NIEER

NATIONAL INSTITUTE FOR EARLY EDUCATION RESEARCH

EVALUATION OF WEST VIRGINIA UNIVERSAL PRE-K. SECOND YEAR LONGITUDINAL OUTCOMES

August 2019

Milagros Nores, Ph.D., Kwanghee Jung, Ph.d., Shannon Riley-Ayers, Ph.D., Andrea Kent, M.Ed., Carol Contreras, B.A, The National Institute for Early Education Research

Mindy Allenger, Ed.D., Marshall University

About the Authors

Milagros Nores, Ph.D. is Co-Director for Research at The National Institute for Early Education Research (NIEER) at Rutgers University. Dr. Nores conducts research at NIEER on issues related to early childhood policy, programs, and evaluation, both nationally and internationally.

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

Shannon Riley-Ayers, Ph.D. is a Senior Program Officer at the Nicholson Foundation. At the time of this research, Dr. Riley-Ayers was an associate research professor at The National Institute for Early Education Research (NIEER) at Rutgers University, with her work centered on program evaluation, early literacy, and developmentally appropriate contents in K-3.

Andrea Kent, M.Ed. is a Project Coordinator at The National Institute for Early Education Research (NIEER) at Rutgers University. Her work includes coordination of field work in studies in New Jersey and West Virginia.

Carol Contreras, B.A. is Program Manager at The National Institute for Early Education Research (NIEER) at Rutgers University.

Mindy Allenger, Ed.D. is Assistant Professor at the College of Education & Professional Development, Marshall University. Dr. Allenger's research and teaching focuses on literacy education, diagnosis, and intervention.

Suggested citation: Nores, M., Jung, K, Riley-Ayers, S., Kent, A., Contreras, C. & M. Allenger (2019). Evaluation of West Virginia Universal Pre-K. Second Year Longitudinal Outcomes. New Brunswick, NJ: National Institute for Early Education Research.

Table of Contents

Introduction

The West Virginia (WV) pre-K program serves 65 percent of 4-year-olds in the state and ranks 6th in the nation in access to preschool for 4-year-olds (Friedman-Krauss et al., 2018). West Virginia has shown gains in terms of quality standards in the last few years and currently meets all 10 of NIEER's minimum quality standards benchmarks due to WV's newest requirement for assistant teachers to have at least a Child Development Associate (CDA) credential. WV meets nine of the 10 newly developed and more robust benchmarks (Friedman-Krauss et al., 2018). As state pre-K initiatives have expanded to serve more and more children during the years before kindergarten, it is especially important to study the quality and effectiveness of such programs. It is also central to understand how children who attend the program progress through the elementary years, and likewise understand the quality of the K-12 education they receive.

In the fall of 2015, a five-year longitudinal study of the WV pre-K program was started with the goal of estimating the effects of the WV pre-K program, and the extent to which initial benefits result in persistent educational advantages. This report focuses on the second year of the study, measuring the impacts of pre-K during the children's kindergarten year. This follows a first year report showing large effects of WV pre-K on various children's domains. In this report, impacts have been examined across various child developmental domains. We also report classroom quality in kindergarten tracking the longitudinal quality experienced by the children in the sample through the years. Finally, we report how the impacts varied across children, trying to understand to what extent impacts are sustained over time and for whom.

Overall, the report highlights positive impacts of pre-K on children's learning and development as demonstrated in language and literacy that persist at kindergarten entry, as well as evidence that lower income children benefited the most from the program in literacy, math, and executive functioning (as measured by DCCS). Girls also showed more benefits from the program, particularly in receptive vocabulary. The results provide evidence of program impact of the WV pre-K program on children's learning through kindergarten entry; impacts that may lead to increased school success. However, the positive impacts diminished by the end of kindergarten year. Observations of classroom quality following the longitudinal sample showed that classroom quality in kindergarten may be one of the determinants of this fade-out, as kindergarten classroom quality proved to be lower than the preschool quality experienced by the longitudinal cohort.

This study has some limitations that should be highlighted. The study is centered in children and classrooms from seven counties in WV with lower participation rates in the program to therefore allow for the recruitment of a comparison group. These counties were therefore lagging in terms of expansion of the pre-K program. Thus, generalizability to the rest of the state should not be presumed, to the extent that there are large differences between these counties and those with larger enrollment rates. In addition, we were not able to capture a comparison group in the preschool year, therefore analyses do not include a pretest at age 4.

Study Methods

In the fall of 2015, the research team started a five-year longitudinal study of the WV pre-K program. The goal of the study is to estimate the effects of the WV pre-K program, including the extent to which initial benefits result in persistent educational advantages for children through K-

3. This report focuses on the second year of the study which assessed children in their kindergarten years, examining the following research questions.

1. What is the impact of the prekindergarten program on children's language, math, literacy, and executive functions skill measures at kindergarten entry and at the end of kindergarten year?
2. Are there child subgroups (as defined by low income or child gender) that benefit more from the program than others?
3. What is the overall observed quality of pre-K and kindergarten classrooms in 2016-17?

Sample

In fall of 2015, at the outset of this study, we randomly selected two groups of children as our initial sample: 599 children who were just beginning the WV pre-K program and 573 children who had attended the pre-K program the previous year and were beginning kindergarten. In the 2016-2017 school year, we followed the pre-K group of children from the previous year into kindergarten and also randomly selected additional entering kindergarteners who had not attended the WV pre-K program. The final sample consists of 605 kindergarten children who attended WV pre-K program the previous year[1] and another group of 366 kindergarten children who did not attend WV pre-K. Table 1 reports demographics for the sample of 971 children in the study. The sample is predominantly White (91.9%) and low income (65.6%), with balance across genders (49% female).

Participating children were comparable to the average children in these districts in terms of gender and race, but control group of children were less likely to be low income, although more similar to average children in the counties. We assessed between 1 and 13 children per kindergarten classroom.

Table 1. Child demographics for sample, N=971.

Child Characteristics	Total sample N=971		Comparison K sample N=366		Treatment K sample N=605		WV school average for these districts*
	N	%	N	%	N	%	
Gender							
Male	504	51.9%	183	50.0%	321	53.1%	51.4%
Female	467	48.1%	183	50.0%	284	46.9%	49.1%
Low Income							
Low Income	622	65.6%	205	56.3%	417	71.4%	68.1%
Other	326	34.4%	159	43.7%	167	28.6%	31.9%
Race/Ethnicity							
White	892	91.9%	340	92.9%	552	91.2%	93.1%
Black	31	3.2%	7	1.9%	24	4.0%	3.5%
Other	48	4.9%	19	5.2%	29	4.8%	3.3%

*Source: WV Department of Education, https://zoomwv.k12.wv.us/Dashboard/portalHome.jsp.

---

[1] Additional pre-K attenders were assessed in this round, who were originally identified as non-attenders, but then tracked in the West Virginia Education Identification System (WVEIS) as pre-K attenders.

In addition, we conducted classroom observations on 127 pre-K classrooms and 140 kindergarten classrooms. In pre-K, ECERS-3 data were collected in 127 classrooms and CLASS data were collected in 123 classrooms in two separate visits.[2] In kindergarten, 140 classrooms were observed using both the APEEC and CLASS simultaneously in one visit. Table 2 reports the sample of observed classrooms for both pre-K and kindergarten.

Table 2. Classroom Sample by Grade and Instrument Used

	Pre-K			Kindergarten
	Observed ECERS-3/CLASS	Observed ECERS-3	Observed CLASS	Observed APEEC/CLASS
Putnam	20	21	21	33
Kanawha	16	16	16	20
Nicholas	7	11	8	5
Roane	6	6	6	7
Wood	32	33	33	38
Greenbrier	16	17	16	16
Fayette	23	23	23	21
Overall	120	127	123	140

Measures

Measures on Children

This evaluation measured child outcomes in receptive vocabulary (using the Peabody Picture Vocabulary Test), literacy (using the Woodcock-Johnson Tests of Achievement: Letter-Word subtest and Passage Comprehension subtest), and math (using the Woodcock-Johnson Tests of Achievement Applied Problems subtest). In addition, it measured executive functioning (EF) using two measures: the Dimensional Change Card Sort Game (DCCS) and the Peg Tapping task (PT). These measures as described below.

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. The *Applied Problems* and *Letter-Word Identification, and Passage Comprehension* 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 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. 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 ages 48-50 months means by age were 1.33, for 51-53 they were 1.42, for 54-56 they were 1.58, for 57-59 they were 1.62, for 60-62 they were 1.80, for 63-65 they were 1.84, for 66-68 they were 1.90, for 69-71 they were 2.09 and for more than 65 months they were 2.17 (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. The task requires two abilities, the ability to hold tapping rules in mind and the ability to exercise inhibitory control over one's proponent behavior (the natural tendency is 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. The final score is a sum of all 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 ages 48-50 months means by age were 4.57, for 51-53 they were 6.02, for 54-56 they were 7.87, for 57-59 they were 8.80, for 60-62 they were 10.33, for 63-65 they were 11.17, for 66-68 they were 13.25, for 69-71 they were 13.85 and for more than 65 months they were 14.35 (Meador et al., 2013).

Measures on Classrooms

Early Childhood Environment Rating Scale—Third Ed. (ECERS-3; Harms, Clifford & Cryer, 2015). 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 (in November 2016)[3] 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.

Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008). The CLASS is an observational system that assesses classroom practices in preschool and kindergarten by measuring the interactions between students and adults. Observations consist of four to five 20-minute cycles followed by 10-minute coding periods. Scores (codes) are assigned during various classroom activities, and then averaged across all cycles for an overall quality

---

[3] Published online at https://www.ersi.info/ecers3_notes.html

score. Interactions are measured through 10 dimensions, which are divided into three domains. CLASS uses a 7-point Likert-scale, for which a score of 1 or 2 indicates low quality and a score of 6 or 7 indicates high quality.

Assessment of Practices in Early Elementary Classrooms (APEEC; Maxwell, McWilliam, Hemmeter, Ault & Schuster, 2001). The APEEC assesses quality in the early elementary environment, kindergarten to third grade, with a focus on developmentally appropriate practices (DAP; Copple & Bredekamp 2009). The APEEC is comprised of 16 items which are rated on a 7-point scale. A score of 1 indicates inadequate quality, a score of 5 indicates good quality and a score of 7 indicates excellent quality.

Procedures

NIEER and Marshall University worked collaboratively to hire and train data collectors on the child assessment and classroom observation measures. For child assessments, data collectors received a two-day training on the measures. Following a two-day training, data collectors were shadowed by trained staff in the field for reliability.

For classroom observation measures, training in administering the protocol that includes the ECERS-3 and the CLASS for pre-K classrooms and the APEEC and the CLASS for kindergarten classrooms was provided in two weeks. ECERS-3 observers were trained by an ECERS-3 certified trainer to meet ERSI reliability requirements for observer certification.[4] Data collectors met ECERS-3 reliability requirements with agreement percentages between 86-96%. CLASS observers were trained by a CLASS certified trainer and met the Teachstone reliability requirements for observer certification.[5] Data collectors met CLASS reliability requirements with agreement percentages ranging between 86-98% for CLASS PreK and 84-92% for CLASS K. APEEC observers were similarly trained and met agreement percentages with the trainers between 85-92%. All observation data was cleaned and entered at NIEER by trained staff.

