Abstract Title Page

Title: A Regression Discontinuity Analysis of the Impact of *Building Blocks* in an Urban Public Prekindergarten Program and Associations between Fidelity-to-Curriculum and Child Outcomes

Author(s):
Christina Weiland, Harvard University (PRESENTER) (christina_weiland@mail.harvard.edu)
Hadas Eidelman, Harvard University (hadas_eidelman@mail.harvard.edu)
Hirokazu Yoshikawa, PhD, Harvard University (hirokazu_yoshikawa@gse.harvard.edu

Abstract Body

Background / Context:

The number of states offering publicly funded prekindergarten programs increased from 10 in 1980 to 38 in 2009 (Gormley, Gayer, Phillips, & Dawson, 2005; NIEER, 2009) and as of 2009, approximately 30% of U.S. 4-year-olds were enrolled in public prekindergarten programs (NIEER, 2009). Yet, only a handful of studies have examined the causal impacts of these programs on child school readiness (Gormley et al., 2005; Hustedt, Barnett, Jung & Goetze, 2009; Hustedt, Barnett, Jung & Thomas, 2007; Wong, Cook, Barnett, & Jung, 2007). Using regression discontinuity, these studies have found small to moderate positive impacts on children's language, literacy and numeracy skills.

However, due to data limitations, published studies of large-scale publicly funded public prekindergarten programs have not fully addressed questions regarding under what conditions these programs achieved impacts at scale. None of the examined contexts in the studies published to date had a consistent curriculum in place. This is an important gap in the literature for several reasons. First, theory and some empirical research suggest that implementing an intentional curriculum may improve child outcomes by helping to ensure program quality, by keeping children engaged and challenged and by building specific skills targeted by the curriculum (Klein & Knitzer, 2006; NAEYC & NAECS/SDE, 2003). Second, the limited evidence regarding the treatment conditions in evaluated prekindergarten programs is particularly problematic from a district and policy perspective. Public prekindergarten programs are being increasingly held to state standards for both literacy and mathematics instruction (NIEER, 2008), and implementing curricula for these domains in preschools is an increasing reality and requirement for districts and teachers.

Purpose / Objective / Research Question / Focus of Study:

Using data from an urban public pre-k program, we add to and extend the emerging evidence base of the effects of public prekindergarten programs on child school readiness. We also use data collected in treatment classrooms to examine associations between teacher characteristics, fidelity-to-curricula, dosage and child outcomes. Our primary research questions are:

1) What is the causal impact of attending a prekindergarten program that implemented the *Building Blocks* mathematics curriculum at scale across an urban public school district on children's mathematics, language, literacy, executive function and emotional development?
2) Within the treatment group, are teacher characteristics predictive of fidelity-to-curriculum and dosage?
3) Is higher *Building Blocks* fidelity-to-curricula and dosage associated with higher student outcomes, controlling for student- and teacher-level characteristics?

Setting:

Research took place in a large urban public school district in the Northeast. All prekindergarten programs were located in public elementary schools.

Population / Participants / Subjects:

Teacher-level. In spring 2009, we invited all district elementary schools with prekindergarten classrooms to participate in a fidelity-of-implementation study. In total, 64% of eligible elementary schools Principals (N=41) agreed to participate and 61% (N=74) of BPS

prekindergarten teachers agreed to participate. Within schools in which principal agreed to participate, 82% of teachers agreed to be observed. There were no significant differences between participating and non-participating schools on four school-level characteristics and there were no significant differences between participating and non-participating teachers on 8 teacher-level characteristics.

Child-level. In Fall 2009, children in a citywide 4-year-old prekindergarten program and all children who attended the program in the previous year were eligible for the study. For a child to participate in the study, the principal, classroom teacher, and parent/guardian of the child had to consent to participate. Out of 79 elementary schools with eligible children, 12 principals declined to participate (15%). Over 93% of eligible teachers in participating schools agreed to participate. Within participating classrooms in the 67 participating schools, 69.2% of eligible children returned consent forms, for a total sample size of 2,018. This represents 55% of eligible children in the district. As evident in Table 1, the final sample of participating children is racially and linguistically diverse. Our sample for RQs 2 and 3 consists of the treatment group children tested in Fall 2009 for the impact study (N=707) who were also enrolled in classrooms (N=74) in which the teacher participated in the fidelity data observations in Spring 2009.

