4.00	2.60	2.00	4.00	2.67
9. Toileting/diapering	2.50	2.40	1.00	2.00	2.33
10. Health practices	2.50	2.60	3.00	3.67	3.00
11. Safety practices	4.00	3.40	4.00	5.00	5.67
Language and Literacy	4.70	2.88	4.60	3.20	3.53
12. Helping children expand vocabulary	3.50	3.40	5.00	3.33	3.33
13. Encouraging children to use language	5.50	4.00	6.00	3.67	4.33
14. Staff use of books with children	6.00	2.00	4.00	3.00	2.67
15. Encouraging children's use of books	6.50	3.00	4.00	4.33	4.67
16. Becoming familiar with print	2.00	2.00	4.00	1.67	2.67
Learning Activities	3.50	3.06	3.36	2.38	2.43
17. Fine motor	5.00	4.60	5.00	3.67	4.00
18. Art	4.50	3.80	4.00	3.33	3.33
19. Music and movement	4.50	4.00	4.00	3.33	2.00
20. Blocks	3.50	2.00	3.00	1.00	1.67
21. Dramatic Play	3.50	2.80	6.00	2.33	1.67
22. Nature/science	3.00	2.80	4.00	1.67	2.00
23. Math materials and activities	1.50	2.00	1.00	1.67	1.67
24. Math in daily events	2.50	3.00	3.00	2.67	3.00
25. Understanding written numbers	1.50	1.20	1.00	1.00	1.67
26. Promoting acceptance of diversity	5.50	4.40	5.00	3.67	3.33
Interaction	5.30	3.84	5.60	4.60	4.53
27. Appropriate use of technology	N/A	N/A	1.00	N/A	N/A
28. Supervision of gross motor	5.50	2.80	5.00	4.33	3.00
29. Individualized teaching and learning	4.50	4.00	4.00	3.67	5.00
30. Staff-child interaction	4.00	4.60	7.00	6.00	4.33
31. Peer interaction	6.00	4.20	6.00	5.00	5.33
32. Discipline	6.50	3.60	6.00	4.00	5.00
Program Structure	5.84	3.80	4.67	4.89	4.00
33. Transitions and waiting times	6.50	3.80	5.00	5.67	4.67
34. Free play	6.00	4.40	5.00	4.00	4.00
35. Whole -group activities for play and learning	5.00	3.20	4.00	5.00	3.33

Table C.2. CLASS Dimension and Domain Means by Agency, N = 14

CLASS Dimensions and Domains	Agency 1	Agency 2	Agency 3	Agency 4	Agency 5
Emotional Support Domain	6.25	5.99	6.81	6.06	6.15
1. Positive Climate	5.88	5.70	7.00	5.58	5.75
2. Negative Climate*	7.00	6.70	7.00	7.00	6.83
3. Teacher Sensitivity	6.25	5.70	6.50	5.67	6.08
4. Regard for Student Perspectives	5.88	5.85	6.75	6.00	5.92
Classroom Organization Domain	6.38	5.40	6.33	5.44	5.64
5. Behavior Management	6.50	5.40	6.75	5.50	5.67
6. Productivity	6.88	5.90	6.50	5.58	6.08
7. Instructional Learning Formats	5.75	4.90	5.75	5.25	5.17
Instructional Support Domain	2.75	2.53	3.25	2.53	2.72
8. Concept Development	2.13	2.05	2.25	1.92	2.17
9. Quality of Feedback	2.63	2.60	3.75	2.17	2.67
10. Language Modeling	3.50	2.95	3.75	3.50	3.33

*The Negative Climate dimension was transposed so that on here, high represents “good”