Results

Below we address the research questions regarding the impacts of the pre-K program on child outcomes at kindergarten entry and through kindergarten progress. The main results are summarized in this report with additional analyses included in the appendices.

1. What is the impact of the prekindergarten program on children's language, math, literacy, and executive functions skill measures at kindergarten entry and at the end of kindergarten year?

We first present descriptive results from the 2016-2017 evaluation showing average scores for the whole sample, and then separated by treatment and comparison groups at kindergarten entry and at the end of kindergarten. We then estimated through multi-level regression analysis the impact of the program in children's language, literacy, math, and executive function skills. From the WVEIS (WV education information system) we obtained

---

[4] ERSI is the company to which ECERS-3 belongs. More information about the tool and reliability guidelines can be found at http://www.ersi.info/

[5] Teachstone is the company that owns CLASS products and trainings, certifications etc. Training reliability is monitored and reported directly by them. http://www.teachstone.com/about-teachstone/

administrative information on children's race/ethnicity (White, African American, or Other), gender and low-income status. These are included as controls in the estimations further below (coded as dichotomous variables with values 0 or 1).

Table 3 presents summary statistics for the child outcome measures for the overall sample and the treatment and comparison groups. Raw and standard scores are presented for the PPVT and WJ-III subtests. The PPVT and the WJ are standardized with a mean of 100 and a standard deviation of 15. Children scoring above the mean score above average children in the published norms for their age. That is, above what is expected due to natural maturation. Children scoring below the mean score below the norm for their age. Fall scores are kindergarten entry, and spring scores are those at the end at kindergarten. Treatment PPVT standard scores are slightly higher at kindergarten entry, and so are WJ AP standard scores at kindergarten entry and end of kindergarten, as well as LW and PC standard scores at both kindergarten entry and end of kindergarten.

Table 3. Average child scores across the different measures for the total sample, treatment, and comparison groups, N=971

	Total sample		Comparison		Treatment	
	Mean	SD	Mean	SD	Mean	SD
PPVT Raw Fall Score	99.79	19.70	99.81	20.29	99.78	19.35
PPVT Raw Spring Score	110.20	18.32	111.84	18.66	109.20	18.06
PPVT SS Fall Score	105.02	14.77	104.42	15.35	105.39	14.41
PPVT SS Spring Score	106.32	13.46	107.05	13.50	105.88	13.43
WJ-AP Raw Fall Score	16.80	4.19	16.95	4.31	16.72	4.12
WJ-AP Raw Spring Score	20.62	3.67	20.64	3.69	20.62	3.67
WJ-AP SS Fall Score	102.41	12.79	102.00	13.44	102.66	12.38
WJ-AP SS Spring Score	105.40	12.33	104.55	12.98	105.91	11.90
WJ-LW Raw Fall Score	12.96	5.44	13.09	5.86	12.88	5.17
WJ-LW Raw Spring Score	23.56	6.89	23.68	6.87	23.48	6.90
WJ-LW SS Fall Score	97.06	12.29	96.25	13.61	97.55	11.41
WJ-LW SS Spring Score	107.98	13.12	107.23	14.31	108.44	12.34
WJ-PC Raw Fall Score	5.39	2.14	5.48	2.16	5.33	2.12
WJ-PC Raw Spring Score	9.30	4.16	9.49	4.02	9.18	4.25
WJ-PC SS Fall Score	95.45	10.05	94.66	10.83	95.93	9.53
WJ-PC SS Spring Score	98.92	13.60	98.37	14.54	99.25	13.01
DCCS New* Fall Score	16.90	4.22	16.75	4.43	16.99	4.09
DCCS New* Spring Score	18.09	3.70	18.10	3.79	18.09	3.65
DCCS Fall Score	1.97	0.51	1.96	0.56	1.97	0.48
DCCS Spring Score	2.14	0.54	2.15	0.54	2.14	0.54
Peg Tapping Fall Score	13.24	3.75	13.48	3.77	13.09	3.74
Peg Tapping Spring Score	14.43	2.77	14.61	2.51	14.32	2.92

*DCCS New is the sum of the correct sorts in DCCS. This scoring fully accounts for all the positives and negatives and better demonstrates the variance in this measure.

Next, we present findings from multivariate analyses. We examine the association between children's learning and program features while simultaneously controlling for children's characteristics. In the estimation models, we include information on the age of children, gender, race and ethnicity, duration of school days, low income, and IEP status. Teacher characteristics included teacher education attainment, teaching experiences, and early childhood education certification. Program features include class size, number of children with disabilities per classroom, and classroom quality either represented by the CLASS domains or the APEEC.

Results are represented in estimation results and effects sizes. Effect sizes are the estimated effect (or β) expressed in terms of standard deviations of the control group (children who did not go to pre-K). To contextualize results it is helpful to interpret them in the context of current achievement gaps. The current gap at kindergarten entry, which changes very little over the early elementary years, between the lowest income quintile and the highest income quintile is about one standard deviation nationally.

Kindergarten Entry

Table 4 presents the estimates of the associations between pre-K program participation and child learning and development, along with child characteristics. Children who participated in WV pre-K program evidence higher literacy and language development at kindergarten entry.

Low income children evidence statistically significant lower scores across all of the outcome measures. Children with an IEP (Individualized Education Program) showed significantly lower scores in literacy, language, and executive functions. Girls showed significant benefits of pre-K program participation in executive functions as measured by Peg Tapping. Children identified as White evidence small pre-K benefits on executive functions as measured by Peg Tapping. Statistically significant effects are bolded. Full estimations and sensitivity analyses are shown in appendix.

Table 4. Multivariate analyses of children's 2016 fall standard score in relation to child characteristics

	Rec. Vocabulary (PPVT/TVIP)	Literacy (WJ/WM-LW)	Language (WJ/WM-PC)	Math (WJ/WM-AP)	DCCS	PT
Treatment	0.098	**0.144***	**0.144***	0.065	0.070	-0.108
	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)	(0.07)
Age	-0.020*	**-0.069***	**-0.089***	-0.058***	0.011	**0.020***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female	0.006	0.096	-0.002	0.038	0.117	**0.154***
	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)
White	0.088	-0.074	-0.034	0.151	0.000	**0.001***
	(0.13)	(0.11)	(0.08)	(0.12)	(0.00)	(0.00)
Low Income	**-0.400***	**-0.425***	**-0.197**	**-0.378***	**-0.222**	**-0.235**
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
IEP	-0.042	**-0.356***	**-0.370***	-0.027	**-0.328**	**-0.463***
	(0.22)	(0.07)	(0.08)	(0.26)	(0.10)	(0.11)
N	967	967	967	965	968	968

* p<0.05; ** p<0.01; *** p<0.001. Note: Reference groups omitted from the estimation are Males, Non-White, middle to high income. Schools included as control. Standardized scores are used for PPVT, and WJ or WM. Errors are clustered at the classroom level.

In sum, pre-K effects persisted through to kindergarten in literacy and language. There are some small differences for the other measures, but these are not significant enough.

End of Kindergarten

Table 5 presents the estimates of the associations between the pre-K program, kindergarten program features and child characteristics, with children's development at the end of kindergarten year. The positive impact of pre-K participation on literacy and language is no longer evident by the end of this school year. In these estimations classroom features and quality are included to account for children's kindergarten experiences. There is positive relationship between reported number of inclusion children in a classroom and peg tapping. However, a negative association was found between one of the CLASS domains and math scores: lower Classroom Organization scores is related to higher math scores.
Similar to kindergarten entry estimations, a negative association is found between low income status, as well as IEP status, and all child outcomes. Gender and race/ethnicity do not show particular associations with child outcomes. Teachers master's degree is associated with higher Peg Tapping scores. Teacher's experience does not appear to matter across the board. Statistically significant effects are bolded. Full estimations and sensitivity analyses are shown in appendix.

Table 5. Multivariate analyses of children's 2017 spring standard score in relation to child and classroom characteristics and CLASS domains

	Rec. Vocabulary (PPVT/TVIP)	Literacy (WJ/WM-LW)	Language (WJ/WM-PC)	Math (WJ/WM-AP)	DCCS	PT
Treatment	-0.064	0.079	0.034	0.077	0.026	-0.045
	(0.08)	(0.06)	(0.06)	(0.07)	(0.08)	(0.08)
Age	-0.022*	**-0.085*****	**-0.104*****	-0.065***	**0.028****	**0.034*****
	(0.01)	**(0.01)**	**(0.01)**	(0.01)	**(0.01)**	**(0.01)**
Female	-0.020	0.053	0.069	-0.062	0.085	0.075
	(0.07)	(0.06)	(0.06)	(0.07)	(0.08)	(0.09)
White	0.273	-0.016	-0.004	0.224	0.347	0.066
	(0.17)	(0.11)	(0.12)	(0.18)	(0.18)	(0.16)
Low Income	**-0.378*****	**-0.427*****	**-0.361*****	**-0.336*****	**-0.256****	**-0.267****
	(0.07)	**(0.07)**	**(0.07)**	(0.07)	**(0.09)**	**(0.10)**
IEP	**-0.311****	**-0.387*****	**-0.280****	**-0.324****	**-0.382****	**-0.510****
	(0.12)	**(0.08)**	**(0.09)**	(0.10)	**(0.15)**	**(0.17)**
Teacher Education	0.043	-0.143	-0.108	-0.160	0.038	**0.215***
	(0.08)	(0.08)	(0.07)	(0.09)	(0.08)	**(0.08)**
Experience_6_10	0.092	-0.107	-0.123	-0.050	0.020	**-0.218***
	(0.10)	(0.10)	(0.09)	(0.12)	(0.10)	**(0.10)**
Experience_10more	-0.016	-0.086	-0.013	-0.208*	-0.089	**-0.245***
	(0.09)	(0.09)	(0.08)	(0.10)	(0.09)	**(0.12)**
Certification	-0.260**	**0.237***	0.020	0.151	0.032	**0.220***
	(0.09)	**(0.09)**	(0.08)	(0.10)	(0.08)	**(0.10)**
Class Size	0.003	-0.137	0.123	0.008	-0.065	0.301
	(0.14)	(0.15)	(0.12)	(0.16)	(0.16)	(0.21)
Inclusion_large	0.032	0.125	0.171	0.101	0.019	**0.172***
	(0.07)	(0.10)	(0.11)	(0.11)	(0.10)	**(0.08)**

CLASS_ES	0.127	0.148	0.023	0.075	-0.013	-0.009
	(0.09)	(0.09)	(0.09)	(0.11)	(0.08)	(0.10)
CLASS_CO	-0.032	-0.094	-0.048	-0.175*	0.001	-0.058
	(0.06)	(0.07)	(0.06)	(0.08)	(0.07)	(0.08)
CLASS_IS	-0.051	0.066	0.107	0.049	-0.046	0.002
	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)
N	806	806	805	806	805	803

* $p<0.05$; ** $p<0.01$; *** $p<0.001$. Note: Reference groups omitted from the estimation are Males, Non-White, middle to high income, Teacher Education less then Master's degree, Teacher experiences 0-5 years. Other controls are schools and indicators for missing income and missing teacher characteristics and class features. Standardized scores are used for PPVT, and WJ or WM. Errors are clustered by classroom.