Intervention / Program / Practice:

Any child within the city who turns four by September 1 in a given year can apply for the prekindergarten program. All prekindergarten classrooms in the districts are staffed with at least one teacher with at least a B.A. and one paraprofessional (adult-child ratio is about 1:10). Teachers are paid on the same scale as K-12 teachers. Intending to promote classroom quality, the district implemented the literacy curriculum *Opening the World of Learning (OWL)*[1] (Schickedanz & Dickinson, 2005) and the mathematics curriculum *Building Blocks* (Clements & Sarama, 2007a) system-wide in 4-year-old classrooms in the 2007-2008 school year. Treatment children in our study attended the program in the 2008-2009 school year, while control children attended the program in the 2009-2010 school year.

Research Design:

To address RQ1, we employ a regression discontinuity design to obtain causal child-level estimates, with the birthday cutoff for entry into the program in a given year as the source of exogeneity. Importantly, the district strictly enforces the cutoff; in recent years, no child has been admitted into the program when their birthdate suggests they should not.

To address RQ2 and RQ3, we fit the path model shown in Figure 1. This model allowed us to incorporate multiple outcomes and mediators and to simultaneously examine relationships between the hypothesized teacher characteristics, fidelity mediators, and child-level outcomes.

Data Collection and Analysis:

Fidelity-of-curricula, dosage, and quality. For *Building Blocks*, we measured fidelity-to-curriculum using the developers' fidelity measure (*TRIAD Near Fidelity;* Sarama & Clements, 2009), which includes a general curriculum section along with sections that focus on specific components of the curriculum. The measure includes separate sections and items for each component of the *Building Blocks* curriculum and items are scored either as dichotomously (yes/no) or using a five-point Likert scale (where 1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree). We also created a dosage index for *Building Blocks* based on coach-

[1] We also collected dosage and fidelity-to-curricula data on the *OWL* but given the topic of the panel, we will not present these data and analyses at this SREE conference.

completed ratings of how often teachers implemented a given component of each curricula in a given week. From these ratings, we constructed a dosage index, indicating what portion of the full intended *Building Blocks* curriculum was delivered in a given classroom. We also created a general classroom quality rating scale, with items scored using a five-point Likert scale (where 1=almost never, 2=rarely, 3=sometimes, 4=usually, 5=almost always). This measure included items asking about the positive climate in the classroom, teacher sensitivity, and teacher reflectiveness on his/her teaching practice.

We trained nine district Early Childhood Coaches to conduct classroom observations during the treatment year. All early childhood coaches held masters degrees. On average, early childhood coaches had previously taught in early childhood classrooms for 8.8 years (range 2-20 years; standard deviation of 4.9 years) and had worked as a district early childhood coach an average of 3.3 years (range .5-7 years; standard deviation of 2.2 years).

Once all data were collected, two early childhood coaches, two study team members, and the curricula developers independently rated each *Building Blocks* items as either "specific to the curricula" or "general early childhood practice." This was in accordance with Cordray and Hulleman (2009) and Munter and Garrison (2010) who suggest focusing on components of fidelity that are specific to an intervention. Items specific to either curricula were broken into three scales within each curricula: teaching strategies, classroom structures, and curricula materials. In confirmatory factor analysis, we found no support for a three-factor structure. Rather, fidelity within each curricula was best represented as a unitary construct. Likewise, general quality was best represented as a unitary construct. Scores reported in this study are unit-weighted averages of 20 *Building Blocks* items.[2] We used MPlus 6.0 to fit the path model shown in Figure 1. See Table 2 for descriptive statistics on the *Building Blocks* fidelity-to-curricula and dosage variables.

