Abstract Title Page
Not included in page count.

Title: The Impact of an Urban Universal Public Prekindergarten Program on Children's Early Numeracy, Language, Literacy, and Executive Function Outcomes

Author(s): Christina Weiland and Hirokazu Yoshikawa

Abstract Body

Limit 5 pages single spaced.

Background / Context:
Description of prior research and its intellectual context.

According to the literature, high-quality early childhood education equips children with the cognitive and academic skills required to be successful in elementary school and beyond (Gilliam, 2005; National Forum on Early Childhood Program Evaluation, 2007). Landmark studies of particularly intensive interventions have found positive impacts that last into adulthood and that are highly cost effective (Campbell, Ramsey, Pungello, Sparling & Miller-Johnson, 2002; Currie, 2001; Karoly, Kilborn & Cannon, 2005; Schweinhart, Barnett & Belfield, 2005; Yoshikawa, 1995), with estimated cost-benefit ratios ranging from about $3 to $12 for every $1 spent (Heckman, Moon, Pinto, Savelyev & Yavitz, 2010; Temple & Reynolds, 2007).

It is unclear in the literature, however, whether or how such impacts can be maintained at scale. Generalizing the results of small-scale experiments to at-scale programs raises a host of problems, as small-scale programs are generally more resource-intensive and have extensive involvement of the creators of the intervention that simply is not possible at scale (Murnane & Willett, 2010; Yoshikawa, Rosman & Hsueh, 2002). This may explain why many large-scale preschool programs often have weaker effects than model programs (Barnett, 1995; Currie & Thomas, 1995).

Research on an increasingly common large-scale preschool model - state-funded prekindergarten – has lagged behind its policy scale up. The number of states offering such programs increased from 10 in 1980 to 38 in 2009 (Gormley, Gayer, Phillips, & Dawson, 2005; NIEER, 2009). Yet, only a handful of studies have examined the causal impacts of these programs (Gormley, Gayer, Phillips, & Dawson, 2005; Hustedt, Barnett, Jung & Goetze, 2009; Hustedt, Barnett, Jung & Thomas, 2007; Wong, Cook, Barnett, & Jung, 2007). Encouragingly, these studies have found small to moderate positive impacts on children's language, literacy and numeracy skills. However, no study to date has examined the causal impact of a large-scale state-funded public prekindergarten program on child executive functioning, a domain many developmentalists consider to be an important component of school readiness.

Existing studies of large-scale state-funded public prekindergarten programs also do not fully address the question of under what conditions these programs can achieve impacts at scale, as there was no consistent curriculum in place in any of the programs examined in this literature (Gormley, Gayer, Phillips, & Dawson, 2005; Hustedt, Barnett, Jung & Goetze, 2009; Hustedt, Barnett, Jung & Thomas, 2007; Wong, Cook, Barnett, & Jung, 2007). 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). Finally, existing prekindergarten studies have not explored sensitivity of results to some common methodological challenges, including possible non-comparability of treatment and control children due to differential time to join and attrit from the intervention.

Purpose / Objective / Research Question / Focus of Study:
Description of the focus of the research.

We add to and extend the emerging evidence base of the effects of public preschool programs on child school readiness. Using a quasi-experimental, Regression Discontinuity (RD) design, we estimate the impacts of a year of a universal preschool program on children's early numeracy, language, literacy, and executive function skills, both for the overall population and for several subgroups. While we find impacts similar to those reported in other public prekindergarten RD studies, we make a unique contribution to the literature, as ours is the first causal study of a universal prekindergarten program in which a uniform curricula was in place across the district and in which we have information about the type of care experienced by control children during the treatment year. At SREE, we will also present results from robustness checks new to this literature, including sensitivity of results to differential time to join and attrit from the intervention.

Setting:
Description of the research location.

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:
Description of the participants in the study: who, how many, key features or characteristics.

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 our study, so long as they were placed in non-special education only classrooms (criterion 1), they did not permanently withdraw from the district before October 1 of their prekindergarten year (criterion 2), and they were enrolled in school at least seven days (criterion 3). Special education-only classrooms were excluded due to concerns about the appropriateness of the assessment battery for children who were not mainstreamed. Criteria (2) and (3) were employed to filter out children who register for prekindergarten but never show up or show up for so few days that they were exposed to very little of the program and could not be located to be assessed.

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.

Intervention / Program / Practice:
Description of the intervention, program or practice, including details of administration and duration.

