EVALUATION OF WEST VIRGINIA UNIVERSAL PRE-K. FIRST YEAR OUTCOMES

May 2017

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

About the Authors

Milagros Nores, Ph.D. Dr. Nores is a 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.

Shannon Riley-Ayers, Ph.D. Dr. Ayers is Associate Professor at The National Institute for Early Education Research (NIEER) at Rutgers University and conducts research and technical assistant for the Center for Enhancing Early Learning Outcomes (CEELO). Her research spans preschool through grade three with a focus on teaching quality, assessment, and professional learning.

Andrea Kent, M.Ed. Ms. Kent is 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.

Kwanghee Jung, Ph.D. Dr. Jung 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.

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.

Carol Contreras, B.A.. Ms. Contreras is Senior Assistant Project Coordinator at The National Institute for Early Education Research (NIEER) at Rutgers University.

Suggested citation: Nores, M., Riley-Ayers, S., Kent, A., Jung, K, M. Allenger & Contreras, C. (2017). Evaluation of West Virginia Universal Pre-K. First Year Outcomes. New Brunswick, NJ: National Institute for Early Education Research.

Table of Contents

Introduction

WV pre-k program serves 66 percent of four-year-olds in the state and ranks 6th in the nation in access to preschool for four-year-olds (Barnett, et al., 2017). 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 because of WV's new requirement for assistant teachers to have at least a Child Development Associate (CDA) credential. WV meets 9 of the 10 newly developed and more robust benchmarks. The benchmarks include aspects of class size, ratio, qualifications (lead teacher requires a bachelor's degree; assistant teacher requires a CDA), in-service training, screening and referral services, meals, and monitoring. Only five other states meet all 10 minimal standards benchmarks (Alabama, Mississippi, Louisiana's NCESD program, North Carolina, and Rhode Island). The passage of SB 146 (2016) strengthened West Virginia's program by requiring a minimum of 25 hours of weekly instruction in the universal Pre-K and serves as a model for other states.

A report released this spring (Wechsler, et al., 2016) highlighted the quality of West Virginia's Pre-K program stating how the program benefited from several initial design choices. These included a realistic timeline for program expansion and the integration of the program into the K-12 school aid funding formula. WV has focused part of its efforts in a continuous quality improvement cycle that is data driven, with local inputs, in addition to the development of early learning standards, professional development supports, coaching and technical assistance. "Since WV Pre-K's inception, the state has invested significant resources in building the program, gradually achieving universal access and improving quality standards." (Wechsler, et al., 2016, p.3).

In 2015, the National Institute for Early Education Research (NIEER) at the Rutgers University and Marshall University embarked in the first year of a what is to be a five-year evaluation study of West Virginia's Universal Pre-K Program. This report presents the first year (2015-16) findings from the evaluation. Its focus is on the impact on children's learning across various domains, the classroom quality they experienced, and how the impact varied for different children. In addition to describing these findings, the report provides comparable findings from other preschool studies to contextualize the results.

Overall, the report demonstrates positive impacts of pre-k on children's learning and development across all domains. Moreover, it shows evidence that lower income children benefitted more from pre-k as demonstrated in print knowledge and math assessments. Results for boys were higher than for girls across the all domains. Average quality in pre-k was moderate for the ECERS-3 overall, with many classrooms however still scoring as inadequate. Pre-K scores were moderate on Emotional Support and Classroom Organization, however Instructional Support scores were quite low with only about a third of the classrooms having moderates level of quality as measured using the CLASS. We also assessed quality in kindergarten classrooms. Average APEEC scores were low, however the data show a large amount of variation with many classrooms scoring inadequately. CLASS K scores were more similar across classrooms, demonstrating lower average scores than pre-k for Emotional Support and Instructional Support with fewer classrooms reaching higher quality across the board. The report offers considerable depth into classroom quality observed and recommendations for improvement.

As with all evaluations, this study has limitations and it is important to acknowledge these. First, the research methods used in this evaluation are quasi-experimental and not those of a randomized control trial (the gold standard of research). Universal access to pre-K prevents

using randomized control and thus the quasi-experimental research methods utilized here are the next most rigorous. Second, the sample was purposefully selected in 7 counties in WV, and therefore, may not be representative of the full state. These counties were selected because of their lower participation rate in the pre-K program because this made sense for a longitudinal study which is to follow the pre-K cohort included in this report. However, this means that the program was in the process of growth in these countries and partnerships were newer, which may provide some explanation of the range of quality.

Study Methods

The West Virginia Universal Pre-K evaluation study is a multi-year, multi-site study that includes a combination of designs to assess impact of pre-K on children, continued impact over time as children progress through third grade, and quality of the pre-K program and early elementary classrooms. In the first year of the study, the research team collected child and classroom information to answer the following key questions:

1. What is the impact of the prekindergarten program on children's math, language, literacy, and executive functions skill measures at Kindergarten entry?
2. Are there child subgroups (as defined by low income or child gender) that benefit more from the prekindergarten program than others?
3. What is the overall observed quality of pre-K and kindergarten classrooms in 2015-16?

The main purpose of the WV impact evaluation is to estimate the effects of the program on children's learning and development. To measure the impact on children we used on a regression-discontinuity design (RDD) which is a quasi-experimental design that takes advantages of the exogenous determination on who enters the program provided by the birthdate cutoff eligibility. This design offers protection against selection bias beyond simply controlling for family background in a statistical analysis by employing a statistical model that uses stringent age cut-offs to define groups (Lipsey, et al., 2015). Testing groups using the age cut-offs and then statistically adjusting for age variation reduces the likelihood that selection bias has an appreciable impact on study results. RDD designs have been used to estimate preschool program impacts in various programs across the U.S. (Barnett, et al., 2013; Hill, Gormley & Adelstein, 2015; Huang, Invernizzi, & Drake, 2012; Hustedt, et al., 2007; Weiland & Yoshikawa, 2013; Wong, et al., 2008). The cutoff provides two groups that can be compared: those that made the birthdate cutoff and can attend one year, and those that did not make it and need to wait until next year to attend.

Two groups of children were assessed in the fall of 2015. (1) Children entering the pre-k program and (2) children entering kindergarten who have attended the pre-k program the prior year. Measures and procedures are described below. In addition, to assess program quality, classroom observations were conducted February through the end of May. Quality was assessed using observation protocols widely established in the field. Figure 1 (below) reports the data collection timeline for the school year of 2015-16.

Table 1. Data Collection Timeline.

2015	
September	• Training for data collectors • Initial SPP site information gathered
October	• Parent consent form distribution • Fall assessment visit scheduling • Fall child assessment visits begin
November through December	• Fall child assessment visits continue
2016	
January	• Calls to directors to discuss classroom observations (CLASS & ECERS-3)
February through May	• CLASS & ECERS-3 observations in pre-k classrooms • CLASS & APEEC observations in K classrooms

Sample

In fall of 2015, we recruited children entering the WV Pre-K program and children who had attended the pre-K program the previous year in 7 counties in WV. The study focused on 7 counties (Fayette, Greenbrier, Kanawha, Nicholas, Putnam, Roane and Wood) because the RDD is one piece of a larger design that includes a longitudinal study with a comparison group of non-attenders. Therefore, this required working in counties were the WV Pre-K had not achieved high rates of enrollment. Early in the fall, a random sample of schools in the 7 counties were informed about the study and families were invited to participate. In each school, all pre-k and K classrooms were delivered consent forms to enroll children for participation.

We assessed 1,172 children: 599 children starting in pre-k classrooms which are our "comparison" group (because they have not yet received the "treatment" of the pre-k program) and 573 children starting in kindergarten classrooms which are our "treatment" group having attended pre-k the year before. Participating children were comparable to the average children in these district in terms of gender and race, and slightly more likely to be low income. We assessed between 1 and 16 children per pre-k classroom and between 1 and 14 children per K classroom (average of 6.28 for pre-K and 5.25 for K). The analytic sample of 1,172 is predominantly white (92.4%) and low income (72.8%).

Table 2. Child demographics for sample, N=1,172.

Child Characteristics	Total sample N=1,172		Comparison Pre-K sample N=599		Treatment K sample N=573		WV school average for these districts*
	N	%	N	%	N	%	
Gender							
Male	602	51.4%	309	51.6%	293	51.1%	51.4%
Female	570	48.6%	290	48.4%	280	48.9%	49.1%
Low Income							
Low Income	853	72.8%	436	72.8%	417	72.8%	68.1%
Other	319	27.2%	163	27.2%	156	27.2%	31.9%
Race/Ethnicity							

White	1,083	92.4%	540	90.2%	543	94.8%	93.1%
Black	44	3.8%	28	4.7%	16	2.8%	3.5%
Other	45	3.8%	31	5.2%	14	2.4%	3.3%

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

In addition, we conducted classroom observations on the 130 pre-K classrooms and 98 kindergarten classrooms from which we drew children.