Figures 1 below shows effect sizes. An effect size is the estimated association standardized by dividing it by the standard deviation of the control group. Effect sizes allow viewing results across differing measures in a common metric. The figure reports the effects at kindergarten entry, at the end of kindergarten and gain scores through kindergarten. Beneficial effects of language and literacy at the beginning of the kindergarten diminished through kindergarten year. Positive impacts of pre-K on math were consistent during the kindergarten year but the impacts were not statistically significant.

Figure 1. Effect Size for Receptive Vocabulary, Literacy, Language, and Math
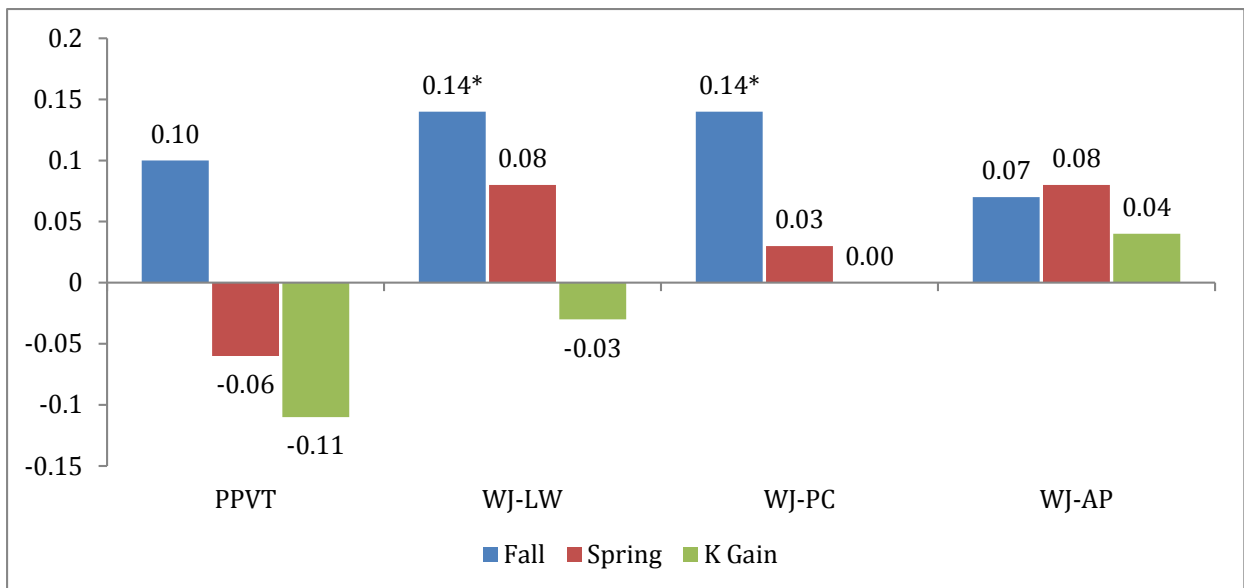
Figure 2 shows effect size for executive functioning skills. Positive pre-K impacts on DCCS decreases during the kindergarten year.

Figure 2. Effect Size for DCCS and Peg Tapping
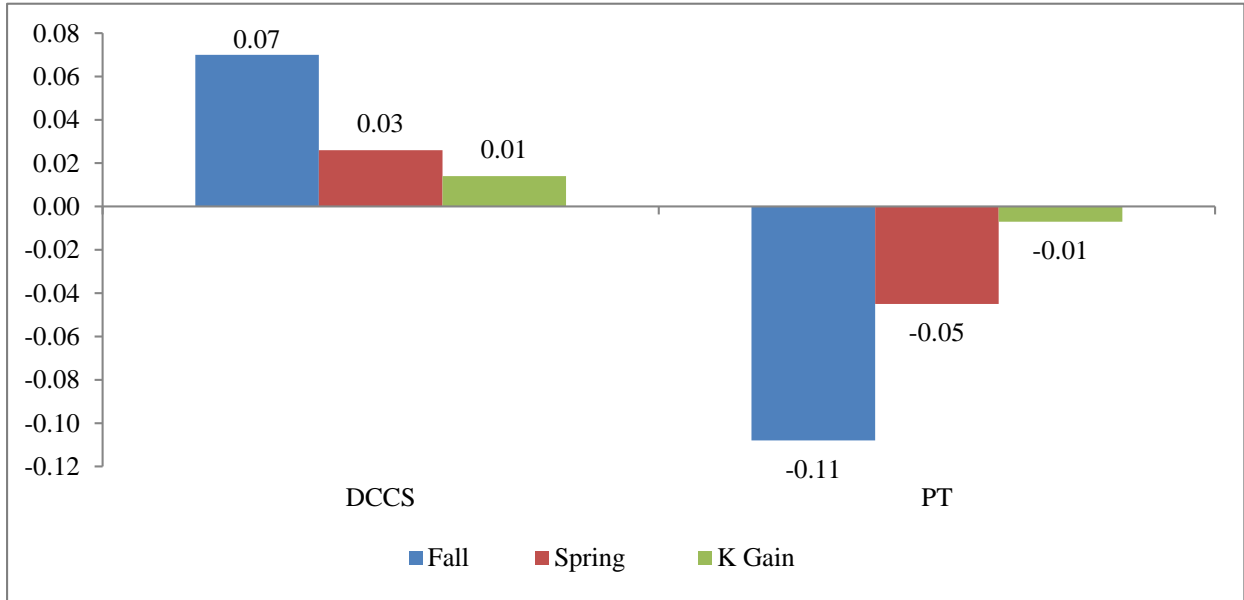
Figure 2. Effect Size for DCCS and Peg Tapping

2. Are there child subgroups (as defined by low income or child gender) that benefit more from the prekindergarten program than others?

We further analyzed the pre-K impacts only for low income children and females. Figures 3 and 4 depict results by outcome and subgroup at kindergarten entry. Figures 5 and 6 replicate this exercise at the end of kindergarten.

Children from low income families show significant benefits from pre-K education as represented in increased literacy scores. Low-income children and females evidenced positive but not statistically significant impact of pre-K. Effects from the pre-K program on receptive vocabulary is highest for females (PPVT) (See figure 3). The pre-K program's impact on DCCS was higher for low-income children (although still not significant).

Figure 3. PPVT, WJ-LW, WJ-PC and WJ-AP Effect Sizes estimations for selected subgroups at kindergarten entry
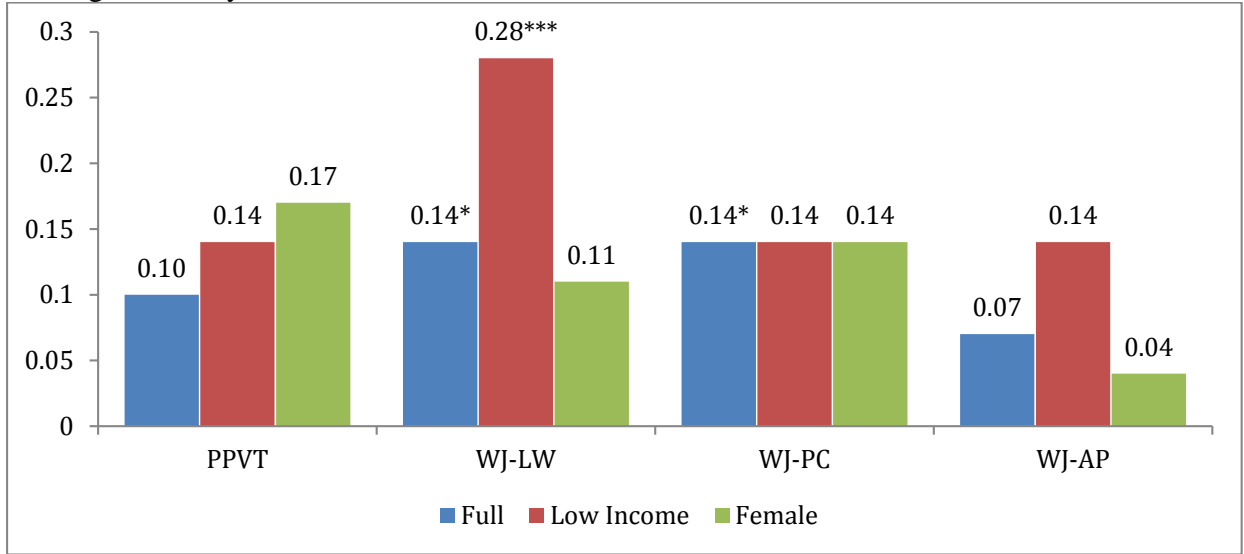
Figure 4. DCCS and PT Effect Size estimations for selected subgroups at kindergarten entry
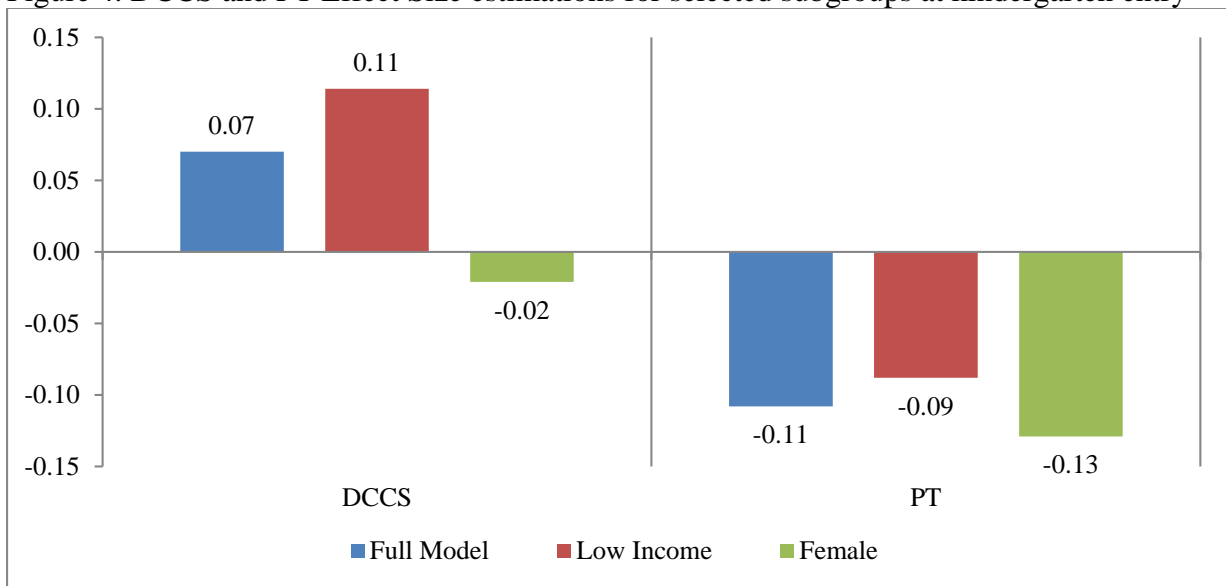
Effect sizes from estimations at the end of the kindergarten school year are presented in Figure 5 and 6 below. Low-income children show stronger effects from pre-K participation sustained through the kindergarten year, than the overall sample in receptive vocabulary, literacy, and math. Females showed higher positive program effects in both executive functioning measures. However, these gains did not retain its significance. That is, attending pre-K did not contribute to a difference in children's development growth rate throughout the kindergarten year.