The program is universal in design, as any child within the city who turns four by September 1 in a given year can apply for the program. However, there is more demand than there is supply, with a lottery system determining which students are allotted prekindergarten seats. Currently, approximately 35% of the city's 4-year-old population attends 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). Prekindergarten 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)* (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:
Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

We employ a Regression Discontinuity Design to obtain causal 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. Approximately 94% of children offered a spot in the 4-year-old program in 2008-2009 enrolled in the program for at least one day.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

Children were tested by study-trained child assessors. All assessors were college educated and approximately one third held masters degrees. On average, the battery of tests took approximately 45-50 minutes to administer. Testers were instructed to test children in one session if possible but to divide the session into smaller segments if children showed signs of fatigue. Because of the session length, we randomly varied the order of the tests to limit the possibility of systematically biasing results due to child fatigue. The assessors visited classrooms in Fall 2009, as close to the start of the school year as possible. Measures of pre-literacy, numeracy and executive function skills were collected. See Table 2 in Appendix B for a list of 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. As shown in Figure 1, on either side of the cutoff, we superimpose a linear regression line and a smoothed, locally weighted non-parametric regression line on a scatterplot of the raw data. 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. We compare relative fit across models and purposely overspecify the model as a robustness check. Although less efficient than underspecifying, overspecifying yields more unbiased coefficients (Trochim, 1984) and has been used as a strategy in other early childhood RD designs (Gormley et al., 2005; Wong et al., 2007).

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 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. In all regression models, we use multiple imputation (with 50 imputations) to account for missing data in accordance with Graham (2009).

We extend our approach to examine effects by subgroups by fitting separate regressions for each subgroup. Subgroups of interest are defined by home language (English, Spanish, other), race (Black, Latino, White, Asian), free/reduced lunch, gender, special needs status, and pre-program childcare experience (Head Start, private, family daycare, none, public).

We also perform a series of robustness checks. Threats to the internal validity of our results include: 1) treatment misallocation at the cutoff; 2) non-smooth or discontinuous variation of observed and unobserved student characteristics around the cutoff; 3) discontinuities in the outcomes at points other than the cutoff; 4) incorrect specification of the functional form; and 5) sensitivity of results to choice of bandwidth. All of these threats could result in either an over- or under-estimation of the true impact of the treatment. We do not have space in this proposal to discuss how we address each threat specifically; however, our robustness checks follow current best practices in the RD literature (Imbens & Lemieux, 2008; Lee & Lemieux, 2009).

At SREE, we will present results from several robustness checks new to the prekindergarten RD literature. For example, we will explore sensitivity of results to differential time to enter the treatment and control groups; in our study, treatment children had a full school year to enter prekindergarten (2008-2009), while control children had only a few months (Fall 2009). To address this challenge, we will examine the sensitivity of results to different definitions of the sample (i.e. children who were enrolled the full time vs. those enrolled at least one month vs. those enrolled at least one day). To equalize attrition time, we will obtain administrative records for the control group that cover the same amount of time as the treatment group (i.e. through December of the kindergarten year) and refit models without study control children who later attrited from the district. Finally, we will address the methodological challenge posed by different start rules in commonly used early childhood assessments for children of different ages.

Findings / Results:

Description of the main findings with specific details.

As evident from Table 3, we found significant ($p<0.05$), small to moderate and positive effect sizes on all assessments. Effect sizes for numeracy skills as measured by the Applied Problems assessment were largest (0.61), with effect sizes for pre-reading and reading skills (0.55) and vocabulary (0.44) somewhat smaller. Executive functioning effect sizes were in the small range but all positive: 0.25 (inhibitory control), 0.27 (cognitive flexibility), and 0.26 (working memory).‡ We also found significant positive impacts on most outcomes for most subgroups, which is notable given our reduced power to detect effects within subgroups (exceptions were White children and children with special needs). Results were robust across

* Allowing the slope to vary on either side of the cutoff is necessary to estimate unbiased estimates of impact at the cutoff (Imbens & Lemieux, 2008).

† Including demographics as covariates increases the precision of estimates (Imbens & Lemieux).

‡ We do not present effect sizes for all outcomes or all subgroups, as some analyses are currently underway. At SREE in March, we will present impacts for all assessments shown in Table 2 and for all subgroups of interest.

multiple bandwidths and model specifications, and in other standard RD robustness checks, we find no reason to doubt our findings.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

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 has positive impacts on child early numeracy, language, literacy, and executive function skills. Results for executive function are particularly of interest, given that ours is the first study to provide causal evidence of the impact of a large-scale publicly funded prekindergarten program on this important developmental domain.