Table 3. Classroom sample by grade and instrument used.

	Pre-K		Kindergarten
	Observed APEEC/CLASS	Observed CLASS	Observed APEEC/CLASS
Fayette	23	23	19
Greenbrier	18	18	16
Kanawha	16	16	13
Nicholas	11	11	8
Putnam	21	15	0
Roane	7	7	8
Wood	34	15	34
Overall	130	105	98

Measures

Measures on Children

The *Peabody Picture Vocabulary Test—Fourth Edition (PPVT-IV;* Dunn & Dunn, 2007) is a 228-item test of receptive vocabulary in standard English. The PPVT is predictive of general cognitive abilities and is a direct measure of vocabulary size. The rank order of item difficulties is highly correlated with the frequency with which words are used in spoken and written language. The test is adaptive (to avoid floor and ceiling problems), establishing a floor below which the child is assumed to know all the answers and a ceiling above which the child is assumed to know none of the answers. The test is reliable based on reported split-half reliabilities or test-retest reliabilities. The PPVT has shown concurrent validity (e.g., Qi, Kaiser, Milan, & Hancock, 2006) and the results of these tests are found to be strongly correlated with school success (Blair & Razza, 2007; Early, et al., 2007).

The *Tests of Preschool Early Literacy* (TOPEL; Lonigan, Wagner, Torgesen, 2007) measures abilities related to early literacy. It contains three subtests: print knowledge that measures early knowledge about written language conventions and alphabet knowledge, definitional vocabulary that measures a child's single word oral vocabulary and definitional vocabulary, and phonological awareness which examines elision and blending abilities. The reliability evidence for the subtests and the composite score are good and range from .87-.96 for internal consistency, .81-.91 for test-retest, and .96-.98 for interscorer differences.

The *Woodcock-Johnson Psycho-Educational Battery—Third Edition (WJ-III;* Woodcock, McGrew, Mather, & Schrank, 2001) includes multiple subtests. Only the *Applied Problems* and *Letter-Word Identification* subtests were used in this study. WJ-III was normed on a stratified

random sample of 6,359 English-speaking subjects in the United States. Correlations of the WJ-R with other tests of cognitive ability and achievement are reported to range from 0.60 to 0.70. This measure has been used in numerous large-scale preschool studies (e.g., Early, et al., 2007; Wong, Cook, Barnett, & Jung, 2008).

Dimensional Change Card Sort Task (DCCS; Zelazo, 2006). This task engages reverse categorization where children must sort a set of cards based on different sorting criteria given by the examiner. Generally, the test assesses attention-shifting. Scores on the DCCS reflect a pass/fail system on each of three levels of increasing difficulty. Raw scores range between 0 and 3, where a score of 0 means a child did not pass the first level which includes a color sorting task. At this first level, children are tasked with sorting two objects by color into a corresponding labeled box. A score of 1 means a child passed the color sort but failed the shape sort, which is the subsequent task and asks children to ignore color and instead sort objects by their shape. A score of 2 means a child passed shape sort but failed advance trials. Lastly, a score of 3 means the child passed advance trials, which ask children to ignore color or shape by adding a border to cards to indicate which attribute to sort by. There are no standard score equivalents. However, a study of test-retest reliability, means by age for children 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 with 8 one-tap and 8 two-tap trials in random sequence. The task requires both the ability to hold two things in mind—the rule to tap once when experimenter taps twice and the rule to tap twice when experimenter taps once, and the ability to exercise inhibitory control over one's proponent behavior, the natural tendency to mimic what the experimenter does. Common errors include: (1) complying with only one of the two rules, (2) tapping many times regardless of what the experimenter did, and (3) doing the same thing as the experimenter, rather than the opposite. The final score for Peg Tapping is a sum of all the 16 items that comprise the test. Again, while there are no standard score equivalents, in a study of test-retest reliability, means by age for children 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, 2014). The ECERS-3 is an observation and rating instrument for preschool classrooms serving children aged three to five. The total ECERS-3 score represents an average of the scores on the 35 items under 6 domains. A rating scale between 1 and 7 is used, where a rating of 1 indicates inadequate quality, a rating of 3 indicates minimal quality, a rating of 5 indicates good quality, and a rating of 7 indicates excellent quality. The ECERS-3 was used only in pre-k classrooms. More details on the ECERS-3 can be found in Appendix B.

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. More details on the CLASS can be found in Appendix B.

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. More details on the APEEC can be found in Appendix B.

Procedures

NIEER and Marshall University worked collaboratively to hire and train data collectors on the child standardized assessment and classroom observation measures. For child assessments, data collectors received a two-day training on the measures. Following the two-day training, data collectors were successfully shadowed by expert staff on two iterations of the assessments for reliability. After two iterations of assessments, each of the data collectors achieved 100% reliability.

For classroom observation measures, initial training was provided in administering the observation protocol that includes the ECERS-3 and the CLASS for pre-k classrooms and the APEEC and the CLASS for kindergarten classrooms. Training took place in separate full-day workshops. ECERS-3 observers were trained by an ECERS-3 certified trainer and met the ERSI reliability requirements for observer certification. The trainee must complete three observations with the trainer with 85% or above exact matches or one-away from the true score. All data collectors met the ECERS-3 reliability requirements with agreement percentages ranging between 85-90%. CLASS observers were trained by a CLASS certified trainer and met the Teachstone reliability requirements for observer certification. All data collectors met CLASS reliability requirements with agreement percentages ranging between 84-92% for CLASS PreK and 80-86% for CLASS K. APEEC observers were trained similarly and data collectors met agreement percentages with the trainer ranging between 85-100%. All observation score sheets were cleaned and entered at NIEER by trained staff.

Results

This evaluation addresses the research questions using a combination of methods. The main results are summarized in this report with additional analyses included in the appendices. The RDD estimation method is summarized in Appendix A.

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

We first present descriptive results from the 2015-2016 evaluation showing average scores for the whole sample, and then separated by treatment and comparison groups. We then run RDD estimations to assess the impact of the program in children math, language, literacy and executive function skills. From the WVEIS (WV education information system) we obtained administrative information on children's race/ethnicity (White, African American or Other), gender and low-income status. These are included as controls (dichotomous variables coded 0,1). RDD results are then compared to those reported by various other studies, as well as those for WV in a previous RDD conducted by NIEER in 2004.

Table 4 presents simple summary statistics for the child outcome measures for the overall sample and the treatment and comparison groups. On average, children in the treatment sample scored 24.51 points higher on the PPVT, 20.73 higher on the TOPEL, 5.47 points on the WJ applied problems, 0.32 on the DCCS task, 2.86 on the DCCS if we recode it to count all the points (this provides more variance), which we define as DCCS New and 4.74 on the Peg Tapping task. The RDD estimation will adjust these differences relative to the cut-off date, and age of children, to differentiate average growth related to age from the impact of the program.

Table 4. Average child scores across the different measures for the total sample, treatment and comparison group, N=1,172.

	Total sample		Comparison Pre-K sample		Treatment K sample	
	Mean	SD	Mean	SD	Mean	SD
PPVT	86.54	23.99	74.56	22.88	99.07	17.97
TOPEL	21.09	11.43	13.63	9.78	28.89	6.94
WJ-AP	13.93	5.12	11.25	4.59	16.72	4.04
DCCS	1.75	0.55	1.60	0.57	1.92	0.48
DCCS New*	15.04	5.33	13.64	5.80	16.50	4.33
Peg Tapping	11.19	5.42	8.87	5.80	13.61	3.66

*DCCS New is another way to code DCCS which fully counts for all the points and better captures variance in this measure.

The regression discontinuity design (RDD) assesses the effects of participation in WV Pre-K on children's skills at kindergarten entry using age as a cut off. That is, the kindergarten sample of children who did attend prekindergarten the previous year is compared to 4-year-old children just entering prekindergarten. Graphical displays on the relation between the different outcomes on either side of the age cutoff and superimposed locally weighted nonparametric regression lines on a scatter plot of the raw data are assessed to identify whether the assumption of a discontinuity being present at the age cutoff is supported. This was effectively the case for all the measures of child learning and development assessed.