Figure 5. PPVT, WJ-LW, WJ-PC and WJ-AP Effect Sizes estimations for selected subgroups at the end of kindergarten
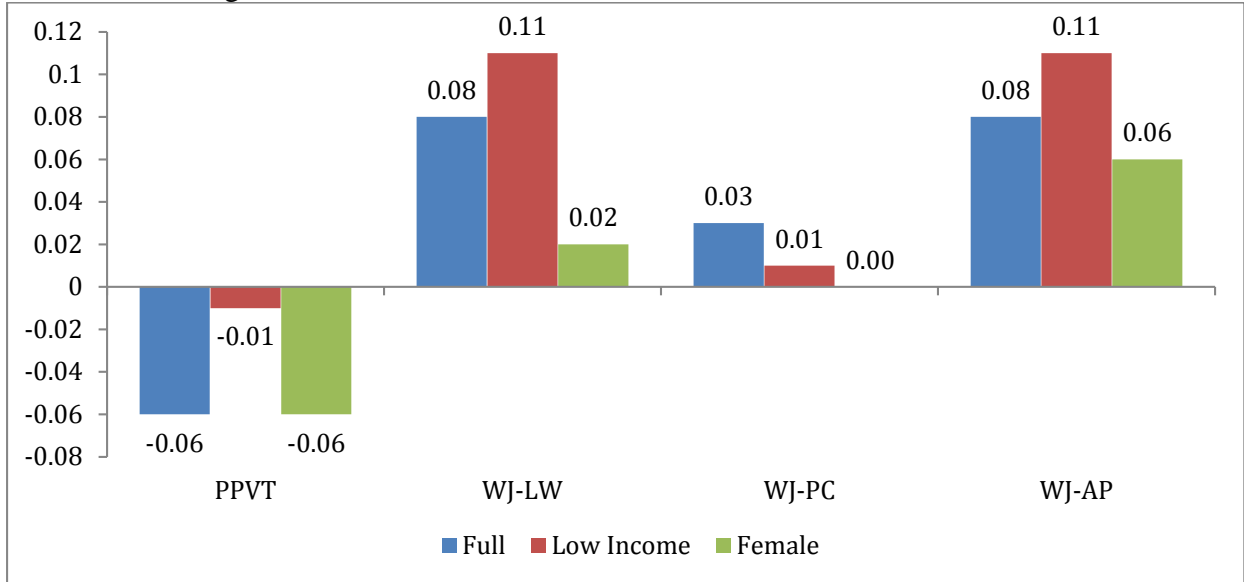
Figure 6. DCCS and PT Effect Size estimations for selected subgroups at the end of kindergarten
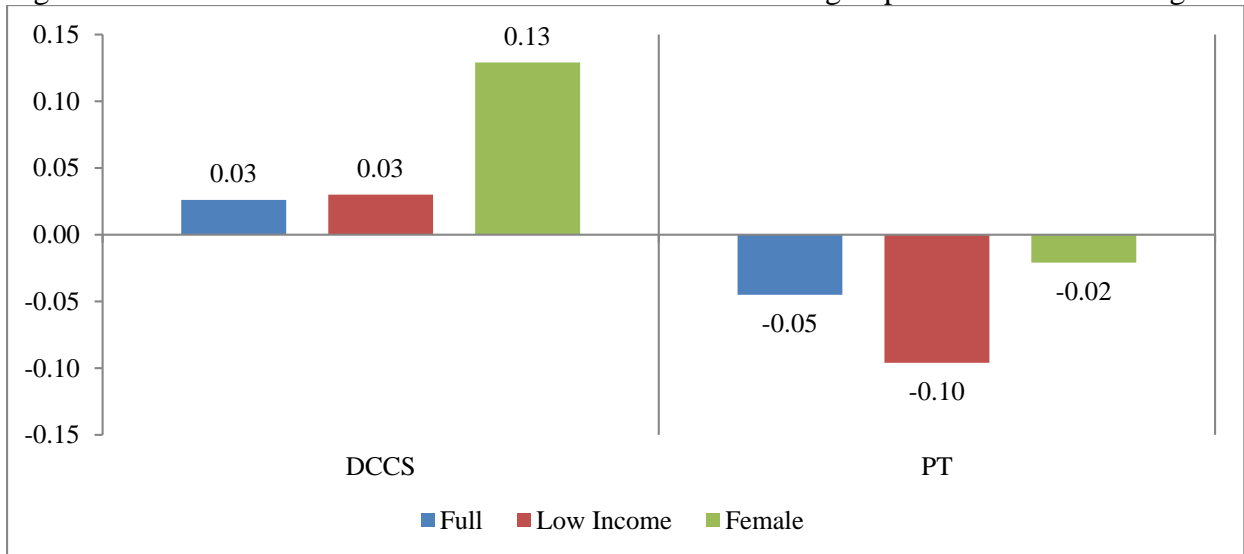
3. What is the overall observed quality of preschool and kindergarten classrooms in 2016-17?

Pre-K Classrooms

Pre-K ECERS-3 Results Spring of 2017

Scores for the 127 classrooms that were observed using the ECERS-3 are presented in Table 6 and Figure 7 and in contrast to the previous year's scores. Highest scoring subscale was Interaction (Mean = 4.32), rated in minimum to good level.

Table 6. ECERS-3 Subscales and Overall Means and Ranges

ECERS-R Item and Subscales	Spring 2016 (N=130)				Spring 2017 (N=127)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Overall	**4.04**	**(0.99)**	**1.60**	**6.00**	**3.51**	**(0.84)**	**1.24**	**5.47**
Space and Furnishings	3.95	(1.03)	1.00	6.29	3.46	(0.92)	1.00	6.00
Personal Care Routines	3.99	(1.20)	1.00	6.25	3.65	(0.99)	1.25	6.00
Language and Literacy	4.48	(1.33)	1.00	6.80	3.86	(1.12)	1.00	6.40
Learning Activities	3.46	(1.08)	1.09	6.20	2.82	(0.94)	1.27	5.60
Interaction	4.81	(1.38)	1.00	7.00	4.32	(1.29)	1.20	6.80
Program Structure	4.41	(1.44)	1.00	7.00	3.99	(1.14)	1.00	6.33

Figure 7 illustrates the overall mean scores distribution for the 127 classrooms observed using the ECERS tool. Majority of the classrooms were in the minimum to good level (3.00-5.00; 66%), followed by inadequate level (1.00-2.99; 28%). Only six percent of classrooms scored at the excellent level (5.01-7.00). There was a decrease in the overall ECERS-3 scores between 2016 and 2017.

Figure 7. ECERS-3 Distributions, Spring of 2016 and 2017
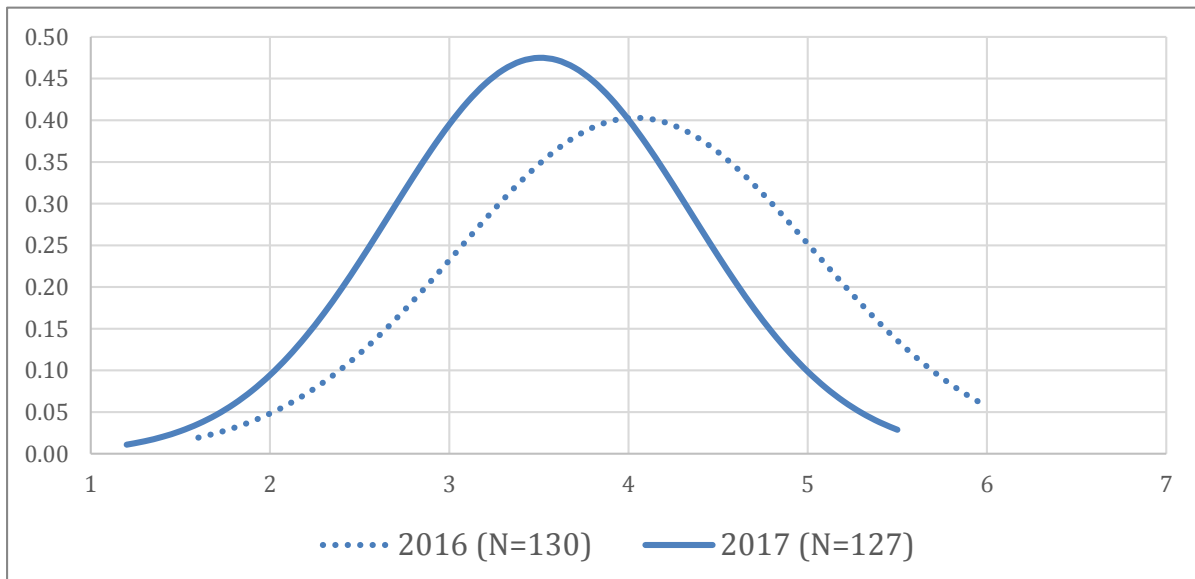
Figure 8 provides ECERS-3 scores for WV 2016 and 2017 observations as well as ECERS scores from other studies: GA, WA, PA, three-state, NJ Abbott districts (Spring 2017 and Spring 2016) for comparison purposes.

Figure 8. ECERS-3 scores and subscale scores for WV Pre-K and for other programs
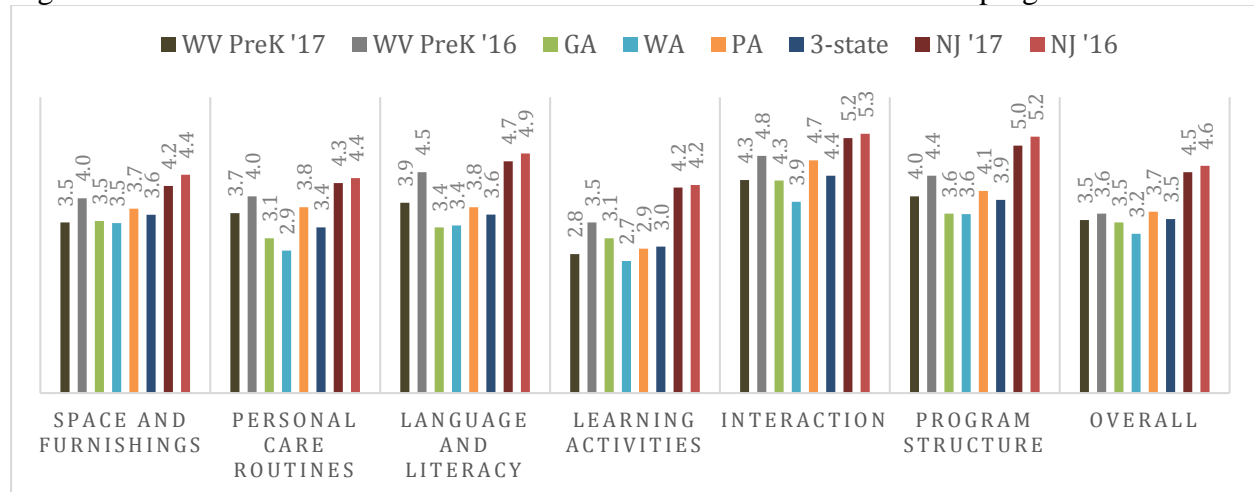
Pre-K CLASS Results Spring 2017

The scores presented below reflect overall means for the 123 pre-K classrooms that were observed using the CLASS instrument. Pre-K CLASS mean scores were 5.93 for Emotional Support (ES), 5.32 for Classroom Organization (CO) and 2.67 for Instructional Supports (IS).

