NIEER
NATIONAL INSTITUTE FOR EARLY EDUCATION RESEARCH

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

August 2019

Milagros Nores, Ph.D., Kwanghee Jung, Ph.d., Carol Contreras, B.A, Elisa Valle, B.S. 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.

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

Elisa Valle, B.S. is Assistant Project Coordinator 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.

Grateful acknowledgment is made to the West Virginia Department of Education's Office of Early Learning. In addition, the authors would like to thank the schools and school districts of Fayette, Greenbrier, Kanawha, Nicholas, Putnam, Roane and Wood, which opened their doors and classrooms to the research team.

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., Jung, K, Contreras, C., Valle, E. & M. Allenger (2019). Evaluation of West Virginia Universal Pre-K. Third Year Longitudinal Outcomes. New Brunswick, NJ: National Institute for Early Education Research.

Table of Contents

Introduction

The West Virginia (WV) Universal Pre-K system continues to serve children across all 55 counties in the state. It currently enrolls over 15,000 children, which is 67 percent of 4-year-olds in the state, and continues to rank 6th in the nation in access to preschool for 4-year-olds (Friedman-Krauss et al., 2019). In the last decade, West Virginia has continuously increased their quality standards and currently meets all NIEER's minimum quality standards benchmarks and nine of the 10 newly developed and more robust benchmarks (Friedman-Krauss et al., 2019). As state pre-K expansion initiatives serve increased children during the years before kindergarten, it is especially critical to study the quality and effectiveness of such programs. It is also essential that evaluations are embedded in the systems so as to understand how children who attend the program progress through the early elementary years, with particular attention to the quality of the K-12 experience.

A five-year longitudinal study of the WV pre-K program was started in the fall of 2015 with the goal of estimating the effects of the WV pre-K program, understand the extent to which initial benefits result in persistent educational advantages, and assess the quality of the educational experiences of children through their P-3 progression. This report presents results for the third year of the study. It reports the results of a longitudinal cohort of children at the end of their first-grade year. The report examines results across various child developmental domains, and how results varied across children. In particular, the reports attempts to understand to what extent pre-K impacts are sustained over time and for whom.

In the previous kindergarten year (2016-2017) report, we highlighted 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 benefitted most from the program. Girls also showed stronger effects. However, positive impacts diminished by the end of kindergarten year, with non pre-K attending children starting to converge with pre-K attending children in the developmental outcomes measured. This report follows this end of kindergarten findings by assessing results at the end of first grade. The convergence trend is found to persist.

As found in the kindergarten classrooms, observations of first grade classroom quality following the longitudinal sample showed that classroom quality experienced by the longitudinal sample in first grade may be a strong contributor to this convergence. First grade classrooms proved to were observed having minimal quality, much lower than what the children experienced in their preschool and even their K year.

This study has some limitations that should be stated. Children and classrooms in this study initially selected from seven counties in WV which had lower enrollment rates in the program, to allow recruitment of a comparison group. However, this condition also means these counties have been the slower ones in in their pre-K program expansion. Thus, generalizability of the findings to the rest of the state is not guaranteed, if there are important differences between these counties and those with larger enrollment rates.

Study Methods

A five-year longitudinal study of the WV pre-K program was started in the fall of 2015. This report presents the third year of the study in which the longitudinal cohort was assessed in the

spring of their first-grade year. The current report estimates the differences between longitudinal children that attended pre-K with those that did not in various developmental areas. The following research questions were examined.

1. What is the impact of the prekindergarten program on children's language, math, literacy, and executive functions skill measures at the end of their first grade 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 experienced by the longitudinal children from pre-K through to first grade?

Sample

At the inception of this study, in the fall of 2015, 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 following 2016-2017 school year, pre-K group of children from the initial sample was followed into kindergarten and another group of children who did not attend WV pre-K was randomly selected from their same classrooms and schools. That year the study sample consisted of 605 kindergarten children who attended the WV pre-K program the previous year[1] and another group of 366 kindergarten children who had not attended the WV pre-K. This report follows these two groups of children into first grade.

The first grade sample consists of 518 first grade children who attended WV pre-K program and another group of 309 first grade children who did not attend WV pre-K. Table 1 reports demographics for the sample of 827 children in the study. This implied an attrition rate of 15%. The sample is predominantly White (92%) and low income (62.2%), with a balanced gender composition (48.4% female). This sample does not differ significantly with the original sample. That is, children assessed in first grade were comparable to the average children in these districts in terms of gender and race. Control group of children were slightly less likely to be low income. We assessed between 1 and 10 children per first grade classroom, following the longitudinal children as the moved across the P-3 system.

In addition, we conducted classroom observations in 142 first grade classrooms. APEEC and CLASS data were collected simultaneously in one visit. In pre-K, ECERS-3 data was collected in 125 classrooms and CLASS data was collected in 120 classrooms in two separate visits.[2] Classroom observation for the preschool sample was reported in a separate report.

[1] Additional pre-K attenders were assessed in this round, who were originally identified as non-attenders, but then track in the West Virginia Education Identification System (WVEIS) as pre-K attenders.
[2] Differences in the final sample were due to interruptions in West Virginia's educational system in the spring of 2018 that created many difficulties observing all classrooms as planned.

Table 1. Child demographics for sample, N=827

Child Characteristics	Total sample N=827		Comparison K sample N=309		Treatment K sample N=518		WV school average for these districts*
	N	%	N	%	N	%	
Gender							
Male	427	51.6%	147	47.6%	280	54.1%	51.4%
Female	400	48.4%	162	52.4%	238	45.9%	49.1%
Low Income							
Low Income	513	62.2%	163	52.9%	350	67.7%	68.1%
Other	312	37.8%	145	47.1%	167	32.3%	31.9%
Race/Ethnicity							
White	760	92.0%	287	92.9%	473	91.3%	93.1%
Black	28	3.4%	7	2.3%	21	4.1%	3.5%
Other	39	4.6%	15	5.1%	24	4.6%	3.3%

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

Measures

Measures on Children

This evaluation assessed children's 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). Moreover, the research team also assessed executive functioning (EF) using two measures: the Dimensional Change Card Sort Game (DCCS) and the Peg Tapping task (PT). We follow with a description of these measures.