Child-level outcomes data collection. Children were tested by study-trained child assessors. All assessors were college educated and approximately one third held masters degrees. The assessors visited classrooms in Fall 2009, as close to the start of the school year as possible. See Table 3 in Appendix B for a list of child-level measures used in our study.

Our implementation of the RD framework is guided by the advice of Lee and Lemieux (2009), by the strategy and organization of Wong et al. (2007), and by the recently released *What Works Clearinghouse* guidelines (Schochet et. al, 2010). We first conduct a graphical analysis of the relationship between the outcome and smoothed function of child age on either side of the cutoff. These graphs give some indication of functional form, as well as whether there is indeed a "jump", or difference between the two groups, at the cutoff. Second, because identifying the correct functional form of the continuous assignment variable is one of the chief challenges in RD analysis (Lee & Lemieux, 2009), we fit a series of regression model specifications, including polynomials, interaction terms and non-parametric models, to the raw data.

Our primary equation for fitting regression models is as follows:

$$OUTCOME_{ij} = \beta_0 + \beta_1 TREAT_{ij} + \beta_2 Age_{ij} + \beta_3 TREAT_{ij} * AGE_{ij} + \delta X_{ij} + (\varepsilon_{ij} + \delta_{ij}) \quad (1)$$

where *OUTCOME* is a child-level test score, *TREAT* is a dummy variable that takes on the value of 1 if the student's birthday is on or before September 1, 2004 and the value of 0 if not, *AGE* is a smooth function of the student's age measured in days and centered on the September 1 birthday cutoff, *TREAT*AGE* is an interaction term that allows the effect to vary on either side of the cutoff, X is a vector of student demographic covariates, ε is the error term associated with

[2] We dropped 3 *Building Blocks* items due to lack of variation and/or skewness.

students and δ is the error term associated with classrooms. Subscript *i* denotes students and subscript *j* denotes classrooms. In all regression models, we adjust standard errors for clustering at the classroom level and include school fixed effects. In all regression models, we use multiple imputation (with 50 imputations) to account for missing data in accordance with Graham (2009).

Findings / Results:

As evident from Figure 2 and 3, we found significant ($p<0.05$), small-to-moderate positive effect sizes on all assessments. The effect size for numeracy skills was 0.58 (Applied Problems) and the effect size for numeracy/geometry was 0.49 (REMA-Short). The effect size for pre-reading and reading skills was 0.62 and vocabulary, 0.45. Executive functioning effect sizes were in the small range but all positive: 0.20 (inhibitory control), 0.27 (cognitive flexibility), and 0.23 (working memory). For emotional development, the effect size was 0.18. Results were robust across multiple bandwidths and model specifications, and in other standard RD robustness checks, we find no reason to doubt our findings.

Regarding the hypothesized links between teacher characteristics and dosage and fidelity-to-curriculum, we found some evidence that holding a Bachelors degree in Early Childhood Education (ECE) was positively related to dosage and fidelity-to-curricula.[3] Regarding the relationship between dosage, fidelity-to-curricula and child outcomes, we found that dosage and fidelity-to-curricula were not significant predictors of children's outcomes within the treatment group, nor was teacher quality. In interpreting these results, we note there was little variation in dosage and fidelity-to-curricula in our sample.

Conclusions:

Our results add to the growing literature on the causal effects of large-scale state-funded prekindergarten programs. We find that a universal publicly funded prekindergarten program had positive impacts on child early numeracy, language, literacy, executive function, and emotional development. In Table 4, we place our main impact language, literacy, and early numeracy results in the context of other RD prekindergarten studies. Our effect sizes are larger than those achieved in any RD prekindergarten study to date, which is particularly notable given that ours is the only RD prekindergarten context in which there were uniform curricula in place.