Table 7. CLASS Domains Means and Ranges

CLASS Domains	Spring 2016 (N=105)				Spring 2017 (123)			
	Mean	(SD)	Min	Max	Mean	(SD)	Min	Max
Emotional Support	5.66	(0.90)	2.35	6.95	5.93	(0.75)	3.25	7.00
Classroom Organization	5.09	(1.16)	1.33	6.87	5.32	(0.86)	2.93	7.00
Instructional Support	2.65	(0.83)	1.13	5.33	2.67	(0.88)	1.20	5.47

Figure 9 illustrates the distribution of CLASS scores for each dimension and across years. Similar to the score patterns found in national evaluation of pre-K classrooms (OHS, 2015), the high range of quality are found in the domain of Emotional Support and Classroom Organization, and low-to-mid range of quality found in the Instructional Support domain. The mean scores from all of the three CLASS domains are slightly lower than those in the national evaluations (OHS, 2015) but, higher than scores from previous WV pre-K classrooms.

Figure 9. Pre-K CLASS distributions with Means and Standard deviations
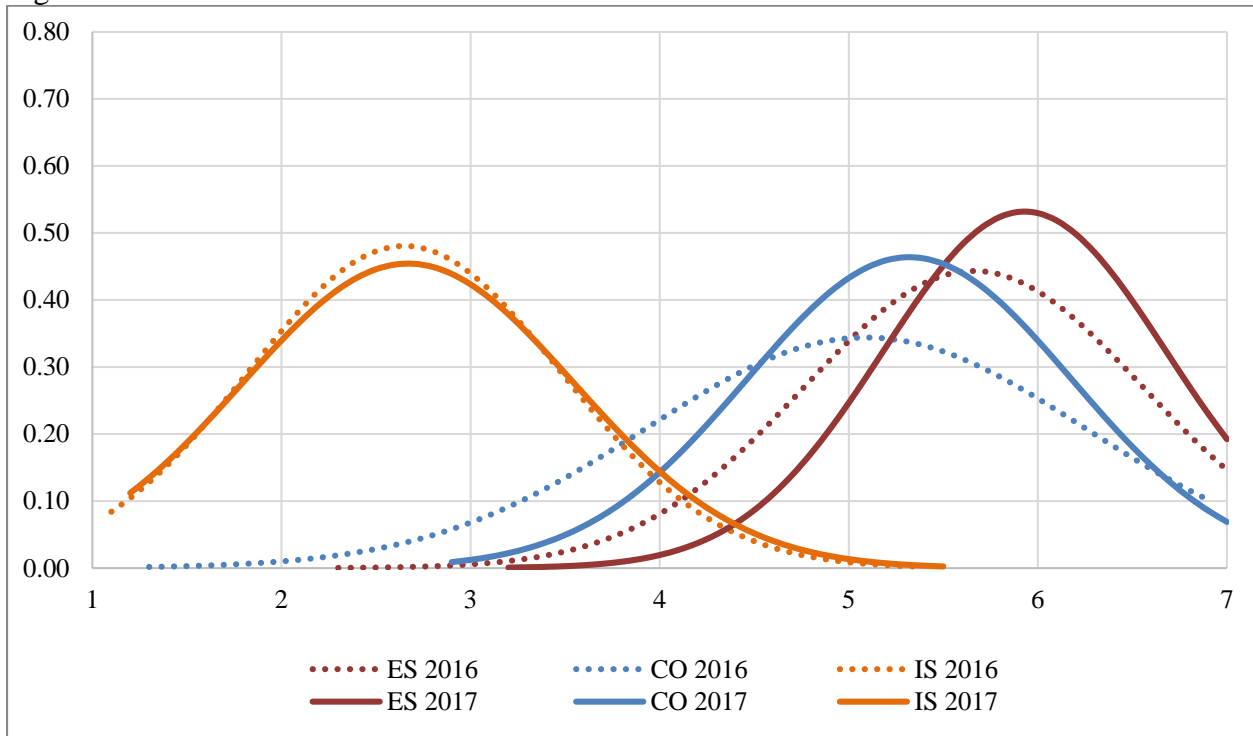
Figure 10 below illustrates West Virginia's pre-K CLASS scores together with those from various other programs in the U.S. as means of comparisons. High-quality programs were included in this comparison. In terms of ES and CO, West Virginia is in the mid-range across these programs. For IS however, West Virginia is among the lowest scoring.

Figure 10. CLASS Scores for various programs
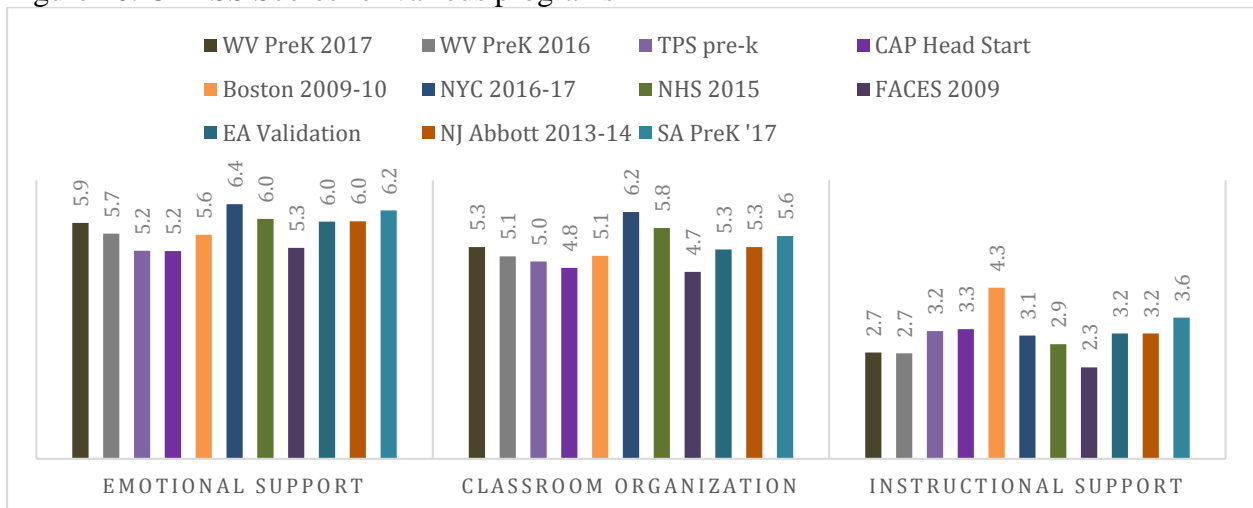
Pre-K ECERS-3 and CLASS Scores by Auspice and Number of IEPs

Table 8 represents the mean scores for each ECERS-3 subscale and the overall score and each CLASS domain delineated by auspice and by number of IEPs reported in the classrooms. For auspice, classrooms were grouped into three categories: private center, public school, or Head Start center. For IEPs, classrooms were categorized by the number of IEPs in the classroom: three or fewer and four or more.

On average, public classrooms outperform private and Head Start classroom, with average ECERS 3 levels of 3.59, CLASS ES levels of 6.04, CLASS CO levels of 5.49 and CLASS IS levels at 2.79.

In addition, we assessed differences in higher versus lower inclusion classrooms. Greater inclusion classrooms are averaging slightly higher scores in ECERS-3 and its subscale. Scores from CLASS domains were almost identical between these classrooms.

Table 8. ECERS-3 and CLASS Scores Categorized by Auspice and Number of IEPs

	Pre-K				
	Auspice			Number of IEP	
	Private	**Public**	**Head Start**	**0-3**	**4+**
ECERS-3	*N=14*	*N=93*	*N=20*	*N=76*	*N=51*
Space and Furnishings	3.40	3.46	3.49	3.28	3.72
Personal Care Routines	3.11	3.71	3.73	3.45	3.94
Language and Literacy	3.43	4.00	3.56	3.67	4.16
Learning Activities	2.36	2.90	2.74	2.71	2.99
Interaction	4.00	4.43	4.07	4.02	4.77
Program Structure	3.60	4.19	3.35	3.73	4.39
Overall	3.15	3.59	3.36	3.32	3.79
CLASS	*N=14*	*N=89*	*N=20*	*N=72*	*N=48*
Emotional Support	5.35	6.04	5.80	5.92	5.94
Classroom Organization	4.80	5.49	4.94	5.32	5.30
Instructional Support	2.26	2.79	2.41	2.66	2.67

Note: the number of IEPs was missing for three classrooms assessed using just the CLASS tool, therefore the total N for CLASS in the Number of IEPs category is 120 rather than 123.

Pre-K *Teacher Demographic Data*

Teacher demographic information gathered during the administration of the classroom observations and presented here. A total of 117 lead teachers and 77 assistant teachers participated in the survey.

Most lead teachers report having attained either a BA or a master's degree. Assistant teachers report on average lower levels of educational attainment, with 86% having attained either a high school diploma or some college or associate's degree, and 12% having attained a BA or a higher degree. About 65% of lead teachers have five years or more experiences in teaching, and almost all of the lead teachers, and about half of the assistant teachers, have certification.

Table 9. Pre-K Lead and Assistant Teacher Demographic Data

		Lead teacher		Assistant Teacher	
		N = 117	%	N =77	%
Teacher Education	GED	-	-	2	2.6%
	High School Diploma	-	-	18	23.4%
	Some college or AA	6	5.1%	48	62.3%
	Bachelor's Degree	53	45.3%	8	10.4%
	Master's Degree or higher	58	49.6%	1	1.3%
Experience in Early Childhood	0 - 5 years	39	33.3%	31	40.3%
	6 - 10 years	30	25.6%	19	24.7%
	More than 10 years	46	39.3%	21	27.3%
	Missing	2	1.7%	6	7.8%
Certification	Yes	113	96.6%	38	49.4%
	No	4	3.4%	39	50.6%

Kindergarten Classrooms

Kindergarten APEEC Results Spring 2017

Scores for the 140 kindergarten classrooms observed using the APEEC are presented in Table 10 and Figure 11. A decrease in kindergarten quality is observed in the Spring of 2017.