The *Peabody Picture Vocabulary Test—Fourth Edition (PPVT-IV;* Dunn & Dunn, 2007) is a test of receptive English vocabulary. The PPVT is predictive of general cognitive abilities and is a direct measure of vocabulary size. 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 test is normed at 100 with a standard deviation of 15.

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. 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; Barnett, et al., 2018). This test is also normed at 100 with a standard deviation of 15.

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. The test assesses attention-shifting and short-term memory. 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. The task requires 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). 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

The experiences of the children in the longitudinal cohort has been assessed in the elementary years using two observational measures:

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 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.

Since the quality of the experiences of the longitudinal cohort has been consistently assessed utilizing the CLASS from preschool through first grade, we are able to track how this quality has changed over time.

Procedures

The classrooms sample was assessed in a collaboration with the WVDE, Marshall's University. NIEER and county coordinators. Training and reliability for both observation measures was delivered by certified NIEER trainers, and reliability was completed with the observers prior to the start of data collection. CLASS observers were trained by a CLASS certified trainer that met the Teachstone[3] reliability requirements for observer certification. All observers were then required to fulfill the Teachstone reliability process. APEEC observers were required to meet

[3] 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/

reliability percentages of 85% with NIEER trainers. Data collectors were shadowed in the field to reduce scoring drift.

Observations were collected between January and May 2018. Marshall University observers called schools in advance to schedule appointments for observations, and teacher names were disclosed at that time. All observation score sheets were cleaned, entered and analyzed at NIEER.

Results

Below we address the research questions on the associations between having attended the pre-K program and children's kindergarten entry, end of kindergarten, and end of first grade developmental outcomes. Main results are presented below 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, at the end of kindergarten, and at the end of first grade year?

We start by reporting the average results for the spring of 2017 for the whole sample, and then separated by treatment and comparison groups. This is the spring of the longitudinal children's first grade year. For the purpose of comparison, descriptive results from kindergarten entry and at the end of kindergarten year are also reported.

We follow this with estimations using multi-level regression analysis on the association between children having attended pre-K (or not) and children's language, literacy, math, and executive function skills. From the WVEIS (WV education information system) we are able capture children's race and/or ethnicity (White, African American or Other), their gender and low-income status. These are included as controls in the estimations (coded as dichotomous variables with values 0 or 1).

Table 2 reports summary statistics on children's outcomes for the overall sample, as well as for the treatment and comparison groups separately. Raw and standard scores are reported for the PPVT and WJ-III subtests. These are standardized with a mean of 100 and a standard deviation of 15. This means that children scoring above the mean of 100 are showing outcomes above average for their age. Similarly, children scoring below the mean of 100 are scoring below the norm for their age.

Fall kindergarten entry scores, spring end of kindergarten scores, and spring end of first grade scores are also reported. At kindergarten entry and end of kindergarten, children in treatment group showed higher standard scores in all of the measures except end of K PPVT scores. This trend is consistent at the end of first grade, by which children in treatment groups showed higher standard scores in in all of the measures except PPVT standard scores.

Table 2. Average child scores across the different measures for the total sample, treatment and comparison group, Kindergarten N=971, First Grade N = 827

	Total sample		Comparison		Treatment	
	Mean	SD	Mean	SD	Mean	SD
PPVT Raw Fall K Score	99.79	19.70	99.81	20.29	99.78	19.35
PPVT Raw Spring K Score	110.20	18.32	111.84	18.66	109.20	18.06
PPVT Raw Spring 1st Score	124.49	17.56	125.60	17.58	123.82	17.53
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
PPVT SS Spring 1st Score	105.18	12.92	105.67	13.34	104.89	12.67
WJ-AP Raw Fall K Score	16.80	4.19	16.95	4.31	16.72	4.12
WJ-AP Raw Spring K Score	20.62	3.67	20.64	3.69	20.62	3.67
WJ-AP Raw Spring 1st Score	24.64	4.31	24.78	4.02	24.55	4.48
WJ-AP SS Fall K Score	102.41	12.79	102.00	13.44	102.66	12.38
WJ-AP SS Spring K Score	105.40	12.33	104.55	12.98	105.91	11.90
WJ-AP SS Spring 1st Score	103.04	15.07	102.87	14.62	103.14	15.35
WJ-LW Raw Fall K Score	12.96	5.44	13.09	5.86	12.88	5.17
WJ-LW Raw Spring K Score	23.56	6.89	23.68	6.87	23.48	6.90
WJ-LW Raw Spring 1st Score	34.11	9.29	34.24	9.38	34.03	9.24
WJ-LW SS Fall K Score	97.06	12.29	96.25	13.61	97.55	11.41
WJ-LW SS Spring K Score	107.98	13.12	107.23	14.31	108.44	12.34
WJ-LW SS Spring 1st Score	105.90	14.54	105.41	15.71	106.20	13.79
WJ-PC Raw Fall K Score	5.39	2.14	5.48	2.16	5.33	2.12
WJ-PC Raw Spring K Score	9.30	4.16	9.49	4.02	9.18	4.25
WJ-PC Raw Spring 1st Score	17.04	5.94	17.17	5.98	16.97	5.93
WJ-PC SS Fall K Score	95.45	10.05	94.66	10.83	95.93	9.53
WJ-PC SS Spring K Score	98.92	13.60	98.37	14.54	99.25	13.01
WJ-PC SS Spring 1st Score	96.02	14.40	95.59	15.63	96.28	13.62
DCCS New Fall K Score	16.90	4.22	16.75	4.43	16.99	4.09
DCCS New Spring K Score	18.09	3.70	18.10	3.79	18.09	3.65
DCCS New Spring 1st Score	19.57	3.99	19.55	4.03	19.58	3.96
DCCS Fall K Score	1.97	0.51	1.96	0.56	1.97	0.48
DCCS Spring K Score	2.14	0.54	2.15	0.54	2.14	0.54
DCCS Spring 1st Score	2.40	0.62	2.39	0.64	2.40	0.61
Peg Tapping Fall K Score	13.24	3.75	13.48	3.77	13.09	3.74
Peg Tapping Spring K 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.