Main estimations for the RDD model are depicted in this section (with methods and tables with full results included in Appendix A). Children that participated in the WV pre-K program demonstrated statistically significantly higher scores in receptive vocabulary, literacy,

math and executive functions in the preferred estimations.[1] 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 (the children entering preschool). To facilitate interpretation, the current gap at Kindergarten entry, between poor and non-poor children is about one standard deviation nationally. Intent-to-treat ITT estimations are defined as estimates including the full sample. ITT estimates is the average difference between children who are eligible and not eligible, regardless of children end up following their assignment rule or not. Instrumental variable (IV) analyses are defined as those in which we treat students' true assignment into pre-K as an instrument for their actual participation (Hahn et al., 2001; Jacob & Lefgren, 2004; Lee & Lemieux, 2010; van der Klaauw, 2008). This corrects for cases in which there are some violations on the birth date cutoff.

Effect sizes are: 0.31-0.33 for receptive vocabulary (PPVT), 1.07-1.16 for literacy (TOPEL), 0.32-.36 for math (WJ Applied Problems). Executive functions analyses show statistically significant, positive impacts on the modified score of the DCCS and on the Peg Tapping task. Effect sizes are: 0.18-0.20 the original DCCS scoring, 0.26-.29 for the modified DCCS score and 0.23-.25 for the Peg Tapping task. These effects are displayed in Figures 1 and 2 below.

Figure 1. PPVT, TOPEL, and WJ-AP Effect Sizes estimations for RDD.

---
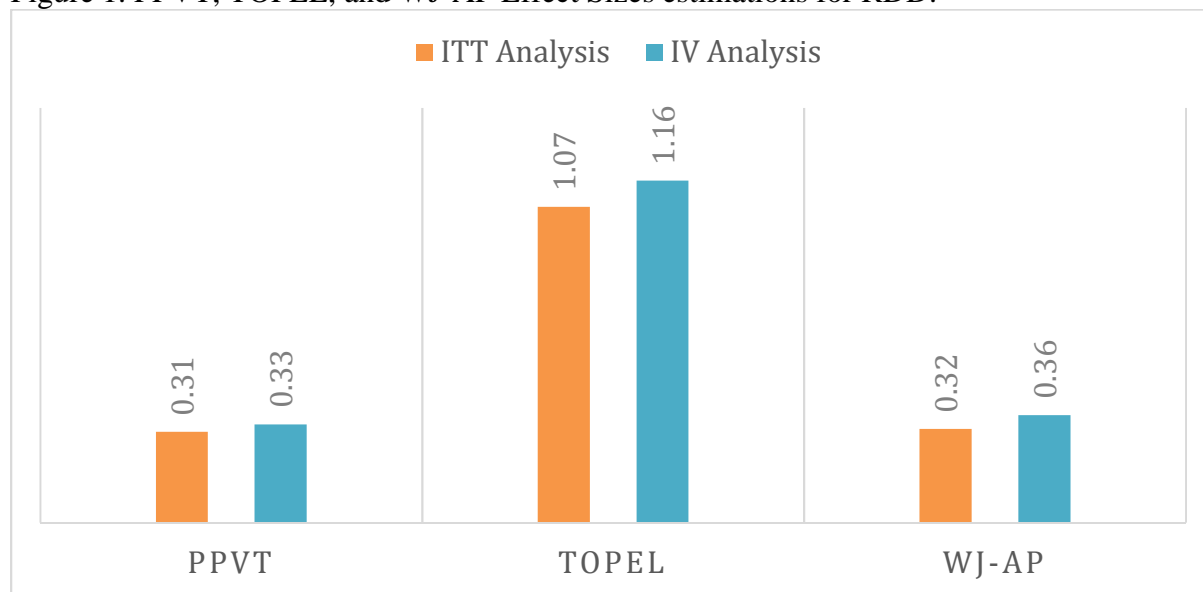
[1] These estimations are preferred for two reasons: between ITT and IV estimations, these are more conservative; and we report 12-month bandwidths although other bandwidths are included in Appendix A. PPVT results showed instability with other bandwidths.

Figure 2. DCCS, DCCS New, and PT Effect Size estimations for RDD.
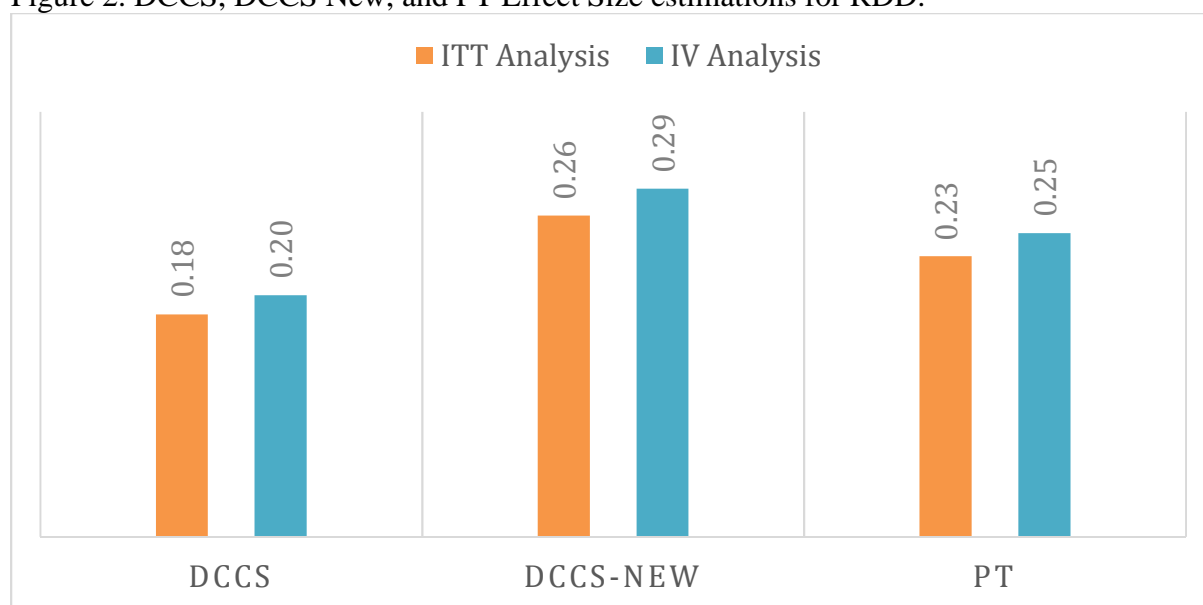
To set these results in context, Table 5 compares these results to those of the study of the Boston preschool program (Weiland & Yoshikawa, 2013), the previous West Virginia RDD as well as the New Jersey RDD (Wong, et al., 2008), and studies of Georgia (Peisner-Feinberg, Schaaf, & LaForett, 2013), Arkansas (Barnett, et al., 2013), North Carolina (Peisner-Feinberg & Schaaf (2011) and Tulsa (Gormley, et al. 2008).

Table 5. Setting results in contexts: RDD results for other preschool programs.

	Language	Literacy	Math	DCCS	PT
WV 2015-16	0.31	1.07	0.32	0.26	0.23
Cities:					
Boston 2008	0.44	0.62[1]	0.59	0.28	0.21
Tulsa 2006	-	0.99	0.36	-	-
States:					
WV 2005	0.13	0.71[2]	0.15	-	-
Arkansas 2006	0.28	1.00[2]	0.27	-	-
Georgia 2012	0.06	0.14[1]	0.18	-	-
New Jersey 2004	0.36	0.50	0.23	-	-
North Carolina 2009	0.27	0.93	0.07	-	-

Notes: [1]Study used the WJ-Letter Word measure. [2]Study used Pre-CTOPP.

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

We next analyzed whether results varied for selected subgroups of the population. We concentrated on low income children, and females as two subgroups of interest and compared the preferred ITT estimations, to ITT estimations for these subgroups only. Figures 3 and 4 depict results by outcome.

While receptive vocabulary results were slightly lower for these two groups, they remained significant. However, low income children showed much higher effects in literacy than the average scores, while females showed slightly lower results. In addition, these same patterns were observed in math, as measured by the WJ-AP.