Table 10. APEEC Overall Mean and Ranges

APPEC	Spring 2016 (N=98)				Spring 2017 (N=140)			
	Mean	(SD)	Min	Max	Mean	(SD)	Min	Max
APEEC Total	3.83	(0.75)	2.31	5.50	3.55	(0.63)	2.25	5.13

Figure 11. APEEC distribution of scores, spring of 2016 and 2017
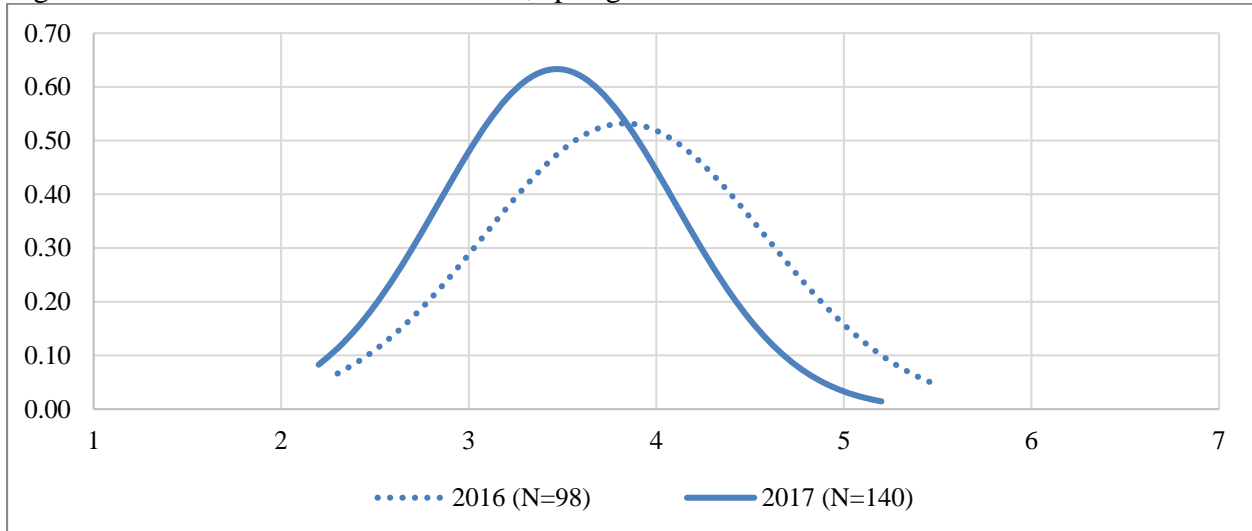
Table 11 below shows WV's kindergarten APEEC scores as well as those from 2 NJ studies and a Kentucky study, all of which included kindergarten classrooms, for comparison purposes. WV kindergarten scores for 2017 are the lowest among the group, even though none of these average scores reach the "good" levels (5).

Table 11. APEEC scores across various studies

	NJ K-3	NJ 2008	Kentucky	North Carolina	WV 2016 Spring K	WV 2017 Spring K
Grades	K-3	K	K-3	K	K	**K**
N	123	135	69	88	98	**140**
Overall	3.90	3.96	3.67	3.6	3.83	**3.55**

Kindergarten CLASS Results Spring 2017

Overall means for the 140 kindergarten classrooms that were observed using the CLASS instrument are reported below. Kindergarten CLASS mean scores were 5.05 for Emotional Supports (ES), 4.81 for Classroom Organization (CO) and 2.05 for Instructional Supports (IS). Scores from current kindergarten classrooms are lower than scores found in last year's kindergarten classrooms (see table 12). Scores for ES are just above the threshold for "good" while CO scores are below 5, and IS scores are below the threshold of 3.

Table 12. CLASS Domains Means and Ranges

CLASS Domains	Spring 2016 (N=98)				Spring 2017 (N=140)			
	Mean	(SD)	Min	Max	Mean	(SD)	Min	Max
Emotional Support	5.49	(0.62)	3.45	6.55	5.05	(0.66)	2.75	6.25
Classroom Organization	5.14	(0.70)	3.33	6.53	4.81	(0.81)	2.27	6.40
Instructional Support	2.23	(0.79)	1.00	4.13	2.06	(0.72)	1.00	4.93

Kindergarten APEEC and CLASS Scores Distributed by IEPs

Table 13 represents the mean scores for APEEC overall score and CLASS domain by number of IEPs reported in the classrooms categorized as three or fewer versus four or more. Eight classrooms did not report IEP numbers and are not included in this table.

APEEC average scores are higher in classrooms with fewer IEPs reported. For CLASS, different patterns are found, with CLASS ES and IS being higher and CLASS CO being lower in classrooms with fewer IEPs.

Table 13. APEEC and CLASS Scores by Number of IEPs, N = 140

	Kindergarten IEP	
	0-3 N = 103	4+ N = 29
APEEC		
Overall	3.56	3.38
CLASS		
Emotional Support	5.07	4.97
Classroom Organization	4.78	4.96
Instructional Support	2.13	1.83

Kindergarten Teacher Demographic Data.

Table 14 presents kindergarten teacher data gathered via survey during the administration of the classroom observations. Data were collected from 128 lead teachers. Teachers in current kindergarten classrooms report similar levels of educational attainment, but slightly lower levels of experience than teachers from the previous year's kindergarten classrooms.

Table 14. Kindergarten Lead and Assistant Teacher Demographic Data

		Lead Teacher		Other	
		N = 128	%	N = 5	%
Teacher Education	Some college or less	-	-	-	-
	Bachelor's Degree	57	44.53%	5	100.00%
	Master's Degree or higher	71	55.47%	-	-
Experience in Early Childhood	0 - 5 years	35	27.34%	3	60.00%
	6 - 10 years	34	26.56%	-	-
	More than 10 years	57	44.53%	2	40.00%
Certification	Yes	123	96.09%	4	80.00%
	No	5	3.91%	1	20.00%

Quality experienced by the longitudinal cohort

Table 15 and Figure 12 present CLASS domain scores and distributions for the longitudinal sample. CLASS scores from previous pre-K classrooms and current kindergarten classrooms are reported and illustrated. Children in current kindergarten classrooms experienced lower classroom quality than their pre-K year across all CLASS domains. The low CLASS IS could strongly contribute to the trends in kindergarten for the pre-K group. This is explained further in the discussion section below.

Table 15. CLASS Domains for Longitudinal Cohort

	Pre-K 2016 (N=105)				K 2017 (N=140)			
	Mean	(SD)	Min	Max	Mean	(SD)	Min	Max
Emotional Support	5.66	(0.90)	2.35	6.95	5.05	(0.66)	2.75	6.25
Classroom Organization	5.09	(1.16)	1.33	6.87	4.81	(0.81)	2.27	6.40
Instructional Support	2.65	(0.83)	1.13	5.33	2.06	(0.72)	1.00	4.93

Figure 12. CLASS Mean Scores for Longitudinal Cohort
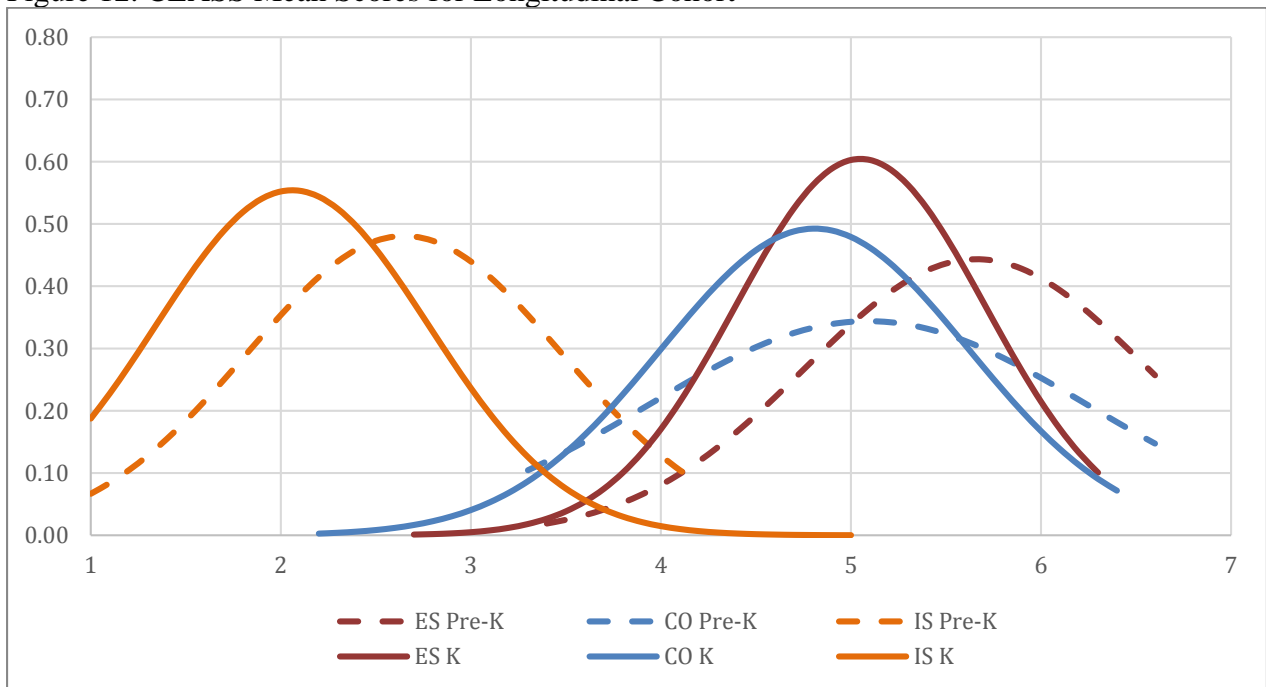
Summary

This report presents the second year (2016-17) findings from the WV Universal Pre-K Program evaluation. We examined the impact of West Virginia's universal preschool program on children's language, math, literacy, and executive functioning skills at kindergarten entry as well as at the end of kindergarten year. We also analyzed whether results varied for low-income children and females. The classroom quality children experienced in pre-K and kindergarten classrooms were also reported. The findings in quality were compared with previous West Virginia classrooms and findings from other studies.

This study demonstrated positive impacts of pre-K program on children's learning and development at kindergarten entry. This study found positive effects of pre-K participation on literacy and language at kindergarten entry. The positive association between pre-K participation and literacy was greater for low-income children. Girls showed beneficial impact of pre-K on receptive vocabulary. When estimates were repeated at the end of kindergarten, the positive impacts of pre-K no longer appeared. Low-income children and girls also showed similar results, with the positive effects of pre-K no longer present.

Pre-K classrooms observations showed moderate levels of quality as measured by the ECERS-3, and the CLASS Emotional Support and Classroom Organization, but low levels of quality on the CLASS Instructional Support domain. Kindergarten classrooms showed minimum levels of quality as measured by the APEEC. Kindergarten classrooms show moderate-low levels of quality on the CLASS Emotional Support and Classroom Organization domains, and low levels of quality on the CLASS Instructional Support domain.