Multivariate analyses examine the association between children's learning and program features while simultaneously controlling for children's characteristics. Estimation models include information on the age of children, gender, race and ethnicity, school days by date of assessment, low income, and IEP status. Teacher characteristics include teacher educational 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 facilitate interpretation, the current gap at kindergarten entry, between the lowest income quintile and the highest income quintile is about one standard deviation nationally.

Multivariate estimations are presented longitudinally, first for kindergarten entry, then for the spring of K and finally for the spring of 1st grade.

Kindergarten Entry

Table 3 presents the estimates of the associations between pre-K program participation and child learning and development, along with associations for child characteristics. Children who participated in WV pre-K program evidence higher literacy and language development at kindergarten entry. This results were reported in Nores, et al. (2019).

Low income children evidence statistically significant lower scores across all 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 slightly higher pre-K benefits on executive functions as measured by Peg Tapping. Statistically significant effects are bolded. Full estimations and sensitivity analyses are reported Nores, et al. (2019).

Table 3. 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.

End of Kindergarten

Table 4 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 significant by the end of kindergarten. In this estimation classroom features and quality are included to accounts for the impact of children's kindergarten experiences on their development level. This results were reported in Nores, et al. (2019). 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 associations with child outcomes. Statistically significant effects are bolded. Full estimations and sensitivity analyses are shown reported in Nores, et al. (2019).

Table 4. 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)**
N	806	806	805	806	805	803

* p<0.05; ** p<0.01; *** p<0.001. Note: Other controls are schools and indicators for missing income as well as variables on teacher characteristics and class features. 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. Standardized scores are used for PPVT, and WJ or WM. Errors are clustered by classroom.

End of 1st grade

Table 5 presents the estimates of the associations between the pre-K program, 1st grade classroom features and child characteristics with children's development at the end of the 1st grade year. At the end of first grade, the impact of pre-K participation is no longer evident in the reported measures. Classroom quality and classroom features are included as children's 1st grade experiences. Teacher education, certification, and experiences are also included in this estimation.

Emotional Support as measured by CLASS was positively associated with receptive vocabulary. However, higher CLASS Instructional Support scores are related to lower receptive vocabulary scores. Having said this, CLASS scores are quite low and evidenced minimal variation.

Consistent with kindergarten estimations, low income status and IEP status are negatively associated with most of child outcomes. Gender and race/ethnicity do not show associations with child outcomes except positive relationship for females in language. Retained children perform on average lower. Teacher's master's degree was associated with higher literacy, language, and math scores. Surprisingly, children with teachers who have six to ten years of experiences showed lower literacy, language, math scores than did children with teachers who have less than five years of experience. Statistically significant effects are bolded. Full estimations and sensitivity analyses are shown in appendix.

Table 5. Multivariate analyses of children's 2018 spring standard score in relation to child and classroom characteristics and CLASS domains

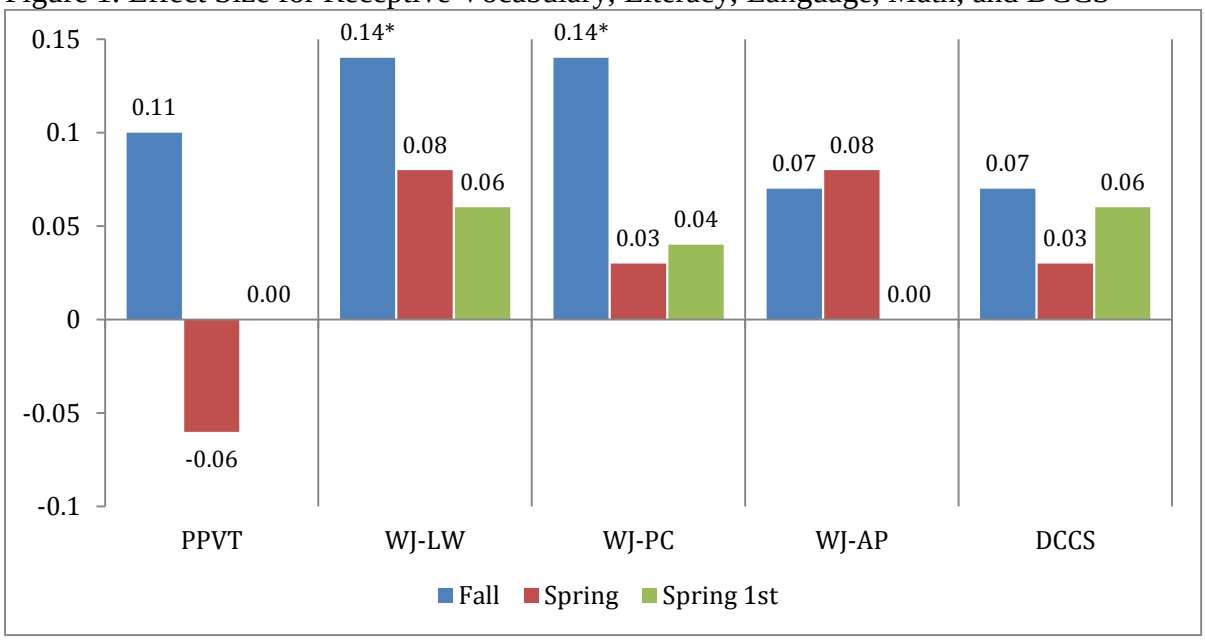
	Rec. Vocabulary (PPVT/TVIP)	Literacy (WJ/WM-LW)	Language (WJ/WM-PC)	Math (WJ/WM-AP)	DCCS
Treatment	-0.006	0.056	0.039	-0.007	0.055
	(0.08)	(0.07)	(0.06)	(0.07)	(0.07)
Retained	-0.650***	-0.887***	-0.806***	-0.758***	-0.451*
	(0.16)	(0.15)	(0.15)	(0.14)	(0.20)
Age	-0.024**	-0.081***	-0.077***	-0.078***	0.002
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female	0.043	0.099	0.151**	-0.069	0.099
	(0.06)	(0.05)	(0.06)	(0.07)	(0.07)
White	0.174	0.082	-0.031	0.233	0.214
	(0.12)	(0.13)	(0.12)	(0.15)	(0.15)
Low Income	-0.419***	-0.396***	-0.433***	-0.418***	-0.284***
	(0.08)	(0.07)	(0.06)	(0.07)	(0.08)
IEP	-0.460***	-0.462***	-0.420***	-0.569***	-0.215
	(0.11)	(0.10)	(0.09)	(0.10)	(0.12)
Teacher: Master's	0.073	0.219**	0.196*	0.172*	-0.017
	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)
Experience: 6-10	-0.106	-0.235*	-0.259*	-0.191*	-0.135
	(0.10)	(0.10)	(0.10)	(0.09)	(0.11)
Experience: 10more	-0.064	-0.031	0.003	0.109	-0.019
	(0.11)	(0.10)	(0.10)	(0.10)	(0.10)
Certification	0.013	0.052	0.078	0.125	-0.027
	(0.13)	(0.12)	(0.12)	(0.13)	(0.13)
Class Size	-0.062	0.026	-0.018	-0.033	0.026
	(0.14)	(0.14)	(0.13)	(0.14)	(0.16)
Inclusion high	0.072	0.001	0.073	-0.048	0.048
	(0.08)	(0.09)	(0.09)	(0.07)	(0.09)
CLASS ES	0.297**	0.215	0.149	0.031	-0.055
	(0.11)	(0.13)	(0.13)	(0.11)	(0.14)
CLASS CO	-0.043	-0.015	0.076	0.112	0.153
	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)
CLASS IS	-0.274*	-0.164	-0.184	-0.037	-0.216
	(0.13)	(0.14)	(0.13)	(0.14)	(0.19)
N	827	827	827	827	827