Figure 3. PPVT, TOPEL, and WJ-AP Effect Sizes estimations for selected subgroups.
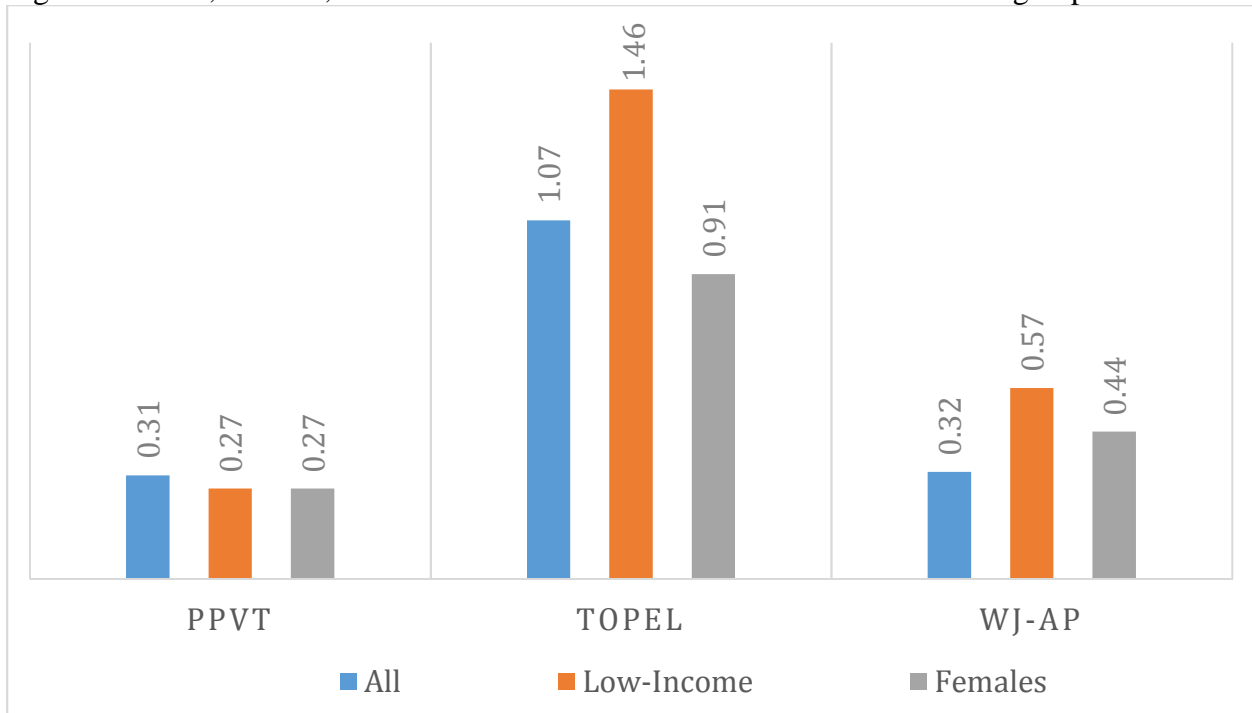
Sample sizes: All=1,174; LI=841, F=571. 12m, ITT, Linear estimations for PPVT and Topel, Quadratic for WJ-Ap.

Figure 4. DCCS, DCCS New, and PT Effect Size estimations for selected subgroups.
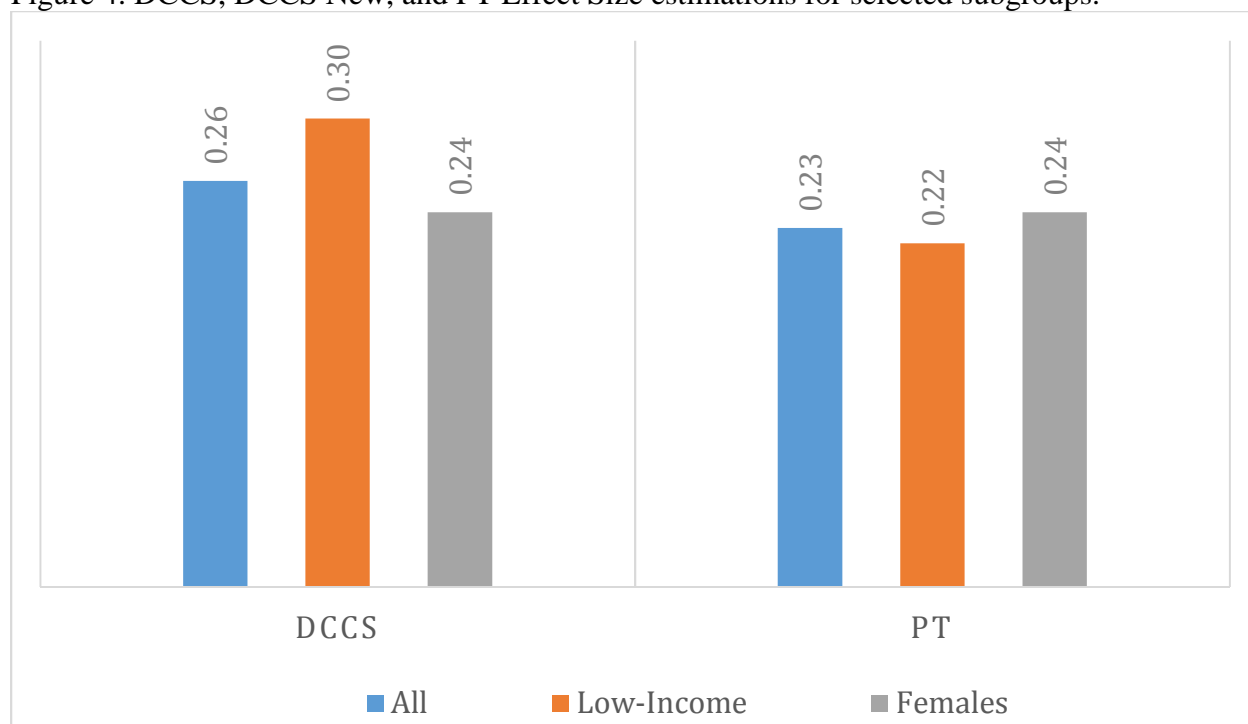
Sample sizes: All=1,174; LI=841, F=565. 12m, ITT, Linear estimations.

3. What is the overall observed quality of preschool and kindergarten classrooms in 2015-16?

In pre-K, ECERS-3 data were collected in 130 classrooms and CLASS data were collected in 105 classrooms in two separate visits. Of the target sample of 132 classrooms, 103 classrooms were observed using both the ECERS-3 and the CLASS, 27 classrooms were observed using ECERS-3 only, and 2 classrooms were observed using CLASS only. In kindergarten, classrooms were observed using both the APEEC and CLASS simultaneously in one visit. Of the target sample of 129 classrooms, data were collected in 98 classrooms. Table 6 represents the full classroom sample of target and observed kindergarten classrooms for both pre-K and kindergarten. (Note: Putnam County declined participation in the kindergarten quality observations for this year.)

Table 6. Classroom Sample by Grade and Instrument Used.

	Pre-K			Kindergarten	
	Target N	ECERS-3 Observed	CLASS Observed	Target N	Observed APEEC/CLASS
Fayette	23	23	23	19	19
Greenbrier	18	18	18	16	16
Kanawha	16	16	16	15	13
Nicholas	12	11	11	8	8
Putnam	22	21	15	27	0
Roane	7	7	7	8	8
Wood	34	34	15	36	34
Overall	132	130	105	129	98

Pre-K Classrooms

Pre-K ECERS-3 Results Spring 2016

The distribution of overall ECERS-3 scores for the 130 classrooms that were observed using the ECERS-3 are presented in Figure 5, together with a superimposed normal distribution. There is a wide dispersion of classrooms across all score levels, with 16% classrooms scoring in the inadequate level, most classrooms scoring in the minimum to good level (60%), and only 23% of classrooms scoring above good. No classrooms scored at the excellent level.