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. We believe there are several possible reasons for this. First, stronger impacts in our study might be explained by differences in the respective counterfactuals. We can only speculate regarding this possibility, as ours is the only RD study to date with information on the counterfactual experienced by the control group and as we have information regarding only the type, not the quality, of control group care settings.

Second, our sample includes only children whose parents gave active consent for them to participate, while many of the other examined contexts required passive consent. This difference raises external validity issues; while our estimates are internally valid, we can generalize our estimates only to children whose parents consented to participate. Thus, it may be that the our sample and samples in other prekindergarten RD studies are not exactly comparable and that this difference fully or partially explains the larger effects estimated in our context.

Finally, as mentioned previously in this paper, ours is the only RD prekindergarten context in which there was a uniform curriculum in place. Stronger effects thus could at least partially be a function of the chosen math and literacy curricula and the district's implementation/ professional development supports. This explanation is consistent with theory and some empirical research that suggests 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 early skills, in domains such as literacy, numeracy, socio-emotional, or self-regulation skills (Klein & Knitzer, 2006; NAEYC & NAECS/SDE, 2003). The overall evidence base for the efficacy of preschool curricula is mixed, as many randomized trials have found small or null effects of curricula on children's math and literacy skills (Judkins et al, 2008; Preschool Curriculum Evaluation Research Consortium, 2008; WWC, n.d.). However, given promising results to date for the specific curricula used in the district (Clements & Sarama, 2007b; Clements & Sarama, 2008; Clements, Sarama, Spitler, Lange & Wolfe, in press; Pearson, n.d.), stronger effects could be explained at least partially by the use of a uniform curricula.

By March 2011, we will also explore a number of methodological issues that have not yet been addressed in prekindergarten RD studies. Chief among these, we will examine the sensitivity of our results to the definition of the sample, due to differential enrollment and attrition timeframes for the treatment and control groups. We are currently awaiting data from the district for these checks but they will be complete by March 2011. Finally, we also explore the sensitivity of our results to different start rules on the PPVT-III for treatment and control children. Consistent with test guidelines, in our study, 5-year-old children started the PPVT-III one set ahead of 4-year-old children.

Appendices

Not included in page count.

Appendix A. References

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, *5*(3), 25-50.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian project. *Applied Developmental Science*, *6*(1), 42-57.

Clements, D. H., & Sarama, J. (2007a). *SRA Real Math, PreK-Building Blocks*. Columbus, OH: SRA/McGraw-Hill.

Clements, D. H., & Sarama, J. (2007b). Effects of a preschool mathematics curriculum: Summative research on the *Building Blocks* project. *Journal for Research in Mathematics Education, 38(2),* 136-163.

Clements, D. H., Sarama, J. H., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-based Early Math Assessment. *Educational Psychology*, *28*(4), 457-482.

Clements, D. H., Sarama, J. H., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (in press). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*.

Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives, 15*(2), 213-238.

Currie, J., & Thomas, D. (1995). Does Head Start make a difference? *American Economic Review*, *85*, 341-364.

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.

Gilliam, W. S. (2005). *Prekindergarteners left behind: Expulsion rates in prekindergarten states.* Retrieved December 9, 2006, from http://www.fcd-us.org/PDFs/ExpulsionFinalProof.pdf.

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.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the High/Scope Perry Preschool Program. *Journal of Public Economics, 94*(1-2), 114-128.

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.

Imbens, G. W., & Lemieux, T. (2008). Regression Discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615-635.

Judkins, D., St. Pierre, R., Gutmann, B., Goodson, B., von Glatz, A., Hamilton, J., et al. (2008). *A study of classroom literacy interventions and outcomes in Even Start (NCEE 2008-4028)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Karoly, L. A., Kilburn, M. R., & Cannon, J. S. (2005). *Early childhood interventions: Proven results, future promise*. Santa Monica, CA: RAND Corporation.

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.

Murnane, R., & Willet, J. (2010). *Method matters: Improving causal inference in educational research.* Oxford University Press.

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 Forum on Early Childhood Program Evaluation. (2007). *Early childhood education program evaluations: A decision-makers guide.* Retrieved April 24, 2008, from http://www.developingchild.harvard.edu/content/downloads/Decision_Guide.pdf.

National Institute for Early Childhood Research (NIEER). (n.d.). *Debunking the myths: Benefits of preschool.* Retrieved December 9, 2006, from http://nieer.org/resources/factsheets/13.pdf.