The results presented here are consistent with most of the literature in preschool programs, where rigorous evidence has shown positive short-terms impacts of such programs on children's development (Yoshikawa, et. al, 2013). The results for quality of classrooms also showed similar patterns to what the literature have found, with instructional supports needing the most improvements. Yoshikawa, et. al (2013) also report that over the elementary grades children that have attended versus that did not attend preschool tend to converge (p.9), but they also see evidence in the literature that despite this, there is evidence of effects on outcomes in adulthood. There is no clarity if this may be or not the case for WV. As for other evaluation, the reasons for converges of test scores may likely be due to the low quality elementary experienced that was observed for the longitudinal group. Such types of experiences "fail to build on the gains created by early childhood education" (p.9). Alternatively, another explanation for convergence proposed by the authors is that having children ready and with positive experiences entering kindergarten allows elementary teachers to focus on non-attenders (less ready) children. This implies teaching to "the bottom".

References

Barnett, W. S., Jung, K., Young, M., & Frede, E. C. (2013). Abbott preschool program longitudinal effects study: Fifth grade follow-up. New Brunswick, NJ: National Institute for Early Education Research, 10, 2001-2004.

Barnett, W. S., Friedman-Krauss, A. H., Weisenfeld, G. G., Horowitz, M., Kasmin, R., & Squires, J. H. (2017). The State of Preschool 2016: State Preschool Yearbook. New Brunswick, NJ: National Institute for Early Education Research.

Barnett, W.S., Jung, K., Frede, E., Hustedt, J.T., Howes, C., & Daniel-Echols, M. (2013). Effects of Eight State Prekindergarten Programs on Early Learning: A Regression-Discontinuity Analysis. 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 and Professional Development (Unpublished). Large Scale Psychometric Assessment of ECERS 3. Seattle: University of Washington

Copple, C. & Bredekamp, S. (2009). Developmentally appropriate practice in early childhood programs serving children from birth through age 8. National Association for the Education of Young Children. 1313 L Street NW Suite 500, Washington, DC 22205-4101.

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., ... & Henry, G. T. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. Child development, 78(2), 558-580.

Early, D. M., Sideris, J., Neitzel, J., LaForett, D. R., & Nehler, C. G. (2018). Factor structure and validity of the Early Childhood Environment Rating Scale–Third Edition (ECERS-3). Early Childhood Research Quarterly, *44*, 242-256.

Edvance Research (2016). Pre-K 4 SA Evaluation Report. YEAR 3. Final Report Submitted to Early Childhood Education Municipal Development Corporation. San Antonio, TX: Author.

Friedman-Krauss, A. H., Barnett, W. S., Weisenfeld, G. G., Kasmin, R., DiCrecchio, N., & Horowitz, M. (2018). *The state of preschool 2017: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.

Harms, T., Clifford, R. M., & Cryer, D. (2014). Early childhood environment rating scale. Teachers College Press.

Hill, C. J., Gormley, W. T., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? Early Childhood Research Quarterly, 32, 60-79.

Huang, F. L., Invernizzi, M. A., & Drake, E. A. (2012). The differential effects of preschool: Evidence from Virginia. Early Childhood Research Quarterly, 27(1), 33-45.

Hustedt, J. T., Barnett, W. S., Jung, K., & Thomas, J. (2007). The effects of the Arkansas Better Chance Program on young children's school readiness. National Institute for Early Education Research.

Lipsey, M. W., Weiland, C., Yoshikawa, H., Wilson, S. J., & Hofer, K. G. (2015). The Prekindergarten Age-Cutoff Regression-Discontinuity Design Methodological Issues and Implications for Application. Educational Evaluation and Policy Analysis, 37(3), 296-313.

Maxwell, K. L., McWilliam, R. A., Hemmeter, M. L., Ault, M. J., & Schuster, J. W. (2002). Predictors of developmentally appropriate classroom practices in kindergarten through third grade. Early Childhood Research Quarterly, 16(4), 431-452.

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

Nores, M., Barnett, W.S., Joseph, G., Stull, S., Jung, K. & Soderberg, J.S. (2017). Year 2 report: Seattle Pre-k program evaluation. New Brunswick, NJ: National Institute for Early Education Research & Seattle, WA: Cultivate Learning.

NYC Department of Education (2017). Pre-K Program Assessments Early Childhood Environmental Rating Scale –Revised (ECERS-R) and Classroom Assessment Scoring System (CLASS) Release. New York: Author. Available at http://schools.nyc.gov/NR/rdonlyres/5FEA3D5B-E615-4E16-83A8-C4E58A4D6F02/0/201516ProgramAssessmentResultsSummary.pdf

Office of Head Start. U.S. A National Overview of Grantee CLASS® Scores in 2015. Washington, D.C.: Department of Health and Human Services. Available at http://eclkc.ohs.acf.hhs.gov/hslc/data/class-reports/docs/national-class-2015-data.pdf

PAKEYS (Unpublished). What does the data tell us? The evolution of environment rating scale (ERS) use within QRIS. Available at https://qrisnetwork.org/sites/all/files/conference-session/resources/651DataTellsUs.pdf.

Peisner-Feinberg, E.S. & Schaaf, J.M. (2008). Evaluation of the North Carolina More at Four Pre-kindergarten Program Year 6 Report (July 1, 2006–June 30, 2007). Chapel Hill, NC: UNC, FPG Child Development Institute.

Peisner-Feinberg, E.S. & Schaaf, J. M. (2011). Effects of the North Carolina More at Four Pre-Kindergarten Program on Children's School Readiness Skills: Summary of Key Findings. Chapel Hill: The University of North Carolina, FPG Child Development Institute.

Peisner-Feinberg, Schaaf, J.M. & LaForett, D.R. (2013). Children's Growth and Classroom Experiences in Georgia's Pre-K Program: Findings from the 2011-2012 Evaluation Study. Chapel Hill: The University of North Carolina, FPG Child Development Institute.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). Classroom Assessment Scoring System: Manual Pre-K. Education Review.

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.

Seplocha, H. & J. Strasser (2008) A Snapshot of Quality in Abbott Kindergarten Classrooms. NJ: William Paterson University.

Wechsler, M., Kirp, D., Ali, T. T., Gardner, M., Maier, A., Melnick, H., & Shields, P. M. (2016). The Road to High-Quality Early Learning: Lessons from the States. Policy Brief, Palo Alto, Washington DC: Learning Policy Institute. Accessed September, 1, 2016.

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

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.

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

Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Tarullo, L., & Burchinal, M. (2010). Quality Dosage Thresholds and Features in Early Childhood Settings. A Review of the Literature (No. cdea2df43cdf4b10bb74af4189cbab8d). Mathematica Policy Research.

Zelazo, P. D. (2006). The dimensional change card sort (DCCS): A method of assessing executive function in children. Nature Protocols, 1, 297-301.

Appendix. Estimation: Full set of results and sensitivity analysis.

In the following appendix tables, we present effect sizes for estimations from standard scores and raw scores including various covariates. In the estimations, child characteristics such as age, gender, race/ethnicity, low income, and IEP status are included in all of the estimation model. First, effect size from full sample presented, followed by group of low income and female. Effect sizes from kindergarten entry, at the end of kindergarten, and gain scores during the kindergarten will be presented in order. Preferred models are those summarized in the main document. These tables provide sensitivity analyses.

Estimations from the Full Sample

Table A1. Effect Size at the Kindergarten Entry

Fall 2016						
	M1	**M2**	**M3**	**M4**	**M5**	**M6**
Receptive Vocabulary	0.11	0.10	0.11	0.13*	0.11	0.12*
Math	0.07	0.07	0.06	0.09	0.08	0.08
Literacy	0.11	0.14*	0.10	0.1	0.13*	0.09
Language	0.06	0.14*	0.07	0	0.09	0.01
DCCS	0.1	0.07	0.08			
PT	-0.03	-0.11	-0.05			
Standard score	x	x	X			
Raw score				x	x	x
no f.e.	x			x		
with school f.e.		x			x	
with county f.e.			X			x

Table A2. Effect Size for Standard Score at the End of Kindergarten

Spring 2017													
Standard score	**M1**	M2	**M3**	M4	**M5**	M6	**M7**	M8	**M9**	M10	**M11**	M12	**M13**
Receptive Vocabulary	-0.04	-0.05	-0.1	-0.05	-0.04	-0.10	-0.06	-0.03	-0.06	-0.04	-0.02	-0.06	-0.03
Math	0.11	0.11	0.08	0.11	0.11	0.07	0.10	0.14	0.08	0.16	0.13	0.08	0.12
Literacy	0.08	0.08	0.09	0.08	0.07	0.07	0.06	0.10	0.08	0.09	0.11[*]	0.08	0.10
Language	0.04	0.04	0.05	0.03	0.02	0.04	0.02	0.04	0.03	0.03	0.04	0.03	0.03
DCCS	0.07	0.06	0.03	0.06	0.07	0.04	0.06	0.05	0.03	0.04	0.04	0.03	0.04
PT	-0.02	-0.04	-0.06	-0.04	-0.02	-0.04	-0.02	-0.01	-0.05	-0.02	-0.01	-0.04	-0.01
no f.e.	x	x			x				x		x		
with school f.e.			x			x			x			x	
with county f.e.				x			x			x			x
with % inclusion		x	x	x	x	x	x	x	x	x	x	x	x
with teacher characteristics					x	x	x	x	x	x	x	x	x
with CLASS dimensions									x	x	x		
with APEEC											x	x	x

Table A3. Effect Size for Raw Score at the End of Kindergarten

Spring 2017													
Raw score	**M1**	M2	**M3**	M4	**M5**	M6	**M7**	M8	**M9**	M10	**M11**	M12	**M13**
Receptive Vocabulary	-0.01	-0.02	-0.07	-0.03	-0.02	-0.06	-0.03	-0.03	-0.06	-0.04	-0.02	-0.06	-0.03
Math	0.11	0.12	0.08	0.11	0.11	0.06	0.10	0.14	0.07	0.12	0.14	0.07	0.12
Literacy	0.10	0.10	0.10	0.09	0.08	0.07	0.07	0.12	0.09	0.10	0.12	0.09	0.11
Language	0.06	0.06	0.06	0.05	0.04	0.04	0.03	0.06	0.05	0.04	0.05	0.04	0.05
no f.e.	x	x			x				x		x		
with school f.e.			x			x			x			x	
with county f.e.				x			x			x			x
with % inclusion		x	x	x	x	x	x	x	x	x	x	x	x

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
with teacher characteristics					x	x	x	x	x	x	x	x	x
with CLASS dimensions									x	x	x		
with APEEC											x	x	x