Table C.3. ECERS-3 Item, Subscale, and Overall Means by Class Size, N=14

ECERS-3 Item and Subscales	Small (≤18) (6 classrooms)	Large (>18) (8 classrooms)
Overall	**3.52**	**3.60**
Space and Furnishings	3.91	3.86
1. Indoor space	6.33	6.50
2. Furnishings for care, play and learning	4.50	4.25
3. Room arrangement for play and learning	3.67	3.63
4. Space for privacy	4.17	4.13
5. Child-related display	3.33	3.38
6. Space for gross motor play	3.50	2.88
7. Gross motor equipment	1.83	2.25
Personal Care Routines	3.33	3.00
8. Meals/ snacks	3.33	2.88
9. Toileting/diapering	2.33	2.13
10. Health practices	3.17	2.75
11. Safety practices	4.50	4.25
Language and Literacy	3.23	3.65
12. Helping children expand vocabulary	3.33	3.63
13. Encouraging children to use language	4.00	4.63
14. Staff use of books with children	3.17	3.00
15. Encouraging children's use of books	4.17	4.25
16. Becoming familiar with print	1.50	2.75
Learning Activities	2.61	3.06
17. Fine motor	4.17	4.50
18. Art	3.50	3.88
19. Music and movement	3.33	3.63
20. Blocks	1.33	2.50
21. Dramatic Play	2.67	2.88
22. Nature/science	2.33	2.63
23. Math materials and activities	1.50	1.88
24. Math in daily events	2.50	3.13
25. Understanding written numbers	1.00	1.50
26. Promoting acceptance of diversity	4.00	4.38
Interaction	4.47	4.50
27. Appropriate use of technology	1.00	1.00
28. Supervision of gross motor	4.00	3.50
29. Individualized teaching and learning	3.83	4.50
30. Staff-child interaction	5.50	4.50
31. Peer interaction	4.83	5.13
32. Discipline	4.17	4.88
Program Structure	4.89	4.09
33. Transitions and waiting times	5.50	4.38
34. Free play	4.83	4.25
35. Whole -group activities for play and learning	4.33	3.63

Table C.4. CLASS Dimension and Domain Means by Class Size, N = 14

CLASS Dimensions and Domains	Small (≤18) (6 classrooms)	Large (>18) (8 classrooms)
Emotional Support Domain	6.08	6.18
1. Positive Climate	5.67	5.91
2. Negative Climate*	1.04	1.22
3. Teacher Sensitivity	5.79	6.00
4. Regard for Student Perspectives	5.88	6.03
Classroom Organization Domain	5.60	5.72
5. Behavior Management	5.71	5.75
6. Productivity	5.92	6.16
7. Instructional Learning Formats	5.17	5.25
Instructional Support Domain	2.43	2.82
8. Concept Development	1.88	2.22
9. Quality of Feedback	2.29	2.84
10. Language Modeling	3.13	3.41

*The Negative Climate dimension was transposed so that on here, high represents “good”

Appendix D. Family Survey Tables.

This appendix first compares survey non-respondents, respondents and non-consented based on DEEL data on gender, children's age, ethnicity, language and FPL for all children on the target sample if 224. It then compares survey non-respondents and respondents for the 192 children for which we were able to collect pre- and post-test data on at least one measure. Groups of respondents and non-respondents differed in ethnicity and FPL. In addition, children for which parents did not consent participation differed significantly in ethnicity, language, and FPL from the overall sample. These children were less likely to be white, more likely to be of Asian origin, less likely to be of unknown or other ethnic background, more likely to be of English or Vietnamese speaking background, and more likely to be above 300 FPL.

The rest of the appendix presents tables for respondents the different indicators and information captured by the family survey. The distribution of respondents and non-respondents statistically differs on race/ethnicity and FPL.

Table D.1.a, Respondents (45.5% of children's families), non-Respondents and Non-Consented from target sample

DEEL Child Information		Respondent %	Non-respondent %	Non-Consented %
Total (N=224)				
Gender	Male (N=108)	49.14	47.06	50.00
	Female (N=116)	50.86	52.94	50.00
Ethnicity*	White (N=60)	12.93	43.14	16.67
	Black (N=56)	35.34	12.75	33.33
	Asian (N=29)	10.34	14.71	33.33
	Hispanic (N=28)	14.66	9.8	16.67
	Other (N=14)	6.03	6.86	0.00
	Unknown (N=37)	20.69	12.75	0.00
Language	English (N=148)	62.07	69.61	83.33
	Spanish (N=17)	7.76	7.84	0.00
	Vietnamese (N=7)	2.59	2.94	16.67
	Other (N=21)	10.34	8.82	0.00
	Unknown (N=31)	17.24	10.78	0.00
FPL*	<100 (N=73)	43.97	19.61	33.33
	100-300 (N=88)	38.79	41.18	16.67
	>300 (N=47)	11.21	30.39	50.00
	Unknown (N=16)	6.03	8.82	0.00

*Respondent versus Non-respondent Distribution was statistically significantly different.