Pearson. (n.d.) Summer school evaluation. Retrieved June 17, 2008 from http://www.pearsonlearning.com/microsites/owl/research.cfm.

Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of preschool curriculum programs on school readiness (NCER 2008-2009)*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Washington, DC: U.S. Government Printing Office.

Schickedanz, J., & D. Dickinson. (2005). *Opening the World of Learning*. Iowa City, IA: Pearson Publishing.

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.

Schweinhart, L.J., Barnett, W.S., & Belfield, C.R. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study through age 40.* Ypsilanti, MI: High/Scope Press.

Temple, J. A., & Reynolds, A. J. (2007). Benefits and costs of investments in preschool education: Evidence from the Child–Parent centers and related programs. *Economics of Education Review, 26*(1), 126-144.

Trochim, W. M. K. (1984). *Research design for program evaluation: The Regression-Discontinuity approach*. Beverly Hills: Sage Publications.

What Works Clearinghouse. (n.d.). *Early childhood education.* Retrieved April 21, 2008, from http://ies.ed.gov/ncee/wwc/Topic/Topic.aspx?tid=13.

Wechsler, D. (1986). *Wechsler Intelligence Scale for Children—Revised.* New York:

Psychological Corporation.

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.

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

Yoshikawa, H., Rosman, E. A., & Hsueh, J. (2002). Resolving paradoxical criteria for the expansion and replication of early childhood care and education programs. *Early Childhood Research Quarterly*, *17*(1), 3.

Appendix B. Tables and Figures
Not included in page count.

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: 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 (REMA) (Clements, Sarama & Liu, 2008)	Numeracy	Comparing/ordering, verbal counting/counting strategies, arithmetic, number recognition and subitizing, geometric, measuring and patterning capacities
Forward Digit Span (Gathercole & Pickering, 2000; Wechsler, 1986).	Executive function	Working memory (phonological loop)
Backward Digit Span (Gathercole & Pickering, 2000; Wechsler, 1986).	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

Table 3: Estimated effect sizes for sample children who participated in the preschool program across child outcomes and by subgroup

	PPVT-III	WJ Letter Word	WJ Applied Problems	Peg Tapping	Forward Digit Span	Dimensional Change Card Sort
Full Sample	0.440***	0.554***	0.609***	0.249**	0.257*	0.269**
Student-level demographics						
Black	0.397**	0.637**	0.535***	0.395*	0.009	0.042
White	0.098	-0.047	0.035	-0.049	0.001	0.011
Hispanic	0.478***	0.217***	0.164***	0.117**	0.018~	0.067**
Asian	0.644~	0.127	0.266*	0.010	0.028~	0.141*
Free/reduced lunch	0.487***	0.173***	0.165***	0.113***	0.014~	0.072**
Full price lunch	0.311*	0.081	0.083*	-0.005	0.018*	0.030
Male	0.439**	0.166***	0.144***	0.083*	0.026**	0.046
Female	0.445**	0.130**	0.142***	0.076~	0.006	0.076**
Special needs	0.264	0.006	0.079	-0.013	0.009	0.094
No special needs	0.468***	0.174***	0.148***	0.100***	0.017*	0.055*
Home language						
English	0.201~	0.108*	0.079**	0.021	0.013~	0.006
Spanish	0.687***	0.190***	0.204***	0.146**	0.021~	0.102**
Other	0.724**	0.200***	0.216***	0.137*	0.019	0.131**

Note: Effect sizes are from models with linear specification of the distance from the age cutoff and with robust SEs with correction for clustering at the classroom-level. Subgroup models also include an interaction between the dichotomous preschool variable and distance from age cutoff. Full sample models include student-level covariates. Effect sizes calculated using the SD of the control group. ***p<0.001; **p<0.01; *p<0.05; ~p<0.10

Table 4: Comparison of effect sizes across RD prekindergarten studies

	PPVT-III	Letter Word Identification	Applied Problems
Present study	0.44***	0.63***	0.61***
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.
Note: Effect sizes calculated using the SD of the control group.
Citations: Tulsa (Gormley, Gayer, Phillips, & Dawson, 2005); MI, NJ, SC, WV, OK (Wong et al., 2007); NM (Hustedt, Barnett, Jung & Goetze, 2009).

Figure 1: Scatterplot of Peabody Picture Vocabulary Test-III scores, with super-imposed fitted linear regression and lowess regression lines