* p<0.05; ** p<0.01; *** p<0.001. Note: Other control included is school. 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. 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. The figure reports the effects at kindergarten entry, at the end of kindergarten and at the end of 1st grade scores. Beneficial effects of language and literacy at the beginning of the kindergarten diminished

through kindergarten and first grade year. Positive impacts of pre-K on math were consistent during the kindergarten year but the impacts were fade-out at the end of first grade. Positive impacts on executive functioning skills decreases during the kindergarten year, but increases again during first grade year, minimally.

Figure 1. Effect Size for Receptive Vocabulary, Literacy, Language, Math, and DCCS

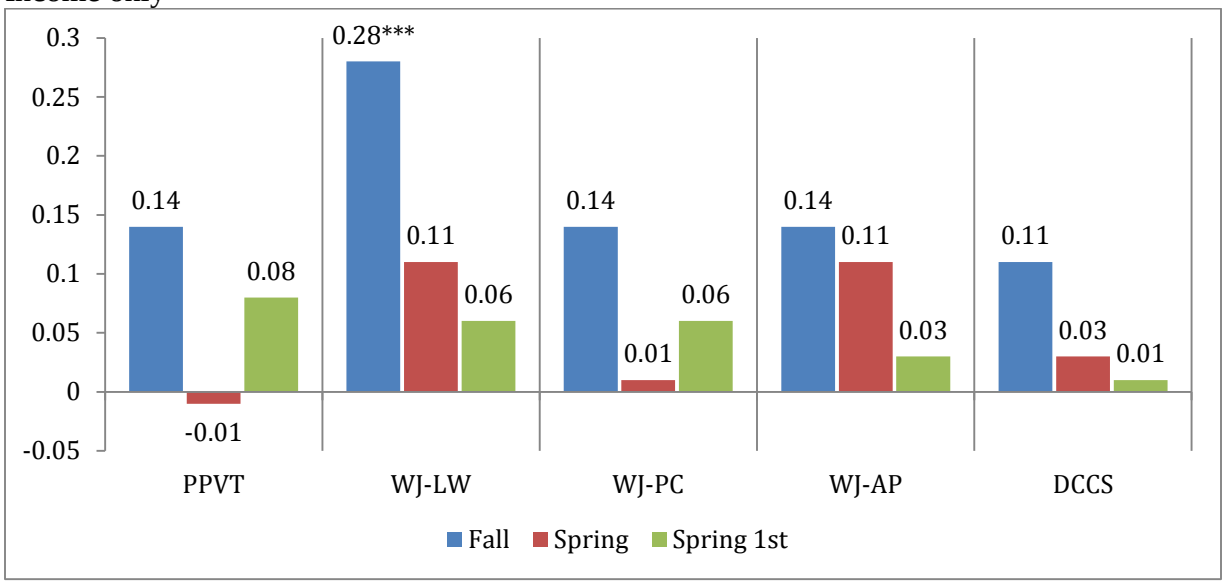


2. Are there child subgroups (as defined by low income or child gender) benefit more from the prekindergarten program than others?

We further analyzed the association between having attended pre-K and children's development levels only for low income children and females. Figures 2 depicts the results from low income children at kindergarten entry, at the spring of kindergarten and at the spring of first grade. Effect sizes from female estimations are presented in Figure 3 below.