Table D.1.b. Respondents (53.1% of children's families) versus Non-respondents, children with pre- and post-test

DEEL Child Information		Respondent %	Non-respondent %
Total (N=192)			
Gender	Male (N=94)	51.11	47.06
	Female (N=98)	48.89	52.94
Ethnicity*	White (N=56)	13.33	43.14
	Black (N=47)	37.78	12.75
	Asian (N=25)	11.11	14.71
	Hispanic (N=23)	14.44	9.8
	Other (N=13)	6.67	6.86
	Unknown (N=28)	16.67	12.75
Language	English (N=129)	64.44	69.61
	Spanish (N=14)	6.67	7.84
	Vietnamese (N=5)	2.22	2.94
	Other (N=19)	11.11	8.82
	Unknown (N=25)	15.56	10.78
FPL*	<100 (N=57)	41.11	19.61
	100-300 (N=80)	42.22	41.18
	>300 (N=41)	11.11	30.39
	Unknown (N=14)	5.56	8.82

*Respondent versus Non-respondent distribution was statistically significantly different.

Socioeconomic indicators

Table D.2. Socioeconomic indicators

	N	Percent
Parent Education		
Less than 9th grade	2	1.8
Some high school	2	1.8
GED	2	1.8
High school diploma	8	7.3
Some college	26	23.9
Associate's degree	11	10.1
Bachelor's degree	26	23.9
Master's degree or higher	32	29.4
Total	109	100.0
Annual household income		
10,000 or less	8	7.3
11,000-20,000	7	6.4
21,000-30,000	7	6.4
31,000-40,000	13	11.9
41,000-50,000	13	11.9
51,000-60,000	9	8.3
61,000-70,000	8	7.3
71,000-80,000	13	11.9
81,000 or more	31	28.4
Total	109	100.0

Welfare

Table D.3. Welfare

	N	Percent Yes
Cash assistance or Public Benefits	111	27.0
Food stamps	24	21.6
WIC	33	64.7
TANF	7	13.2
Early Head Start	7	14.6
Head Start	6	12.8
Medicaid	25	59.5
Medicare	15	29.4
ESEAP	7	14.6
Working Connections	12	25.5
Food Bank	10	20.4
Since birth, serious financial problems?	110	34.6

Language and Immigration

Table D.4. Language and immigration background

		N	Percent	Mean	SD
Primary Language	English	90	81.1		
	Spanish	7	6.3		
	Vietnamese	4	3.6		
	Mandarin	3	2.7		
	Other	7	6.3		
	Total	111	100.0		
Bilingual goal for child		90	81.1		
Child's generational status	First	7	0.0		
	First generation	7	6.4		
	Second generation	34	30.9		
	Third-plus generation	69	62.7		
If first generation, age of arrival		7		1.5	1.4

Table D.5. Family structure and stability

		N	Percent	Mean	SD
Parents currently living at home	Two parents (both biological or adoptive)	65	59.6		
	Two parents (one biological and one other)	6	5.5		
	One parent	34	31.2		
	Other	4	3.7		
	Total	109	100.0		
Years in current residence		110		2.2	1.4
Age of mother at birth/adoption of child		102		32.6	7.7
Marital status	Never married	24.0	21.8		
	Domestic partnership, never married	7.0	6.4		
	Married	56.0	50.9		
	Separated	5.0	4.6		
	Divorced	17.0	15.5		
	Widowed	1.0	0.9		
	Total	110.0	100		

Table D.6. Preschool choices

Importance if cost was not an issue	**N**	**Percent**
Location or convenience	10	9.8
Quality, Curriculum, Teaching	34	33.3
Bilingual	5	4.9
Socio-emotional development	19	18.6
Diversity	6	5.9
Overall development, KG Readiness	21	20.6
Other	8	7.8