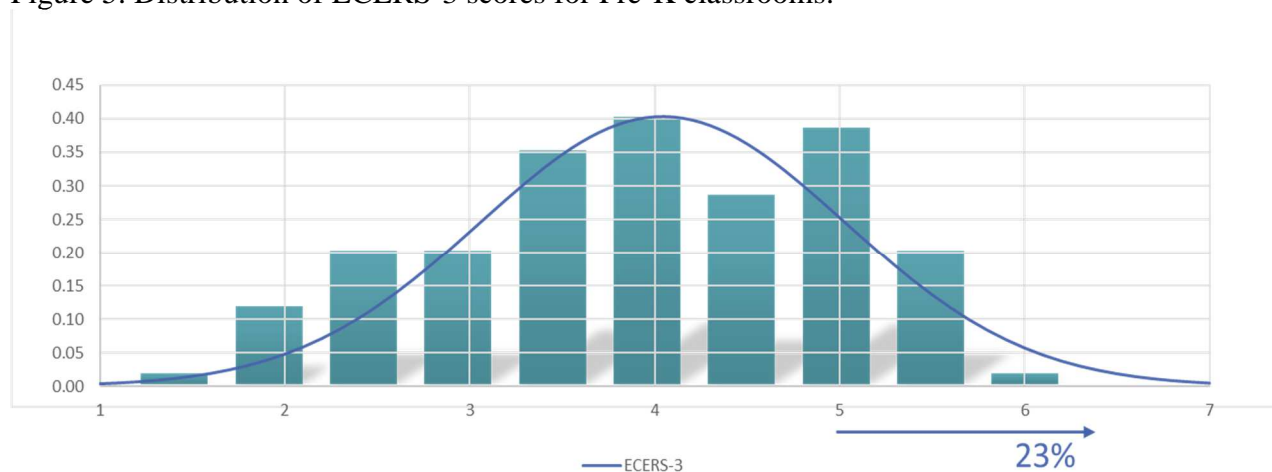Figure 5. Distribution of ECERS-3 scores for Pre-K classrooms.

Figure 6 provides for context average ECERS-3 scores for 2 studies: in GA and NJ Abbott districts (Spring 2016). Since the ECERS-3 is a recent version of the ERS (early rating scales), and widespread use is still not the case, there are not many programs that can be included for comparison purposes.

Figure 6. ECERS-3 scores and subscale scores for WV Pre-K and two other programs.
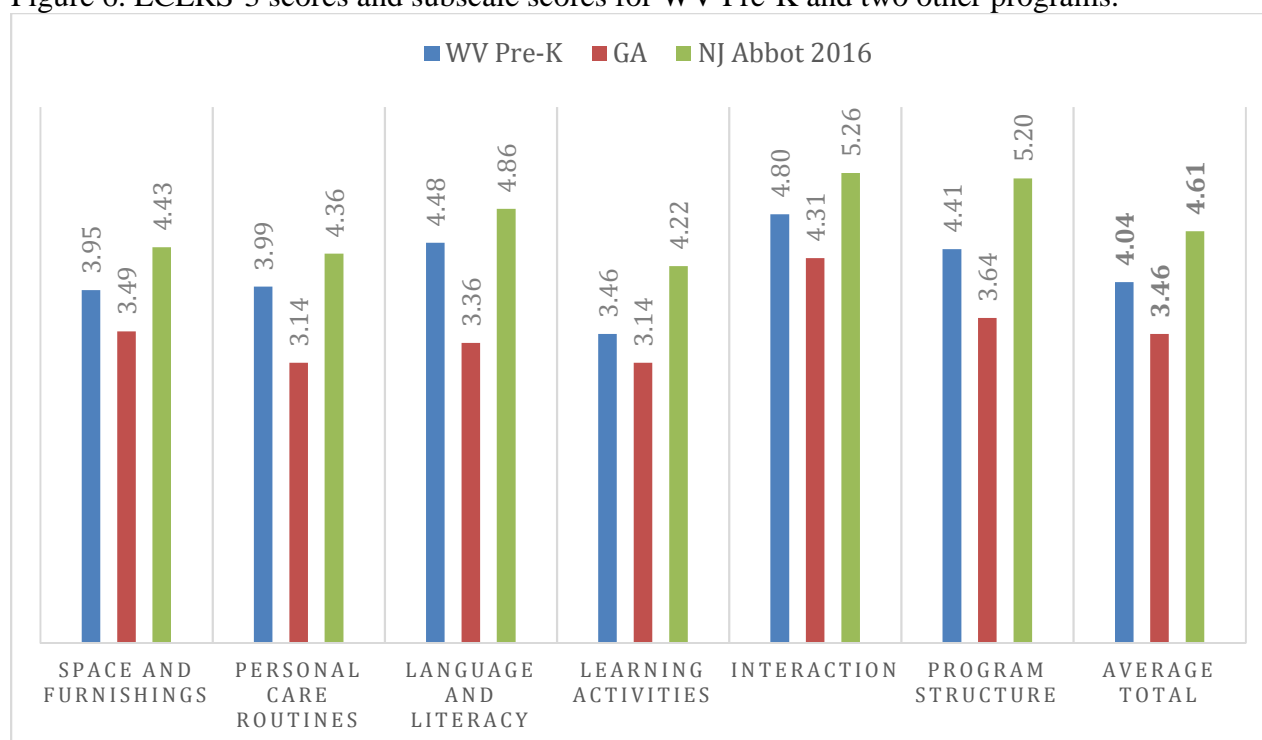
The minimum, maximum, and mean item scores for the 35 ECERS-3 items, six subscales and overall scores are shown in Table 7 below.

Table 7. ECERS-3 Item, Subscale, and Overall Means and Ranges, N = 130.

| ECERS-3 Item and Subscales	Mean	Minimum	Maximum
Overall	**4.04**	**1.60**	**6.00**
Space and Furnishings	*3.95*	*1.00*	*6.29*
1. Indoor space	4.75	1.00	7.00
2. Furnishings for care, play and learning	4.40	1.00	7.00
3. Room arrangement for play and learning	4.43	1.00	7.00
4. Space for privacy	4.42	1.00	7.00
5. Child-related display	4.08	1.00	7.00
6. Space for gross motor play	2.90	1.00	7.00
7. Gross motor equipment	2.69	1.00	7.00
Personal Care Routines	*3.99*	*1.00*	*6.25*
8. Meals/ snacks	3.47	1.00	6.00
9. Toileting/diapering	3.81	1.00	7.00
10. Health practices	3.43	1.00	7.00
11. Safety practices	5.26	1.00	7.00
Language and Literacy	*4.48*	*1.00*	*6.80*
12. Helping children expand vocabulary	4.94	1.00	7.00
13. Encouraging children to use language	4.65	1.00	7.00
14. Staff use of books with children	4.38	1.00	7.00
15. Encouraging children's use of books	4.30	1.00	7.00
16. Becoming familiar with print	4.12	1.00	7.00

Learning Activities	*3.46*	*1.00*	*6.20*
17. Fine motor	4.33	1.00	7.00
18. Art	4.08	1.00	7.00
19. Music and movement	3.04	1.00	7.00
20. Blocks	3.03	1.00	7.00
21. Dramatic Play	3.64	1.00	7.00
22. Nature/science	3.03	1.00	7.00
23. Math materials and activities	3.06	1.00	7.00
24. Math in daily events	3.72	1.00	7.00
25. Understanding written numbers	2.85	1.00	7.00
26. Promoting acceptance of diversity	4.35	1.00	7.00
27. Appropriate use of technology (N=87)	2.66	1.00	6.00
Interaction	*4.80*	*1.00*	*7.00*
28. Supervision of gross motor	3.55	1.00	7.00
29. Individualized teaching and learning	4.94	1.00	7.00
30. Staff-child interaction	5.50	1.00	7.00
31. Peer interaction	4.95	1.00	7.00
32. Discipline	5.08	1.00	7.00
Program Structure	*4.41*	*1.00*	*7.00*
33. Transitions and waiting times	4.53	1.00	7.00
34. Free play	4.43	1.00	7.00
35. Whole-group activities for play and learning (N=128)	4.30	1.00	7.00

Pre-K CLASS Results Spring 2016

The scores presented here reflect overall means for the 105 pre-K classrooms that were observed using the CLASS instrument (a smaller number of classroom were observed with CLASS due to one district opting out of the observations this year). Pre-K CLASS mean scores were 5.66 for Emotional Supports (ES), 5.09 for Classroom Organization (CO) and 2.65 for Instructional Supports (IS).

In general, the score patterns for this evaluation are slightly under the findings from the National Overview of CLASS in pre-K classrooms in 2015 (OHS, 2015). The highest scores are prevalent in the domain of Emotional Supports, with a national mean of 6.03, mid-high scores in the Classroom Organization section with a national mean of 5.80, and lower scores in the Instructional Support domain with a national mean of 2.88. Moreover, some research seems to support (Burchinal et al. 2010, OPRE, 2010) thresholds for ES and CO above 5 and IS above of 3 (and a minimum of 2.75) as necessary for quality. In the counties observed, 84% of the classrooms were above these levels on ES, 65% were so in CO and only 43% in IS were above the lower threshold of 2.75 further, only 32% above the threshold of adequate levels for IS of 3. The distribution of CLASS scores is illustrated below.