Table A4. Effect Size for Standard Score Gain During Kindergarten

Fall 2016-Spring 2017													
Standard score	**M1**	M2	**M3**	M4	**M5**	M6	**M7**	M8	**M9**	M10	**M11**	M12	**M13**
Receptive Vocabulary	-0.10	-0.11	-0.13	-0.11	-0.10	-0.13*	-0.11	-0.09	-0.11	-0.10	-0.10	-0.11	-0.11
Math	0.07	0.07	0.06	0.07	0.06	0.05	0.06	0.06	0.04	0.04	0.06	0.05	0.05
Literacy	-0.02	-0.02	-0.03	-0.02	-0.04	-0.04	-0.04	-0.03	-0.03	-0.04	-0.02	-0.03	-0.02
Language	0.03	0.03	0.02	0.02	0.01	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.01
DCCS	0.04	0.04	0.02	0.04	0.04	0.02	0.04	0.02	0.01	0.01	0.02	0.02	0.02
PT	-0.01	-0.03	-0.01	-0.03	-0.01	0.01	-0.01	0.00	0.01	-0.01	0.00	0.01	-0.01
no f.e.	x	x			x				x		x		
with school f.e.			x		x				x		x		
with county f.e.				x			x			x			x
with % inclusion		x	x	x	x	x	x	x	x	x	x	x	x
with teacher characteristics					x	x	x	x	x	x	x	x	x
with CLASS dimensions									x	x	x		
with APEEC											x	x	x

Table A5. Effect Size for Raw Score Gain during Kindergarten

Fall 2016-Spring 2017													
Raw score	**M1**	M2	**M3**	M4	**M5**	M6	**M7**	M8	**M9**	M10	**M11**	M12	**M13**
Receptive Vocabulary	-0.10*	-0.11*	-0.12*	-0.11*	-0.10	-0.12*	-0.11*	-0.12*	-0.12*	-0.13*	-0.12*	-0.12*	-0.12*
Math	0.06	0.06	0.05	0.06	0.06	0.03	0.05	0.05	0.02	0.03	0.05	0.03	0.04
Literacy	0.00	-0.01	-0.02	-0.01	-0.03	-0.04	-0.03	-0.02	-0.03	-0.04	-0.01	-0.03	-0.02
Language	0.06	0.06	0.03	0.04	0.03	0.01	0.02	0.04	0.01	0.02	0.04	0.00	0.03

no f.e.	x	x			x			x			x	
with school f.e.			x			x			x			x
with county f.e.				x			x			x		x
with % inclusion		x	x	x	x	x	x	x	x	x	x	x
with teacher characteristics					x	x	x	x	x	x	x	x
with CLASS dimensions								x	x	x		
with APEEC										x	x	x

Estimations from Low Income only

Table A6. Effect Size at the Kindergarten Entry: Low Income only

Fall 2016						
	M1	M2	M3	M4	M5	M6
Receptive Vocabulary	0.16*	0.14	0.15*	0.18*	0.18*	0.17*
Math	0.1	0.14	0.11	0.13	0.17	0.14
Literacy	0.21*	0.28***	0.22**	0.19*	0.26**	0.21**
Language	0.05	0.14	0.06	-0.02	0.07	-0.01
DCCS	0.14	0.11	0.12			
PT	-0.03	-0.09	-0.03			
Standard score	x	x	x			
Raw score				x	x	x
no f.e.	x			x		
with school f.e.		x			x	
with county f.e.			x			x

Table A7. Effect Size for Standard Score at the End of Kindergarten: Low Income only

2017 Spring							
Standard score	**M1**	M2	**M3**	M4	**M5**	M6	**M7**
Receptive Vocabulary	-0.01	-0.03	-0.01	-0.01	-0.02	-0.02	-0.02
Math	0.11	0.14	0.11	0.14	0.14	0.11	0.14
Literacy	0.11	0.14	0.11	0.14	0.15*	0.11	0.15*
Language	-0.00	-0.02	0.01	-0.01	-0.01	0.01	-0.00
DCCS	0.08	0.05	0.03	0.06	0.04	0.03	0.05
PT	-0.08	-0.05	-0.10	-0.05	-0.04	-0.10	-0.04
no f.e.	x	x			x		
with school f.e.			x			x	
with county f.e.				x			x
with % inclusion		x	x	x	x	x	x
with teacher characteristics		x	x	x	x	x	x
with CLASS dimensions		x	x	x			
with APEEC					x	x	x

Table A8. Effect Size for Raw Score at the End of Kindergarten: Low Income only

2017 Spring							
Raw score	**M1**	M2	**M3**	M4	**M5**	M6	**M7**
Receptive Vocabulary	-0.02	-0.03	-0.02	-0.01	-0.02	-0.02	-0.01
Math	0.11	0.15	0.10	0.14	0.14	0.10	0.15
Literacy	0.12	0.17	0.13	0.17	0.17*	0.13	0.18[*]
Language	0.00	-0.02	0.02	-0.01	-0.02	0.01	-0.00
no f.e.	x	x			x		
with school f.e.			x			x	
with county f.e.				x			x
with % inclusion		x	x	x	x	x	x
with teacher characteristics		x	x	x	x	x	x
with CLASS dimensions		x	x	x			
with APEEC					x	x	x

Table A9. Effect Size for Standard Score Gain during Kindergarten: Low Income only

Fall 2016-Spring 2017							
Standard score	**M1**	M2	**M3**	M4	**M5**	M6	**M7**
Receptive Vocabulary	-0.11	-0.11	-0.07	-0.10	-0.16	-0.07	-0.10
Math	0.06	0.06	0.03	0.04	0.06	0.03	0.05
Literacy	-0.05	-0.04	-0.08	-0.05	-0.03	-0.09	-0.04
Language	-0.00	-0.02	-0.01	-0.02	-0.02	-0.01	-0.01
DCCS	0.11	0.02	-0.01	0.02	0.00	-0.02	0.01
PT	0.03	-0.04	-0.06	-0.05	-0.04	-0.07	-0.05
no f.e.	x	x			x		
with school f.e.			x			x	
with county f.e.				x			x
with % inclusion		x	x	x	x	x	x
with teacher characteristics		x	x	x	x	x	x
with CLASS dimensions		x	x	x			
with APEEC					x	x	x

Table A10. Effect Size for Raw Score Gain during the Kindergarten: Low Income only

Fall 2017-Spring 2018							
Raw score	**M1**	M2	**M3**	M4	**M5**	M6	**M7**
Receptive Vocabulary	-0.21***	-0.14*	-0.10	-0.12	-0.13	-0.10	-0.12
Math	0.05	0.04	0.02	0.03	0.05	0.02	0.04
Literacy	-0.03	-0.03	-0.08	-0.05	-0.02	-0.09	-0.03
Language	0.04	-0.02	0.01	-0.02	-0.02	-0.00	-0.01
no f.e.	x	x			x		
with school f.e.			x			x	
with county f.e.				x			x
with % inclusion		x	x	x	x	x	x
with teacher characteristics		x	x	x	x	x	x
with CLASS dimensions		x	x	x			
with APEEC					x	x	x

Estimations from Female only

Table A11. Effect Size at the Kindergarten Entry: Female only

2017 Fall	M1	M2	M3	M4	M5	M6
Receptive Vocabulary	0.09	0.17	0.10	0.13	0.22*	0.15
Math	-0.00	0.04	0.01	0.01	0.06	0.03
Literacy	0.05	0.11	0.05	0.05	0.10	0.05
Language	0.11	0.14	0.11	0.07	0.08	0.07
DCCS	-0.02	-0.03	-0.03			
PT	-0.08	-0.13	-0.07			
Standard score	x	x	x			
Raw score				x	x	x
no f.e.	x			x		
with school f.e.		x			x	
with county f.e.			x			x

Table A12. Effect Size for Standard Score at the End of Kindergarten: Female only

2018 Spring							
Standard score	**M1**	M2	**M3**	M4	**M5**	M6	**M7**
Receptive Vocabulary	-0.14	-0.12	-0.06	-0.11	-0.12	-0.09	-0.11
Math	0.09	0.09	0.06	0.10	0.10	0.05	0.10
Literacy	0.05	0.04	0.02	0.06	0.05	-0.00	0.07
Language	0.03	0.01	0.00	0.03	0.02	-0.03	0.04
DCCS	0.11	0.06	0.13	0.06	0.05	0.12	0.06
PT	0.00	-0.04	-0.02	-0.02	-0.03	-0.02	0.00
no f.e.	x	x			x		
with school f.e.			x			x	
with county f.e.				x			x
with % inclusion		x	x	x	x	x	x
with teacher characteristics		x	x	x	x	x	x
with CLASS dimensions		x	x	x			
with APEEC					x	x	x

Table A13. Effect Size for Raw Scores at the End of Kindergarten: Female only

2018 Spring							
Raw	**M1**	M2	**M3**	M4	**M5**	M6	**M7**
Receptive Vocabulary	-0.10	-0.11	-0.06	-0.10	-0.12	-0.08	-0.11
Math	0.08	0.08	0.04	0.08	0.08	0.03	0.09
Literacy	0.07	0.05	0.04	0.08	0.06	0.01	0.09
Language	0.08	0.05	0.03	0.07	0.06	0.00	0.08
no f.e.	x	x			x		
with school f.e.			x			x	
with county f.e.				x			x
with % inclusion		x	x	x	x	x	x
with teacher characteristics		x	x	x	x	x	x
with CLASS dimensions		x	x	x			
with APEEC					x	x	x

Table A14. Effect Size for Standard Score Gain during the Kindergarten: Female only

Fall 2017-Spring 2018							
Standard score	**M1**	M2	**M3**	M4	**M5**	M6	**M7**
Receptive Vocabulary	-0.21*	-0.18*	-0.15	-0.17*	-0.19*	-0.16	-0.18*
Math	0.07	0.05	0.01	0.04	0.05	-0.02	0.04
Literacy	-0.04	-0.07	-0.08	-0.05	-0.06	-0.09	-0.04
Language	-0.01	-0.03	-0.03	-0.02	-0.02	-0.06	-0.01
DCCS	0.11	0.05	0.13	0.05	0.05	0.12	0.05
PT	0.04	0.01	0.04	0.02	0.02	0.04	0.04
no f.e.	x	x			x		
with school f.e.			x			x	
with county f.e.				x			x
with % inclusion		x	x	x	x	x	x
with teacher characteristics		x	x	x	x	x	x
with CLASS dimensions		x	x	x			
with APEEC					x	x	x

Table A15. Effect Size for Raw Score Gain during the Kindergarten: Female only

Fall 2017-Spring 2018							
Raw score	**M1**	M2	**M3**	M4	**M5**	M6	**M7**
Receptive Vocabulary	-0.21***	-0.22***	-0.19*	-0.22**	-0.22***	-0.20*	-0.21**
Math	0.05	0.01	-0.02	0.01	0.02	-0.02	0.01
Literacy	-0.03	-0.08	-0.10	-0.06	-0.07	-0.11	-0.04
Language	0.03	-0.01	-0.02	0.01	0.00	-0.04	0.02
no f.e.	x	x			x		
with school f.e.			x			x	
with county f.e.				x			x
with % inclusion		x	x	x	x	x	x
with teacher characteristics		x	x	x	x	x	x
with CLASS dimensions		x	x	x			
with APEEC					x	x	x