Parental perceptions on SPP programs and teachers

Table D.7. Perceptions on the child's development and the program

	N	Strongly disagree	2	3	4	5	Strongly agree
Positive changes on child since SPP enrollment							
Language	107	5.6	0.0	5.6	13.1	24.3	51.4
Physical Development	108	1.9	1.9	7.4	20.4	33.3	35.2
Behavioral/ Socio-Emotional	109	3.7	1.8	3.7	17.4	37.6	35.8
Literacy	109	4.6	3.7	8.3	12.8	29.4	41.3
Math	107	6.5	2.8	16.8	11.2	29.0	33.6
Science	107	7.5	5.6	16.8	20.6	22.4	27.1

Regarding the child's program	N	Strongly disagree	2	3	4	5	Strongly agree
I feel connected with my child's teacher	109	4.6	0.9	10.1	8.3	21.1	55.0
I feel connected with my child's preschool	109	3.7	0.9	6.4	11.0	27.5	50.5
I have received work samples	106	3.8	4.7	7.5	3.8	14.2	66.0
I have received assessment results	107	4.7	2.8	3.7	10.3	21.5	57.0
I know about the curriculum that is used	107	1.9	4.7	10.3	13.1	18.7	51.4
I feel welcome at the preschool	109	0.9	1.8	2.8	2.8	17.4	74.3
I have received feedback about my child's performance	109	2.8	0.9	0.9	7.3	21.1	67.0

Table D.8. Perceptions of the teacher

Regarding the child's teacher	N	No	Yes
Talks to me each day	110	15.5	84.6
Uses a curriculum for teaching	101	5.0	95.1
Teaches my child behavioral/social/emotional skills	107	2.8	97.2
Teaches my child academic skills	106	5.7	94.3
Tracks my child's progress	107	5.6	94.4
Is fluent in my child's primary home language	107	12.2	87.9
Has a Bachelor's degree	58	8.6	91.4
Engages in training opportunities	77	3.9	96.1

Table D.9. Perceptions of the child's feelings towards the program

How often did your child do any of the following regarding his/her SPP preschool?	N	Never	1-2 times	3-4 times	Every day
Said good things about the school	108	2.8	18.5	38.0	40.7
Was upset or didn't want to go to the preschool	110	62.7	31.8	3.6	1.8
Said he/she liked his or her teacher(s)	110	13.6	23.6	25.5	37.3
Pretended to be sick to stay home from the preschool	110	87.3	11.8	0.9	0.0
Was excited about going to the preschool	111	3.6	23.4	24.3	48.7

Table D.10. Perceptions of feeling welcome or unwelcome by the program

How welcome do you feel in this preschool?	N	Percent
Very uncomfortable	16	14.4
Somewhat uncomfortable	2	1.8
Somewhat comfortable	16	14.4
Very comfortable	77	69.4
Total	111	100.0

Table D.11. Parental participation in the program

How often do you…	N	Never	< 1 per week	Once per week	> once per week
Volunteer in your child's school?	110	76.4	20.9	0.9	1.8
Volunteer as a chaperone on field trips?	109	75.2	22.9	0.9	0.9
Talk with your child's teacher at pick up/drop off?	111	0.9	6.3	10.8	82.0
Talk with your child's preschool director at pick up/drop off?	110	14.6	21.8	18.2	45.5
Communicate with your child's teacher by phone?	111	19.8	55.0	8.1	17.1

Parenting practices

Table D.12. Parenting activities with the child

In a typical week, how often do you…	N	Not at all	1-2 times per week	3-6 times per week	Every day
Read books to your child	111	1.8	15.3	32.4	50.5
Tell stories to your child	110	3.6	25.5	27.3	43.6
Sing songs and/or dance with your child	110	1.8	19.1	40.0	39.1
Help your child to do arts & crafts	109	6.4	41.3	42.2	10.1
Write with your child	111	12.6	39.6	37.8	9.9
Involve your child in household chores	111	1.8	18.9	40.5	38.7
Take your child on errands	111	0.9	23.4	37.8	37.8
Play pretend or role playing games	111	4.5	34.2	31.5	29.7
Watch TV with your child	110	7.3	27.3	36.4	29.1
Play video games with your child	110	68.2	20.0	9.1	2.7
Do puzzles with your child	111	16.2	62.2	17.1	4.5
Talk about numbers and/or shapes with your child	108	2.8	25.0	36.1	36.1
Talk about nature or do science projects with your child	110	10.0	40.0	30.0	20.0
Build or play construction toys with your child	110	9.1	46.4	30.9	13.6
Take your child to the library	110	33.6	62.7	2.7	0.9
Go for a walk/play outside with your child	111	3.6	35.1	38.7	22.5
Take your child to the park or playground	111	3.6	59.5	27.0	9.9
Take your child to museum/zoo/other ed. site	107	22.4	69.2	5.6	2.8
Play a sport or exercise together	109	22.0	57.8	13.8	6.4
Engage in faith-based activities	108	59.3	26.9	7.4	6.5
Visit relatives or friends	110	5.5	60.9	24.6	9.1
Extra academic program	111	88.3	4.5	4.5	2.7