Due to the RD design of our study, we are unable to fit causal mediation models that would help us to definitively identify the causal mechanisms underlying our results. The path model we presented here using fidelity-to-curricula and dosage data on the treatment children in their treatment year found no significant associations between fidelity-to-curricula and dosage and child outcomes. By the SREE Fall Conference, we will have fine-tuned these path models further and will explore using Structural Equation Modeling with latent representations of fidelity-to-curricula. SEM analysis would give us more precise results by removing measurement error from the latent constructs in our structural models. We will test whether fidelity-to-curricula and dosage mediate effects on executive function outcomes, given that mathematics draws heavily on executive skills like working memory and inhibitory control. Further analysis will allow us to better examine whether stronger effects in our study compared to previous prekindergarten RD studies thus could at least partially be a function of the chosen math and literacy curricula and the level of curriculum implementation in the district.

[3] We also fit path models that included direct paths from teacher level of education to children's outcomes, none of which were statistically significant at the 0.05 level. Additionally, we tested for potential moderation between teacher quality and fidelity of curriculum implementation, and found no evidence to support a moderation effect.

Appendices

Appendix A. References

Cordray, D.S. & Hulleman, C. (2009). Assessing Intervention Fidelity in RCTs: Models, Methods and Modes of Analysis. Presentation for the *IES Research Conference*.

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), 315-344.

Dunn, L.M., & Dunn, L.M. (1997). Peabody Picture Vocabulary Test-Third Edition. Bloomington, MN: Pearson Assessments.

Frye, D., Zelazo, P.D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development,* 10(4), 483-527.

Gathercole, S.E. & Pickering, S.J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Education Psychology, 70*, 177-194.

Gormley, W.T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology, 41*(6), 872-884.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549-576.

Hustedt, J. T., Barnett, W. S., Jung, K. & Goetze, L. D. (2009). *The New Mexico PreK evaluation: Results from the initial four years of a new state preschool initiative - final report.* Retrieved August 29, 2010, from http://nieer.org/pdf/new-mexico-initial-4-years.pdf .

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.* The National Institute for Early Education Research, Rutgers University. Retrieved August 30, 2010 from http://www.arkansas.gov/childcare/abc/pdf/longreport.pdf.

Klein, L., & Knitzer, J. (2006). *Effective preschool curricula and teaching strategies.* Retrieved May 7, 2008, from http://www.nccp.org/publications/pdf/text_668.pdf .

Lee, D. S., & Lemieux, T. (2009). Regression Discontinuity designs in economics. *NBER Working Paper no. 14723.* Retrieved August 29, 2010, from http://www.nber.org/papers/w14723.

Munter, C. & Garrison, A. (2010). Evaluating Math Recovery: A case of measuring implementation fidelity of an unscripted, cognitively-based intervention. Presentation, Society for Research on Educational Effectiveness.

National Association for the Education of Young Children (NAEYC) and National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). (2003). *Early childhood curriculum, assessment, and program evaluation: Building an effective, accountable system in programs for children birth through age 8.* Retrieved May 7, 2008, from www.naeyc.org/about/positions/pdf/CAPEexpand.pdf .

National Institute for Early Childhood Research (NIEER). (2008). "The State of Preschool 2007." Retrieved May 8, 2008 from http://nieer.org/yearbook/.

National Institute for Early Childhood Research (NIEER). (2009). *The state of preschool 2009.* Retrieved August 29, 2010, from http://nieer.org/yearbook/.

Ribordy, S.C., Camras, L.A., Stefani, R. & Spaccarelli, S. (1988). Vignettes for recognition research and affective therapy with children. *Journal of Clinical Child Psychology, 17,* 322-325.

Sarama, J. & Clements, D. (2009). *Fidelity of Implementation (TRIAD Near Fidelity)*.

Schochet, P., et al. (2010). *Standards for Regression Discontinuity designs.* Retrieved August 29, 2010, from http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf.

Weiland, C., Wolfe, C., Hurwitz, M., Clements, D., Sarama, J., & Yoshikawa, H. (2011). Psychometrics and validation of the short form of an early mathematics assessment. Poster presented at the biennial meeting of the Society for Research in Child Development, Montreal.

Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson Tests of Achievement*. Itasca, IL: Riverside Publishing.