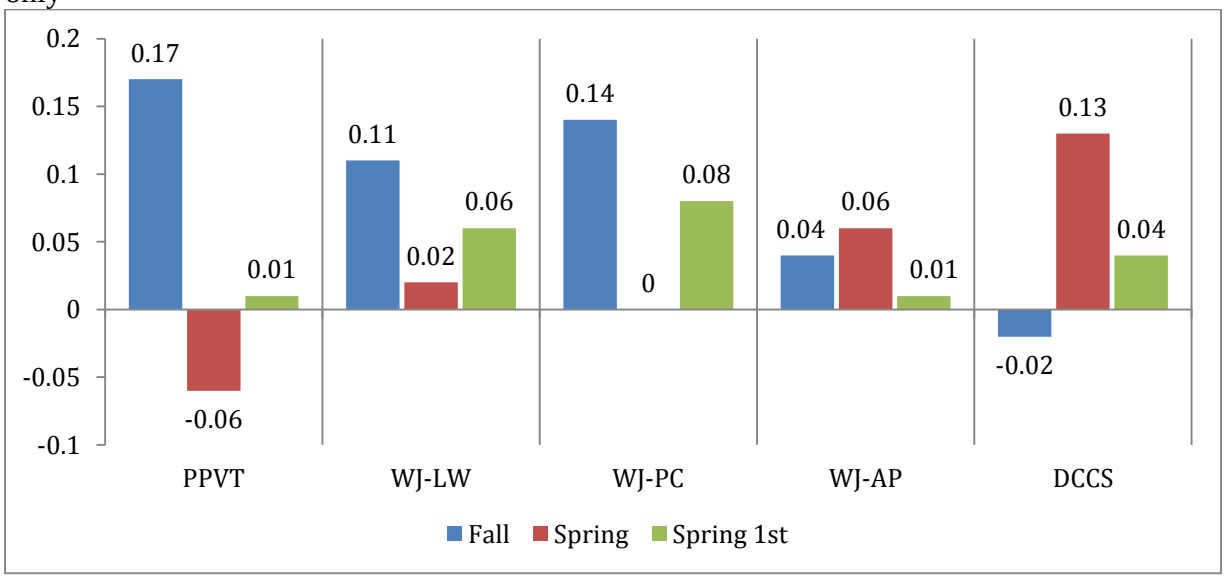
At kindergarten entry, children identified as low-income families show significant benefits in literacy from having attended pre-K. These effects are not observed for vocabulary or math. At the spring of kindergarten, low income children show small effects from pre-K participation sustained through the kindergarten year in literacy and math (see figure 2).

Figure 2. Effect Size for Receptive Vocabulary, Literacy, Language, Math, and DCCS: Low Income only



At kindergarten entry, pre-K program effects on receptive vocabulary is highest for girls, though not significant (See figure 3). At the spring of kindergarten, females showed small positive program effects in executive functioning measures. However, the gains were not significant. At the spring of first grade, females showed small positive program effects in most of the outcomes including literacy, language, and executive functions. None of the gains in first grade outcomes are significant. Overall convergence across the pre-K attending and non-attending group seems to be present across all measures assessed.

Figure 3. Effect Size for Receptive Vocabulary, Literacy, Language, Math, and DCCS: Females only



3. What is the overall observed quality experienced by the longitudinal children from pre-K through to first grade?

Table 6 and Figures 4-6 report the quality of the experiences over the years for the longitudinal cohort that attended pre-K, using the CLASS. CLASS scores from pre-K classrooms in the spring of 2016, kindergarten classrooms in the spring of 2017, and first grade classrooms in the spring of 2018 are reported and illustrated below. Children in that attended pre-K experienced lower classroom quality in kindergarten across all CLASS domains, and even lower quality in 1st grade. Particularly low are the CLASS IS quality in 1st grade. The low CLASS IS could strongly contribute to the converging trends in K and 1st grade for the pre-K group. This is explained further in the discussion section below. Statistically significant differences in scores across the years are marked with an asterisk.

Table 6. CLASS Domains across the years for pre-K attenders in the Longitudinal Cohort

	Pre-K 2016 (N=105)				K 2017 (N=140)				1st grade (N=142)			
	Mean	(SD)	Min	Max	Mean	(SD)	Min	Max	Mean	(SD)	Min	Max
Emotional Support	5.66***a	(0.90)	2.35	6.95	5.05	(0.66)	2.75	6.25	5.21*b	(0.64)	3.60	6.70
Classroom Organization	5.09*a	(1.16)	1.33	6.87	4.81	(0.81)	2.27	6.40	5.20***b	(0.66)	2.93	6.67
Instructional Support	2.65***a	(0.83)	1.13	5.33	2.06	(0.72)	1.00	4.93	1.66***b	(0.31)	1.07	2.53

* p<0.05; ** p<0.01; *** p<0.001. Note. [a]Y1 and Y2 means are significantly different. [b]Y2 and Y3 means are significantly different.

Figure 4. CLASS ES Domain across the years for pre-K attenders in the Longitudinal Cohort