Table D.13. Number of books in the home

Number of books in the home	N	Percent
1 to 10	7	6.3
11 to 20	11	9.9
More than 20	93	83.8
	111	100.0

Table D.14. Screen time

Screen time on his/her own	N	Mean	SD
Typical week day	109	1.5	1.1
Typical weekend day	111	2.7	2.1

Other care and past care

Table D.15. Out-of-home care used in addition to SPP

Other out-of-home care in addition to SPP?	N	Percent	Mean	SD
No	90	80.4		
Yes	22	19.6		
Extended day child care	14	63.6		
Developmental preschool	1	4.6		
With a relative	2	9.1		
With a friend or neighbor	1	4.6		
Childcare someplace else	4	18.2		
Total	112	100.0		
Average number of hours	18		12.2	8.3

Table D.16. Out-of-home care used prior to SPP

Experiences prior to SPP	Birth-1	1-2 yr old	2-3 yr old	3-4 yr old
At home with parent	68.8	50.5	32.1	25.0
At home under non-relative care	5.5	10.1	9.2	2.8
Relative care (at home or other)	17.4	16.5	11.9	4.6
Childcare or center-based care	21.1	28.4	42.2	47.2
Family day care	3.7	7.3	11.0	4.6
Early Head Start	0.9	3.7	6.4	19.4

Child's health and abilities

Table D.17. Parental perceptions of child's abilities

Child's abilities compared to peers	N	Not as well as other children	As well as other children	Better than other children
Following adult directions	109	7.3	64.2	28.4
Playing with other children	111	6.3	66.7	27.0
Talking with others	111	6.3	40.5	53.2
Listening to others	111	5.4	68.5	26.1
Running, jumping, skipping	111	0.0	59.5	40.5
Drawing	111	8.1	68.5	23.4
Writing	108	15.7	67.6	16.7
Reading	108	16.7	69.4	13.9
Counting	111	5.4	64.9	29.7

Table D.18. Parental reports on child's wellness checks, physical and mental health

Health	N	Percent
Health visit past 12 month	111	99.1
Dentist visit past 12 month	111	95.5
Diagnosed with a developmental delay	110	4.6
Health		
Fair	1	0.9
Good	10	9.3
Very good	37	34.3
Excellent	60	55.6
IEP	107	4.7
Witnessed violence in the community	109	8.3

Food fragility

Table D.19. Food fragility as measured by parental reports on affording meals

Food fragility	N	Never	Sometimes	Often
We worried food would run out	107	80.4	15.9	3.7
The food we bought just didn't last	107	86.9	10.3	2.8
We couldn't afford balanced meals	106	85.9	13.2	0.9
We relied on only a few kinds of low-cost meals	107	79.4	19.6	0.9

Financial security and other concerns

Table D.20. Parental financial security concerns and other uncertainties

How concerned are you with…	N	Not at all concerned	Not very concerned	Somewhat concerned	Extremely concerned
Setting a good financial example for my child	110	15.5	11.8	29.1	43.6
Saving enough to be able to retire	109	7.3	7.3	39.5	45.9
Rising costs of health care	110	5.5	16.4	41.8	36.4
Saving for my child's college education	110	4.6	9.1	24.6	61.8
Saving for emergencies	110	1.8	12.7	34.6	50.9
Financial account security	110	6.4	13.6	38.2	41.8
Being able to pay bills month to month	110	11.8	30.9	33.6	23.6
Taking on too much debt	109	17.4	22.9	34.9	24.8
Supporting other adult family members	110	28.2	40.0	26.4	5.5
Natural disasters	111	15.3	39.6	34.2	10.8
Getting bitten by a dog	110	60.9	31.8	0.9	6.4