Wong, V. C., Cook, T. D., Barnett, W. S. & Jung, K. (2007). *An effectiveness-based evaluation of five state prekindergarten programs using Regression Discontinuity.* Retrieved April 20, 2008, from http://nieer.org/research/topic.php?TopicID=1024.

Appendix B. Tables and Figures

Table 1: Descriptive characteristics of sample

Variable	Overall (N=2018)	Born before cutoff (N=969)	Born after cutoff (N=1049)
Attendance zone is the North Zone	0.28	0.29	0.26
Attendance zone is the East Zone	0.44	0.45	0.44
Attendance zone is the West Zone	0.28	0.26	0.30
English only home language	0.50	0.48	0.53
Spanish home language	0.27	0.28	0.27
Other home language	0.22	0.24	0.20
Black	0.27	0.28	0.25
White	0.18	0.18	0.19
Hispanic	0.41	0.39	0.42
Asian	0.11	0.11	0.11
Other race/ethnicity	0.03	0.03	0.03
Special Needs	0.09	0.11	0.08
Free/reduced lunch receipt	0.69	0.72	0.66
Male	0.51	0.52	0.50
Previously attended family daycare	0.07	0.08	0.06
Previously attended Head Start	0.16	0.16	0.16
Did not attend any care program previously	0.34	0.34	0.33
Previously attended public preschool	0.11	0.11	0.11
Previously attended private center care	0.33	0.31	0.35

*Note: one child born after the cutoff is missing all information in this table. 76 children (4% of sample) are missing pre-program care data.

Table 2: Descriptive statistics of teachers' ratings on adherence and dosage and quality (*n*=74)

Fidelity Measure	*Mean*	*Range*	*SD*
BB Adherence	3.01	1.54-3.87	0.49
BB Dosage	73.51	0-100	23.69
Quality	4.16	2.69-5.00	0.61

Table 3: Child Assessment Battery

Name of Assessment	Domain	Specific construct
Peabody Picture Vocabulary Test – III (PPVT-III) (Dunn & Dunn, 1997)	Language	Receptive vocabulary
Woodcock-Johnson Letter-Word Identification (Woodcock, McGrew & Mather, 2001)	Pre-Literacy	Pre-reading and reading
Woodcock-Johnson Applied Problems (Woodcock, McGrew & Mather, 2001)	Numeracy	Early math reasoning and problem-solving abilities
Research-based Elementary *Math* Assessment Short (REMA) (Weiland, Wolfe, Hurwitz, Clements, Sarama & Yoshikawa, 2011)	Numeracy	Comparing/ordering, verbal counting/counting strategies, arithmetic, number recognition and subitizing, geometric, measuring and patterning capacities
Forward Digit Span (Gathercole & Pickering, 2000)	Executive function	Working memory (phonological loop)
Backward Digit Span (Gathercole & Pickering, 2000).	Executive function	Working memory (central executive)
Dimensional Change Card Sort (DCCS) (Frye, Zelazo & Palfai, 1995)	Executive function	Attention Shifting
Pencil Tapping (Diamond & Taylor, 1996)	Executive function	Inhibitory control
Emotion Recognition Questionnaire (Ribordy, Camras, Stafani & Spacarelli, 1998)	Emotional development	Emotion identification/labeling

Table 4: Comparison of effect sizes across published RD prekindergarten studies

	PPVT-III	Letter-Word Identification	Applied Problems
Boston	0.45***	0.62***	0.58***
Tulsa	--	0.80***	0.38*
Michigan	-0.16	--	0.47*
New Jersey	0.36*	--	0.23*
South Carolina	0.05	--	--
West Virginia	0.14	--	0.11
Oklahoma	0.29*	--	0.35
New Mexico, Year 1	0.35+	--	0.38+
New Mexico, Year 2	0.25+	--	0.50+
New Mexico, Year 3	0.17+	--	0.43+

***p<0.001; **p<0.01; *p<0.05 ; + results statistically significant but level of significance not reported.
Citations: Tulsa (Gormley, Gayer, Phillips, & Dawson, 2005); MI, NJ, SC, WV, OK (Wong et al., 2007); NM (Hustedt, Barnett, Jung & Goetze, 2009).
Note: All cited studies use the standard deviation of the control group as the denominator in calculating effect sizes.