Figure 7. Pre-K CLASS distributions with Means and Standard deviations.
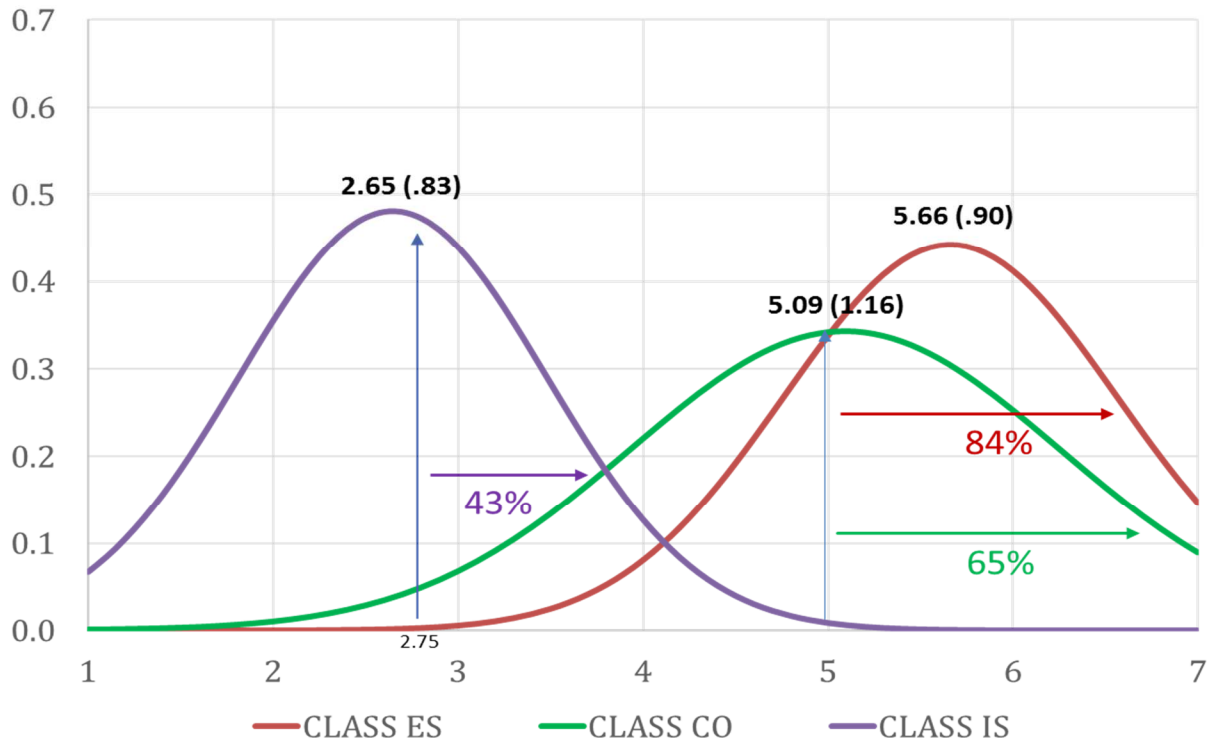
It is a common pattern for preschool evaluations to demonstrate most classrooms in the high range of quality on the Emotional Support (ES) and Classroom Organizations (CO) dimensions, but in the low- to mid-range of quality on Instructional Support (IS). However, pre-k CLASS scores in the WV counties are quite low relative to high quality programs elsewhere in the U.S. Figure 8 below illustrates WV's Pre-K CLASS scores together with those from various other programs in the U.S. for comparison purposes, including high quality programs.
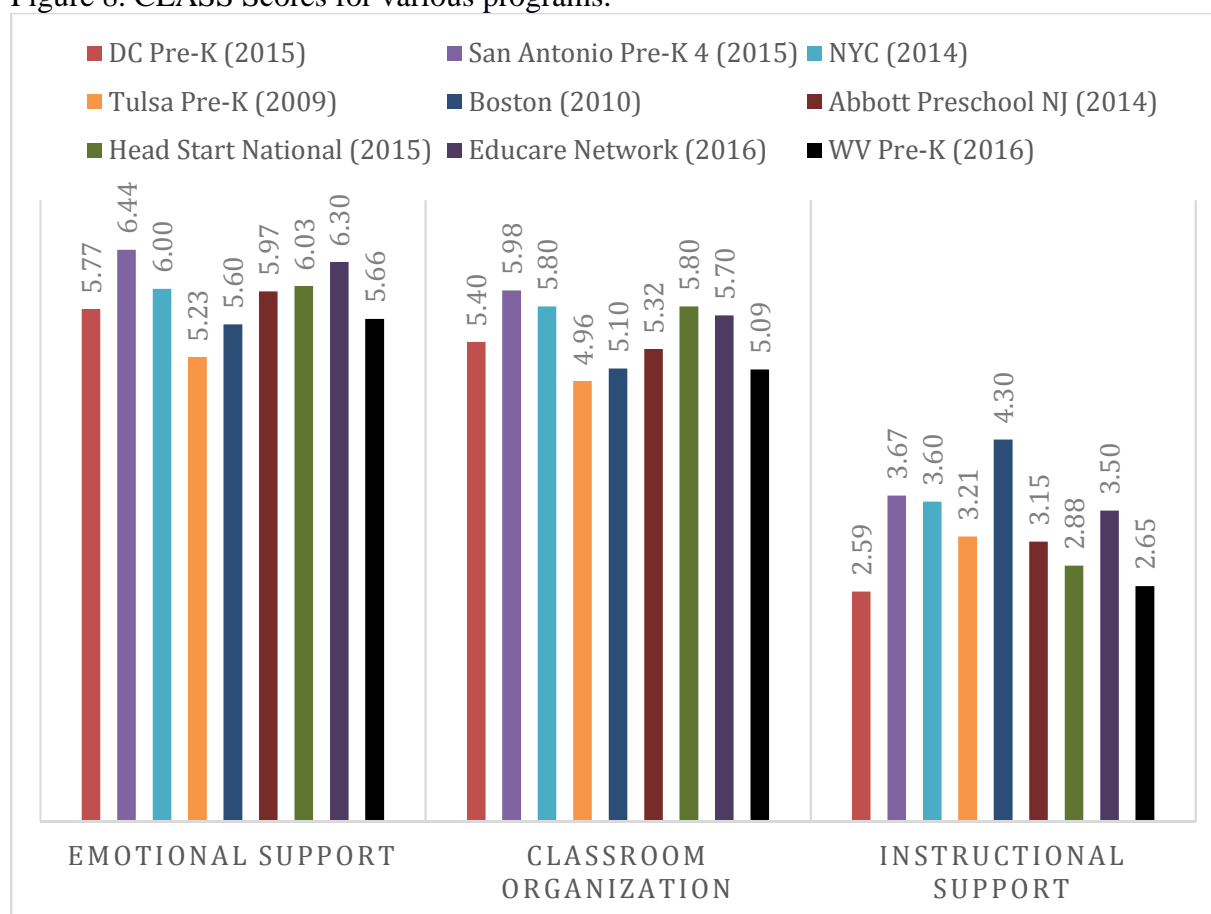
Figure 8. CLASS Scores for various programs.

Table 8 presents the minimum, maximum, and means for the 10 CLASS dimensions and the three domains.

Table 8. Pre-K CLASS Dimension and Domain Means and Ranges, N = 105.

CLASS Dimensions and Domains	Mean	Minimum	Maximum
Emotional Support Domain	5.66	2.35	6.95
1. Positive Climate	5.81	2.40	7.00
2. Negative Climate*	6.67	3.00	7.00
3. Teacher Sensitivity	5.37	1.60	7.00
4. Regard for Student Perspectives	4.80	1.20	7.00
Classroom Organization Domain	5.09	1.33	6.87
5. Behavior Management	5.29	1.00	7.00
6. Productivity	5.37	1.60	7.00
7. Instructional Learning Formats	4.62	1.40	6.60
Instructional Support Domain	2.65	1.13	5.33
8. Concept Development	2.54	1.00	5.20
9. Quality of Feedback	2.62	1.00	5.00
10. Language Modeling	2.78	1.00	6.00

*The Negative Climate dimension is reverse scored so that a high score represents "good."

Note: A lower N is available for CLASS scores than for ECERS scores due to one county opting not to participate.

Pre-K ECERS-3 and CLASS Scores by Auspice and Number of IEPs

Table 9 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 and show quality levels closer (while still on the lower end) with higher quality programs such as those reported above, with average ECERS 3 levels of 4.23, CLASS ES levels of 5.74, CLASS CO levels of 5.16 and CLASS IS levels at 2.74.