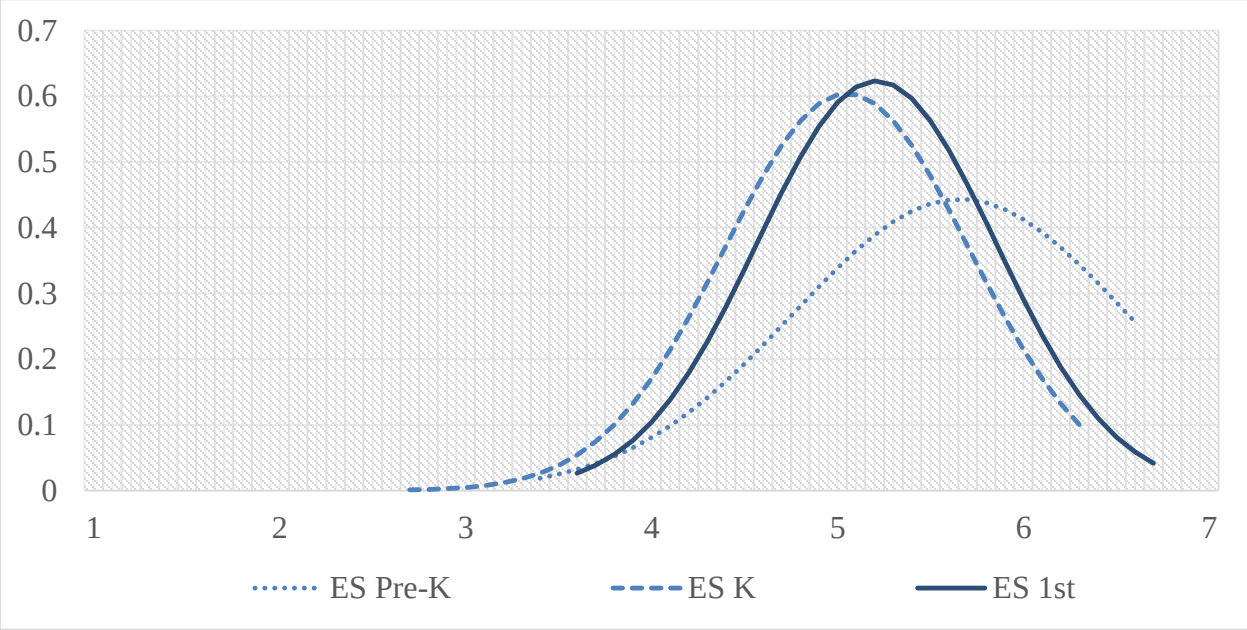


Figure 5. CLASS CO Domain across the years for pre-K attenders in the Longitudinal Cohort

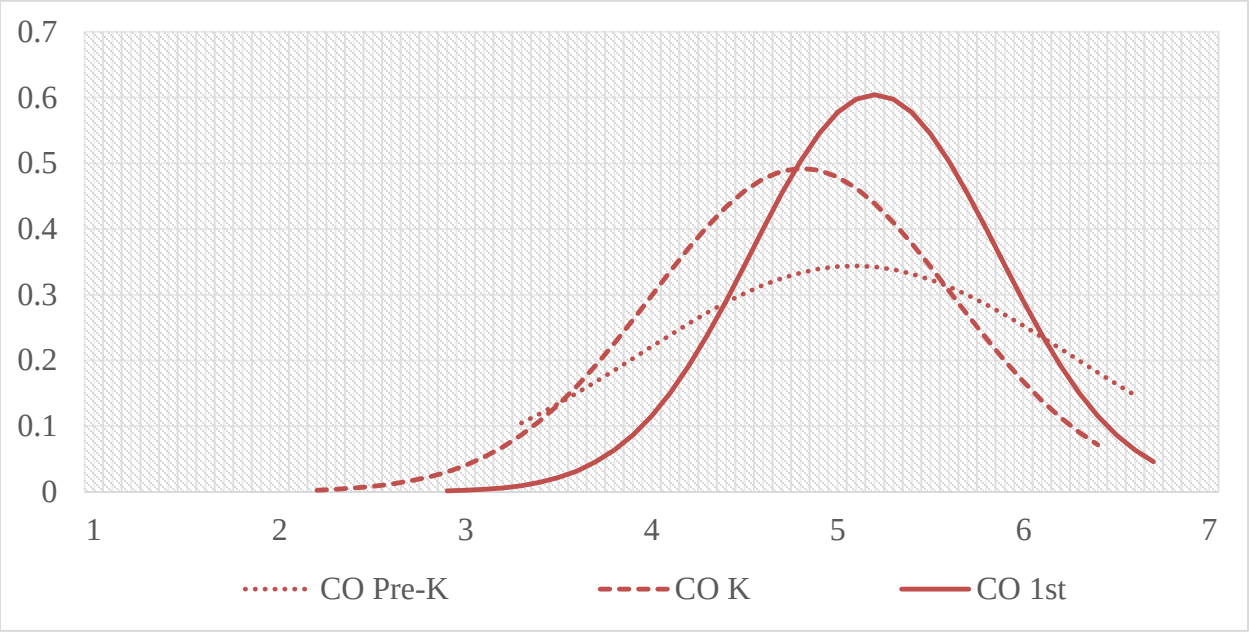
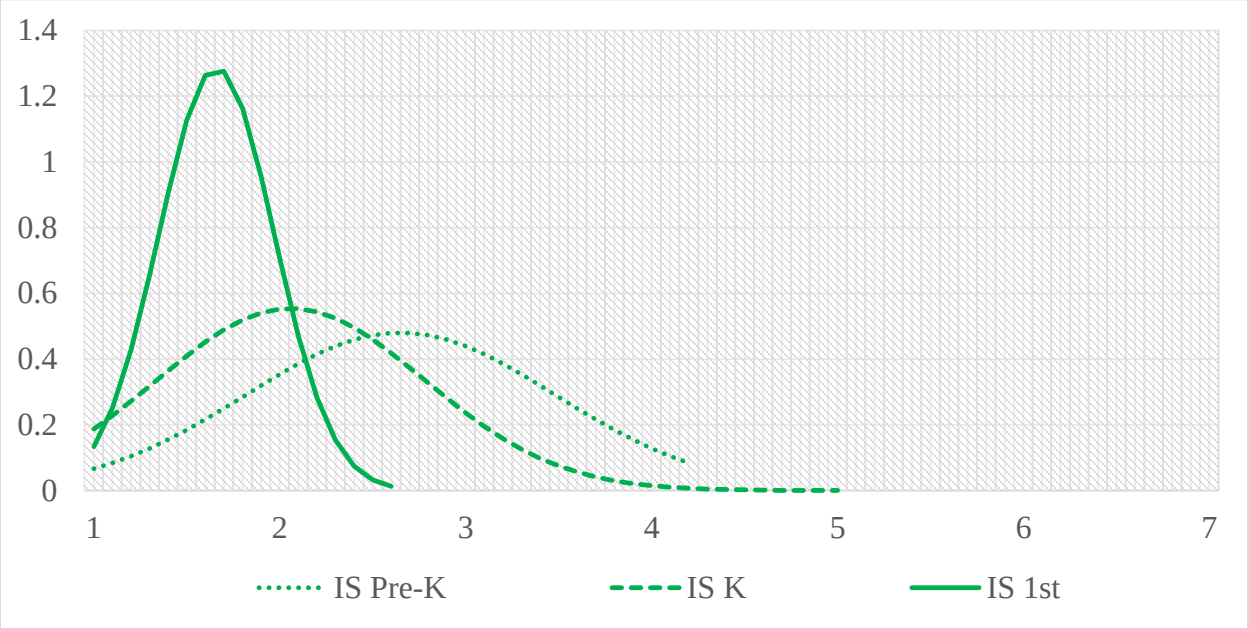


Figure 6. CLASS IS Domain across the years for pre-K attenders in the Longitudinal Cohort



Summary

In this report we present findings from the third year (2017-18) of the WV Universal Pre-K Program evaluation. We investigated the impact of West Virginia's universal preschool program on children's language, math, literacy, and executive functioning skills at first grade and compare this to the trends in kindergarten. We also examined whether results varied for low income children and females. The classroom quality children experienced in pre-K, kindergarten and 1st grade classrooms were also reported.