Figure 1: Path diagram showing fitted coefficients for the relationships between teacher level of education, teacher quality, fidelity-to-curriculum, and children's mathematics outcomes for children who participated in the prekindergarten program in 2008-2009

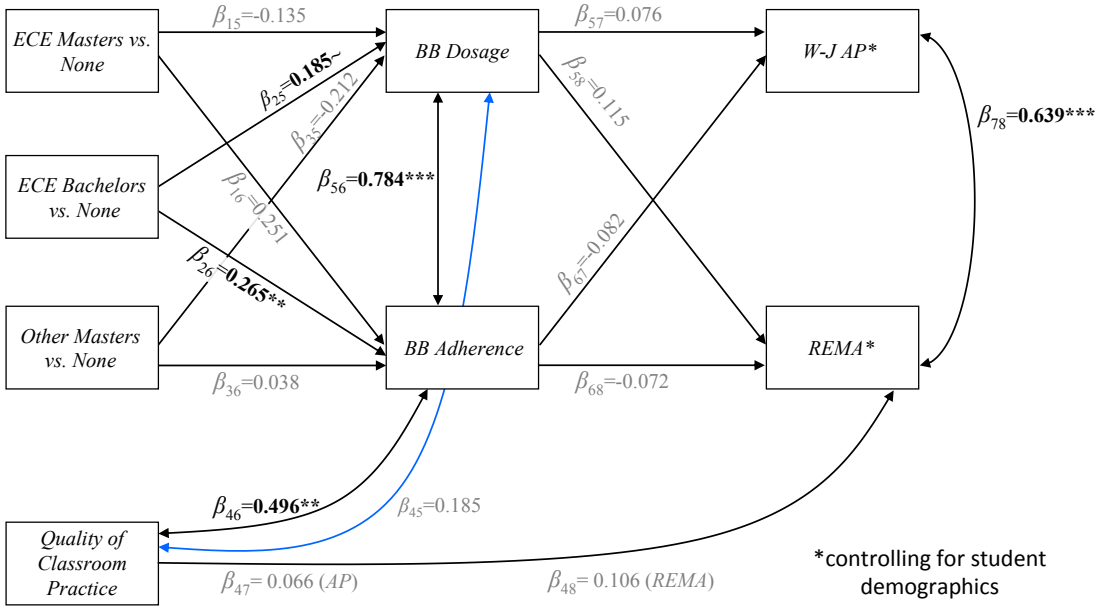
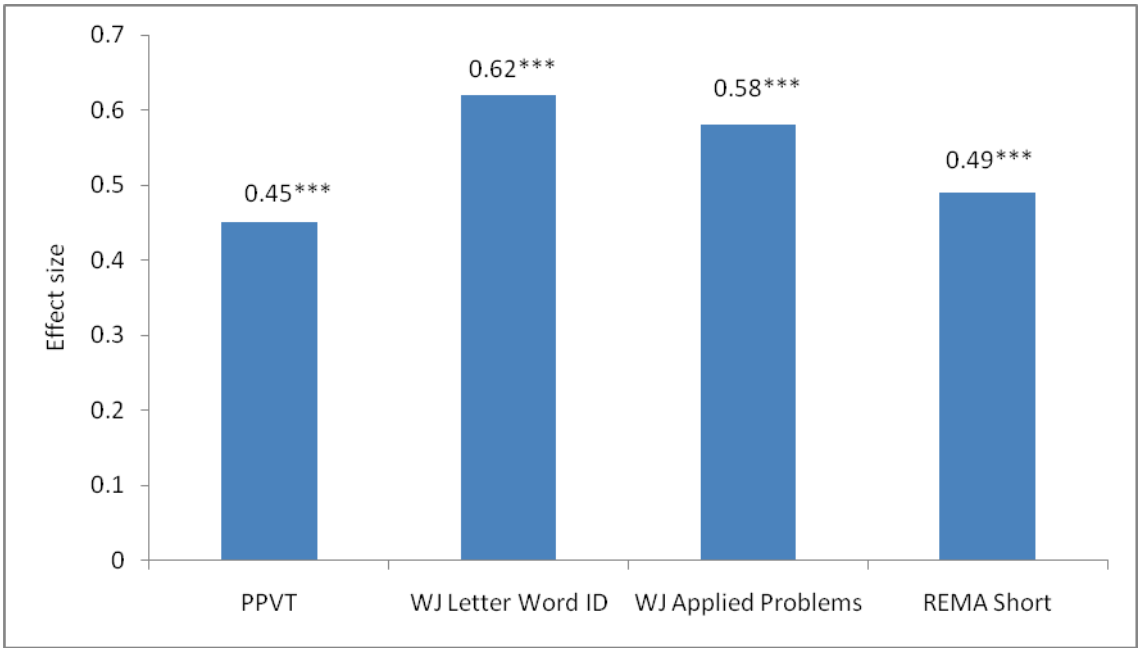
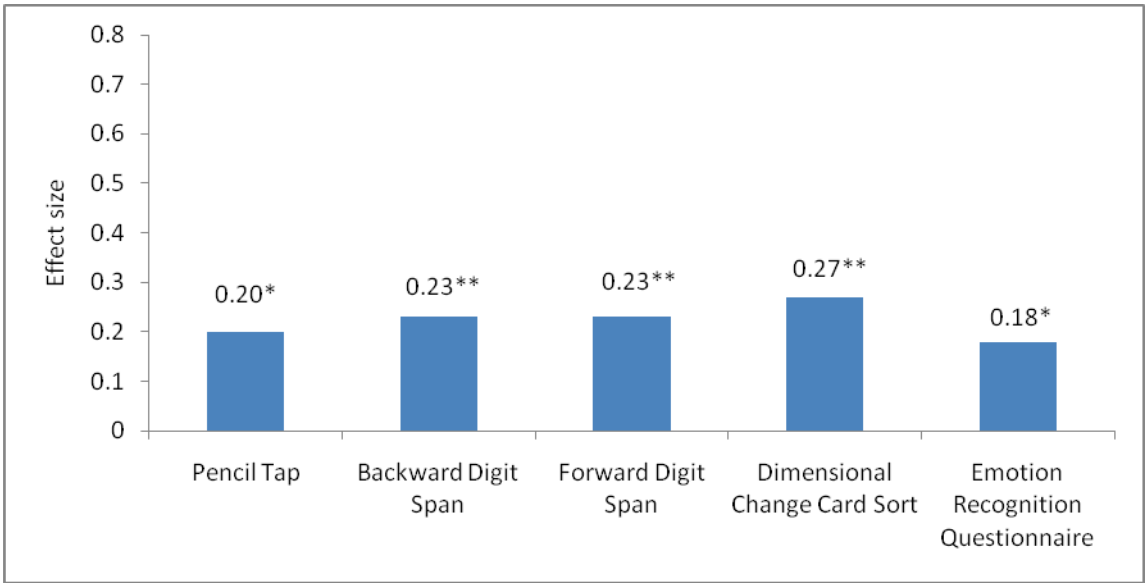


Figure 2: Language, literacy, and mathematics effect sizes for children who participated in the prekindergarten program in 2008-2009



Note: Effect sizes calculated using the standard deviation of the control group. Effect sizes shown are based on models with a bandwidth of 365 days on either side of the cutoff and are robust to functional form and bandwidth choice.
***p<0.001; **p<0.01; *p<0.05

Figure 3: Executive function and emotional development effect sizes for children who participated in the prekindergarten program in 2008-2009



Note: Effect sizes calculated using the standard deviation of the control group. Effect sizes shown are based on models with a bandwidth of 365 days on either side of the cutoff and are robust to functional form and bandwidth choice.
***p<0.001; **p<0.01; *p<0.05