In addition, with the recent focus on moving strongly towards inclusion, we assessed differences in higher versus lower inclusion classrooms. Higher inclusion classrooms are averaging lower scores in ECERS-3 and its subscales, but this is not the case for CLASS ES and CLASS IS and only very slightly the case for CLASS CO.

Table 9. 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=15*	*N=94*	*N=21*	*N=75*	*N=55*
Space and Furnishings	3.07	4.08	4.02	4.00	3.89
Personal Care Routines	3.10	4.12	4.07	4.10	3.85
Language and Literacy	3.99	4.77	3.50	4.59	4.32
Learning Activities	2.78	3.65	3.12	3.59	3.28
Interaction	4.07	4.91	4.85	4.92	4.64
Program Structure	3.42	4.63	4.14	4.73	3.98
Overall	3.29	4.23	3.80	4.33	3.99
CLASS	*N=11*	*N=74*	*N=20*	*N=65*	*N=38*
Emotional Support	5.37	5.74	5.52	5.66	5.67
Classroom Organization	4.61	5.16	5.09	5.10	5.07
Instructional Support	2.47	2.74	2.41	2.63	2.64

Note: the number of IEPs was missing for two classrooms assessed using just the CLASS tool, therefore the total N for CLASS in the Number of IEPs category is 103 rather than 105.

Pre-K *Teacher Demographic Data*

Table 10 presents all pre-k teacher data gathered via survey during the administration of the classroom observations. Data were collected from 122 lead teachers and 117 assistant teachers. Two long-term substitutes were observed but were not included in this analysis of teacher demographic data.

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 89% having attained either High School diploma or some college or associate's degree, and only 9% having attained a BA or a higher degree. Almost 60% of both, lead and assistant teachers, have over 5 years of experience, and certification is nearly universal for lead teachers, and about half of the assistant teachers have certification.

Table 10. Pre-K Lead and Assistant Teacher Demographic Data.

		Lead teacher		Assistant Teacher	
		N = 122	%	N = 117	%
Teacher Education	GED	-	-	2	1.71%
	High School Diploma	-	-	26	22.22%
	Some college or AA	2	1.63%	76	64.95%
	Bachelor's Degree	69	56.56%	10	8.55%
	Master's Degree or higher	51	41.80%	1	0.85%
	Missing	-	-	2	1.70%
Experience in Early Childhood	0 - 5 years	49	40.16%	47	40.17%
	6 - 10 years	28	22.95%	31	26.49%
	More than 10 years	43	35.24%	33	29.46%
	Missing	2	1.63%	6	5.13%
Certification	Yes	114	93.44%	59	50.04%
	No	7	5.73%	58	49.57%
	Missing	1	0.81%	0	-

Kindergarten Classrooms

Kindergarten APEEC Results Spring 2016

Scores for the 98 Kindergarten classrooms observed using the APEEC and CLASS are presented in Figure 9 and Table 11. Unlike measures for preschool classrooms, measures of observation in kindergarten are administered less frequently and therefore we cannot provide other studies and/or programs for comparisons. However, the patterns of inadequate (<3), minimum (3-5), good (5-7) and excellent (7) of the APEEC are consistent with those of the ECERS. Figure 9 illustrates the distribution of scores across these levels: 11% of the classrooms were at inadequate levels, 79% at adequate or minimum levels, and 10% at good levels. This distribution is more compressed than the ECERS-3 distribution for preschool classrooms. That is, fewer classrooms of lower and higher quality. The overall mean APEEC score for the K sample was 3.83.
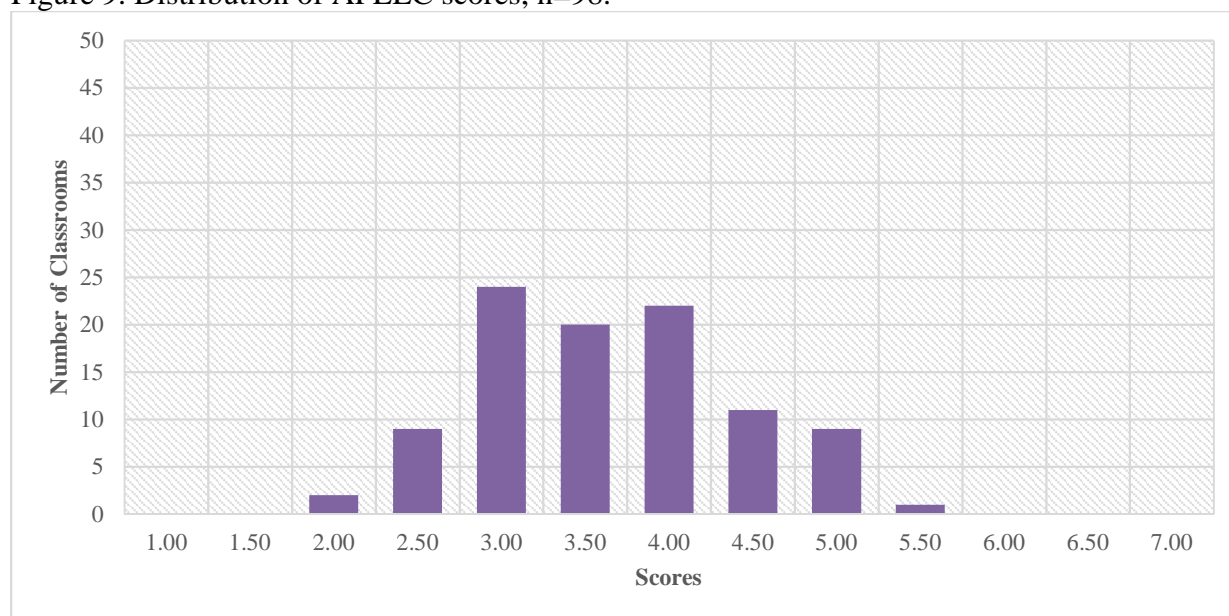
Figure 9. Distribution of APEEC scores, n=98.

Table 11 below shows WV's K APEEC scores as well as those from 2 NJ studies and a Kentucky study, all of which included K classrooms, for comparison purposes.

Table 11. APEEC scores across various studies.

	NJ K-3	NJ 2008	Kentucky	WV K
Grades	K-3	K	K-3	K
N	123	135	69	98
Overall	3.90	3.96	3.67	3.83

Table 12 reports the minimum, maximum, and mean scores for all 16 APEEC items and overall scores.

Table 12. Kindergarten APEEC Item, Subscale, and Overall Means and Ranges, N = 98.

APEEC Items and Subscales	Mean	Minimum	Maximum
Overall	3.83	2.31	5.50
Physical Environment			
1. Room Arrangement	3.30	2.00	7.00
2. Display of Child Products	2.99	1.00	6.00
3. Classroom Accessibility	3.44	1.00	7.00
4. Health and Classroom Safety	4.52	2.00	7.00
Instructional Context			
5. Use of Materials	4.51	1.00	7.00
6. Use of Computers	5.14	1.00	7.00
7. Monitoring Child Progress	4.97	2.00	7.00
8. Teacher-Child Language	3.33	1.00	7.00
9. Instructional Methods	4.18	1.00	7.00

10. Integration and Breadth of Subjects	2.49	1.00	6.00
Social Context			
11. Children's Role in Decision-Making	3.58	1.00	7.00
12. Participation of Children with Disabilities	4.52	1.00	7.00
13. Social Skills	3.91	1.00	7.00
14. Diversity	2.59	2.00	7.00
15. Appropriate Transitions	4.32	1.00	7.00
16. Family Involvement	3.62	2.00	7.00

Kindergarten CLASS Results Spring 2016

The scores reported here are the mean scores for the 98 Kindergarten classrooms that were observed using the CLASS kindergarten instrument. Quality standards for the CLASS K follow those of the CLASS pre-K. Figure 10 illustrates all three distributions for CLASS ES, CO and IS for kindergarten classrooms in the sample. The patterns follow typical patterns with IS scores lower than ES and CO. The CLASS distribution of scores for kindergarten classrooms are more compressed than for the pre-k classrooms (the lower standard deviations are in parenthesis, this means scores are more concentrated around the mean, or classes show less differences between them overall). While this means fewer numbers of low performing classrooms than at the pre-k level, this also means in this case less number of high performing classrooms. Therefore, the ES and IS scores at the K level are lower than those at the pre-K level. In addition, the number of classrooms above a level of 5 is also lower for emotional support (76%) and for classroom organization (61%). This is also the case for the number of kindergarten classrooms reaching levels above 2.75 (27%) and even more so a 3 (20%) in instructional support.