We report positive impacts of pre-k program on children's learning and development at kindergarten entry, in particular, in literacy and language at kindergarten entry. The positive association between pre-K participation and literacy was greater for low income children and for girls on receptive vocabulary. Estimates for the spring of kindergarten and the spring of first grade show that some positive associations remain but are not significant anymore.

These results are consistent with most evaluations of preschool programs, where rigorous evidence has shown positive short-terms impacts of such programs on children's development (Yoshikawa, et. al, 2013). The results also suggest convergence between children that attended the preschool experience and children that did not. This is similar to Yoshikawa, et. al (2013) summary stating that convergence appears to exist over the elementary grades (p.9). The authors state that in many cases, despite this, there is evidence of effects on outcomes in adulthood. In West Virginia, the reasons for convergence of test scores may be due to the low quality of kindergarten and first grade experienced by the longitudinal cohort. These low quality experiences fail to strengthen the gains generated in children by their early childhood education experience. Another explanation for convergence may derive from having children ready and with positive experiences entering K, in parallel to less prepared children. These scenarios may promote that elementary teachers focus on those behind, instead of creating strong classrooms that differentiate and more successfully support growth across of children. Since these counties in these studies did have lower pre-K enrollment rates than other in the state, it may be the case that a larger fraction of children in every classrooms needs strong support due to their lack of previous education experience. Results may differ once a critical mass of students experience the preschool program as these may result in lesser differences between children in their early elementary experiences.

References

Barnett, S., Jung, K., Friedman-Krauss, A., Frede, E., Nores, M, Hustedt, J., Howes, C., & Daniel-Echols, M. (2018). State Prekindergarten Effects on Early Learning at Kindergarten Entry: An Analysis of Eight State Programs. *AERA Open 4*, 1-16.

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.

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.

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.

Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G., & DiCrecchio, N. (2019). *The state of preschool 2018: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.

Maxwell, K. L., McWilliam, R. A., Hemmeter, M. L., Ault, M. J., & Schuster, J. W. (2001). 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

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.

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.

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.

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 standard score at the end of first grade, raw score at the end of first grade, and standard score gain and raw score gain between spring kindergarten and spring first grade will be presented in order. Preferred models are those summarized in the main document. These tables provide sensitivity analyses.

Appendix

Table A1. Effect Size for Standard Score at the End of 1st Grade

Spring 2018									
Standard score	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
Receptive Vocabulary	0.01	0.01	0.00	0.00	-0.01	-0.01	-0.00	0.01	-0.01
Math	0.03	0.00	0.02	0.03	-0.01	0.02	0.03	0.00	0.02
Literacy	0.05	0.07	0.05	0.05	0.06	0.04	0.05	0.07	0.05
Language	0.06	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.05
DCCS	0.05	0.06	0.04	0.05	0.06	0.04	0.05	0.06	0.04
no f.e.	x			x			x		
with school f.e.		x			x			x	
with county f.e.			x			x			x
with % inclusion	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 A2. Effect Size for Raw Score at the End of 1st Grade

Spring 2018									
Raw score	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
Receptive Vocabulary	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.02	0.00
Math	0.05	0.02	0.04	0.04	0.01	0.03	0.04	0.02	0.04
Literacy	0.08	0.10	0.07	0.07	0.08	0.07	0.07	0.10	0.07
Language	0.08	0.08	0.08	0.08	0.06	0.07	0.08	0.07	0.08
no f.e.	x			x			x		
with school f.e.		x			x			x	
with county f.e.			x			x			x
with % inclusion	x	x	x	x	x	x	x	x	x
with teacher	x	x	x	x	x	x	x	x	x

characteristics									
with CLASS dimensions				x	x	x			
with APEEC							x	x	x

Table A3. Effect Size for Standard Score Gain during Kindergarten and 1st Grade Spring 2017-Spring 2018

Standard score	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
Receptive Vocabulary	0.00	0.06	0.02	0.00	0.05	0.01	0.00	0.06	0.01
Math	-0.03	-0.05	-0.03	-0.04	-0.06	-0.04	-0.03	-0.05	-0.03
Literacy	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04
Language	0.01	-0.01	0.01	0.00	-0.02	0.00	0.01	-0.01	0.01
DCCS	0.06	0.05	0.05	0.05	0.04	0.04	0.06	0.05	0.05
no f.e.	x			x			x		
with school f.e.		x			x			x	
with county f.e.			x			x			x
with % inclusion	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 A4. Effect Size for Raw Score Gain during Kindergarten and 1st Grade Spring 2017-Spring 2018

Raw score	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
Receptive Vocabulary	-0.00	0.05	0.01	-0.00	0.04	0.01	-0.00	0.05	0.01
Math	-0.02	-0.04	-0.02	-0.03	-0.05	-0.03	-0.02	-0.04	-0.02
Literacy	-0.03	-0.02	-0.03	-0.03	-0.02	-0.03	-0.03	-0.02	-0.03
Language	0.01	-0.00	0.02	0.01	-0.02	0.01	0.01	-0.00	0.02
no f.e.	x			x			x		
with school f.e.		x			x			x	
with county f.e.			x			x			x
with % inclusion	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