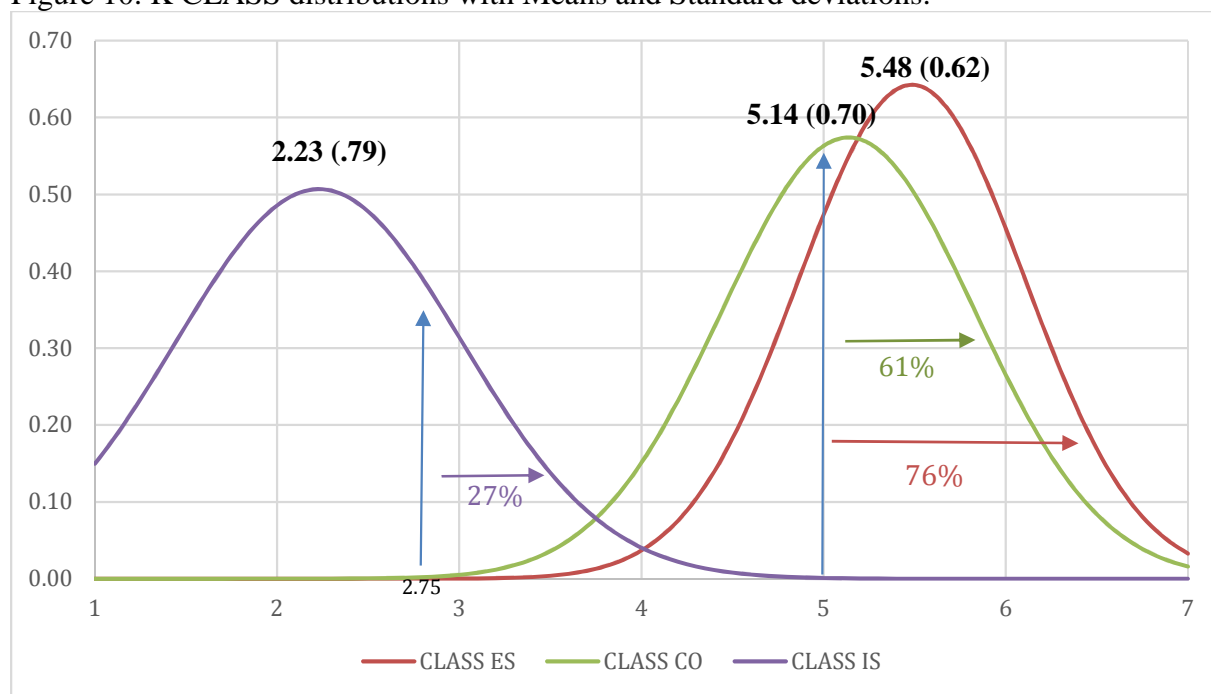Figure 10. K CLASS distributions with Means and Standard deviations.

Table 13 presents the minimum, maximum, and dimension mean scores for all 10 CLASS dimensions and the three domains.

Table 13. Kindergarten CLASS Item, Subscale, and Overall Means and Ranges, N = 98.

CLASS Dimensions and Domains	Mean	Minimum	Maximum
Emotional Support Domain	5.48	3.45	6.50
1. Positive Climate	5.62	2.80	7.00
2. Negative Climate*	6.69	3.40	7.00
3. Teacher Sensitivity	5.26	3.00	7.00
4. Regard for Student Perspectives	4.37	2.00	6.00
Classroom Organization Domain	5.14	3.33	6.53
5. Behavior Management	5.35	3.40	7.00
6. Productivity	5.37	3.00	7.00
7. Instructional Learning Formats	4.70	2.80	6.20
Instructional Support Domain	2.23	1.00	4.13
8. Concept Development	2.05	1.00	4.40
9. Quality of Feedback	2.36	1.00	4.60
10. Language Modeling	2.28	1.00	4.60

*The Negative Climate dimension is reverse scored so that a high score represents "good."

Kindergarten APEEC and CLASS Scores Distributed by IEPs

Table 14 represents the mean scores for the overall score on the APEEC and CLASS domain delineated by number of IEPs reported in the classrooms. The classrooms are categorized by the

number of IEPs in the classroom: three or fewer and four or more. Note that the number of IEPs was missing for eight classrooms observed, therefore the N in the Number of IEPs table is 90 rather than 98.

Much like was the case for pre-K classrooms, in kindergarten, APEEC average scores are higher in classrooms with fewer IEPs. For CLASS, the patterns are however slightly different, with CLASS ES and CO being higher in classrooms with fewer IEPs, with CLASS IS being equally low across the board.

Table 14. APEEC and CLASS Scores by Number of IEPs, N = 90.

	Kindergarten	
	IEP	
	0-3	4+
APEEC	*N=68*	*N=22*
Overall	3.88	3.63
CLASS	*N=68*	*N=22*
Emotional Support	5.46	5.54
Classroom Organization	5.15	4.99
Instructional Support	2.26	2.26

Kindergarten Teacher Demographic Data.

Table 15 presents all kindergarten teacher data gathered via survey during the administration of the classroom observations. Data were collected from 95 lead teachers and 74 assistant teachers.

Teachers in K classrooms show higher levels of educational attainment, while this is less evidently the case for teaching assistants. Teachers in K do appear to report higher levels of experience overall.

Table 15. Kindergarten Lead and Assistant Teacher Demographic Data.

		Lead teacher		Assistant Teacher	
		N = 95	%	N = 74	%
Teacher Education	GED	-	-	-	-
	High School Diploma	-	-	15	20.27%
	Some college or AA	-	-	51	68.92%
	Bachelor's Degree	42	44.21%	4	5.40%
	Master's Degree or	53	55.78%	-	-
Experience in	Missing	-	-	4	5.40%
Experience in Early Childhood	0- 5 years	20	21.05%	12	16.21%
	6 - 10 years	19	20.00%	22	29.72%
	More than 10 years	55	57.89%	33	44.59%
	Missing	1	1.05%	7	9.45%
Certification	Yes	94	98.94%	33	44.59%
	No	-	-	32	43.24%
	Missing	1	1.05%	9	12.16 %

Summary

This report presents the first year (2015-16) findings from the WV Universal Pre-K Program evaluation. The report has provided data analyses on the impact of the program on children's learning on receptive vocabulary and print knowledge, math, and executive functions. The classroom quality children experienced in pre-k, and the classroom quality into which they would transition as kindergarteners. The report also provides insight into how the impact of the program differed for low income children in the seven participating counties, and for females, relative to the average. In addition to describing these findings, the report provides comparable findings from other evaluation studies to contextualize the results.

The analyses reported highlight strong results in children's receptive vocabulary, print knowledge, math and executive function skills, as well as evidence that lower income children benefitted more in print knowledge and math. Results for girls were lower across all outcomes. The results are quite large and comparable to those of high quality programs that have been similarly assessed through RDD.

Pre-K classrooms in participating counties average 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 show lower overall levels of quality as measured by the APEEC and the CLASS Instructional Support, but similar to the pre-k classroom on the CLASS Emotional Support and Classroom Organization domains. Appendix B report provides considerable depth into classroom quality observed and recommendations for improvement in classroom quality.

References

Barnett, W. S., Jung, K., Youn, 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.

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.

Gormley, W. T. (2008). The effects of Oklahoma's pre-K program on Hispanic students. Social Science Quarterly, 89(4), 916-36.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica, 69(1), 201-209.

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.

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. Review of economics and statistics, 86(1), 226-244.

Lee, D. S., & Lemieuxa, T. (2010). Regression discontinuity designs in economics. Journal of economic literature, 48(2), 281-355.

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.

Lonigan, C. J., Wagner, R. K., & Torgesen, J. K. (2007). Test of preschool early literacy: TOPEL. Austin, TX: Pro-ed.

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

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 StudyChapel 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//Reseñas Educativas.

Qi, C. H., Kaiser, A. P., Milan, S., & Hancock, T. (2006). Language performance of low-income African American and European American preschool children on the PPVT–III. Language, Speech, and Hearing Services in Schools, 37(1), 5-16.

Van der Klaauw, W. (2008). Regression–discontinuity analysis: a survey of recent developments in economics. Labour, 22(2), 219-245.

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.

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.

Appendices

Appendices that make part of this report, but are separate documents are:

Appendix A. RDD Methodology and Full estimation: This appendix contains a brief explanation of the RDD estimation method and full set of results.

Appendix B. Classroom Quality Report. This appendix contains in depth explanation of results observed for classroom quality, with detail that would support continuous improvement efforts.