Estimations from Low Income only

Table A5. Effect Size for Standard Score at the End of 1st grade： Low Income Only

Spring 2018									
Standard score	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
Receptive Vocabulary	0.10	0.10	0.11	0.09	0.08	0.10	0.09	0.10	0.10
Math	0.04	0.05	0.06	0.03	0.03	0.05	0.03	0.04	0.05
Literacy	0.07	0.09	0.07	0.06	0.06	0.06	0.06	0.08	0.07
Language	0.07	0.09	0.08	0.05	0.06	0.06	0.06	0.09	0.07
DCCS	0.03	0.02	0.04	0.03	0.01	0.03	0.02	0.02	0.03
no f.e.	x			x			x		
with school f.e.		x			x			x	
with county f.e.			x			x			x
with % inclusion	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 A6. Effect Size for Raw Score at the End of 1st grade: Low Income Only

Spring 2018									
Raw score	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
Receptive Vocabulary	0.11	0.11	0.11	0.10	0.09	0.10	0.10	0.11	0.10
Math	0.06	0.07	0.09	0.05	0.05	0.08	0.05	0.06	0.08
Literacy	0.09	0.11	0.10	0.07	0.09	0.08	0.08	0.11	0.09
Language	0.09	0.10	0.09	0.06	0.07	0.07	0.07	0.10	0.08
no f.e.	x			x			x		
with school f.e.		x			x			x	
with county f.e.			x			x			x
with % inclusion	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 A7. Effect Size for Standard Score Gain during Kindergarten and 1st Grade: Low Income Only

Spring 2017-Spring 2018									
Standard score	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
Receptive Vocabulary	0.12	0.14	0.13*	0.12	0.13	0.13	0.12	0.14	0.12
Math	-0.03	-0.06	-0.01	-0.05	-0.07	-0.02	-0.04	-0.06	-0.01
Literacy	-0.07	-0.06	-0.06	-0.08	-0.07	-0.07	-0.08	-0.06	-0.07
Language	0.03	0.05	0.05	0.02	0.02	0.03	0.02	0.05	0.04
DCCS	0.04	0.03	0.04	0.03	0.01	0.03	0.03	0.03	0.04
no f.e.	x			x			x		
with school f.e.		x			x			x	
with county f.e.			x			x			x
with % inclusion	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 A8. Effect Size for Raw Score Gain During Kindergarten and 1st Grade: Low Income Only

Spring 2017-Spring 2018									
Raw score	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
Receptive Vocabulary	0.13	0.14	0.13*	0.12	0.14	0.13*	0.12	0.14	0.12
Math	-0.02	-0.05	0.00	-0.04	-0.07	-0.01	-0.03	-0.05	-0.00
Literacy	-0.07	-0.05	-0.06	-0.08	-0.07	-0.07	-0.08	-0.05	-0.07
Language	0.04	0.07	0.05	0.02	0.03	0.04	0.03	0.05	0.04
no f.e.	x			x			x		
with school f.e.		x			x			x	
with county f.e.			x			x			x
with % inclusion	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

Estimations from Female only

Table A9. Effect Size for Standard Score at the End of 1st Grade: Female Only

Spring 2018									
Standard score	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
Receptive Vocabulary	-0.02	0.02	-0.02	-0.01	0.01	-0.01	-0.02	0.02	-0.02
Math	0.07	0.01	0.07	0.07	0.01	0.07	0.07	0.01	0.07
Literacy	0.08	0.06	0.08	0.09	0.06	0.08	0.08	0.06	0.07
Language	0.10	0.08	0.11	0.11	0.08	0.11	0.10	0.08	0.10
DCCS	0.03	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.03
no f.e.	x			x			x		
with school f.e.		x			x			x	
with county f.e.			x			x			x
with % inclusion	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 A10. Effect Size for Raw Score at the End of 1st Grade: Female Only

Spring 2018									
Raw score	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
Receptive Vocabulary	-0.01	0.04	-0.01	0.00	0.03	0.00	-0.01	0.04	-0.01
Math	0.08	0.02	0.09	0.09	0.02	0.08	0.09	0.02	0.09
Literacy	0.10	0.07	0.09	0.11	0.07	0.10	0.10	0.08	0.09
Language	0.10	0.07	0.11	0.11	0.07	0.11	0.10	0.07	0.11
no f.e.	x			x			x		
with school f.e.		x			x			x	
with county f.e.			x			x			x
with % inclusion	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 A11. Effect Size for Standard Score Gain during Kindergarten and 1st Grade: Female Only

Spring 2017-Spring 2018									
Standard score	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
Receptive Vocabulary	-0.00	0.05	0.00	0.00	0.04	0.01	-0.00	0.05	0.00
Math	0.02	-0.01	0.01	0.02	-0.01	-0.00	0.02	-0.01	0.01
Literacy	-0.01	-0.02	-0.01	-0.01	-0.02	-0.01	-0.01	-0.02	-0.01
Language	0.04	0.04	0.05	0.05	0.04	0.05	0.04	0.04	0.04
DCCS	-0.02	-0.04	-0.03	-0.01	-0.04	-0.02	-0.02	-0.04	-0.03
no f.e.	x			x			x		
with school f.e.		x			x			x	
with county f.e.			x			x			x
with % inclusion	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 A12. Effect Size for Raw Score Gain during Kindergarten and 1st Grade: Female only

Spring 2017-Spring 2018									
Raw score	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
Receptive Vocabulary	-0.02	0.02	-0.02	-0.02	0.02	-0.01	-0.02	0.03	-0.02
Math	0.04	0.01	0.02	0.04	0.00	0.01	0.04	0.01	0.02
Literacy	-0.01	-0.02	-0.01	-0.01	-0.02	-0.01	-0.01	-0.02	-0.01
Language	0.02	0.03	0.03	0.03	0.03	0.03	0.02	0.03	0.03
no f.e.	x			x			x		
with school f.e.		x			x			x	
with county f.e.			x			x			x
with % inclusion	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