Effects of High-Quality Preschool on Early Literacy and Mathematics Skills—A Regression-Discontinuity Approach

Mariel Kyger, Ph.D. & Melissa Barnhart, Ph.D.

LAUP®

Effects of High-Quality Preschool on Early Literacy and Mathematics Skills—A Regression-Discontinuity Approach

Prepared by Mariel Kyger, Ph.D. & Melissa Barnhart, Ph.D.
Research and Evaluation Department
Los Angeles Universal Preschool

January 2017

Table of Contents

Introduction

A large and growing body of research shows that children who attend high-quality preschool demonstrate a range of positive effects, including improvements in literacy, math, and social-emotional wellbeing (e.g., Slaby, Loucks, & Stelwagon, 2005; Barnett, Lamy, & Jung, 2005; Camilli, Vargas, Ryan, & Barnett, 2010; Yoshikawa et al., 2013). These benefits can last through high school and even into adulthood (e.g., Campbell et al., 2012; Schweinhart et al., 2005; Heckman, 2000). For children from low-income or disadvantaged families, these benefits are even stronger (Palfrey et al., 2005). Ultimately, affordable high-quality preschool narrows the educational gap between high- and low-income children, and in the long term, helps to build a skilled and motivated adult population (Heckman, 2008).

The well-demonstrated effects of early education have led policymakers and child advocates to support programs that provide all children with the opportunity to attend preschool. In 2004, Los Angeles Universal Preschool (LAUP) was created as a result of a grant from First 5 Los Angeles to establish access to preschool for children in Los Angeles County. Studies of students and alumni since LAUP's inception have consistently demonstrated its benefits for preschool children. LAUP students make significant progress in language and math from fall to spring, relative to peers nationally, and meet established targets for cognitive development; they also meet socio-emotional development targets, exceeding the national averages for performance on measures of attention, activity level, and sociability (Xue, Atkins-Burnett, Caronongan, & Moiduddin, 2010; Moiduddin, Xue, & Atkins-Burnett, 2011; Xue, Atkins-Burnett, & Moiduddin, 2012; Atkins-Burnett, Xue, Moiduddin, Aikens, & Cannon, 2013). Longitudinal efforts following LAUP students into second and third grade showed that these students performed as well as or better than their peers on grade-level standardized tests (CSTs) in English and mathematics; low-income children seemed to especially benefit from LAUP participation, performing as well as their higher-income peers on the CSTs (this was not the case for low-income students who did not attend LAUP) (Barnhart & Kyger, 2015; Aikens, Moiduddin, Xue, Chen, & Cannon, 2015).

These studies have revealed that LAUP students consistently meet targets for development in preschool, and that they continue to succeed in elementary school – but these findings cannot determine how students would have performed without the benefit of high-quality preschool. Although we can surmise that participation in preschool is likely to have improved students' outcomes, we cannot state by how much their outcomes have improved. In order to compare the effects of LAUP to the effects of other, similar programs, or to the null effect of no preschool, it is necessary to quantify the impact of LAUP on the population it serves.

One way to statistically quantify the impact of a program on its population is to calculate its effect size. To date, LAUP has been lacking a solid estimate of its effect size, due to the difficulty of identifying and testing a control group. There were three primary barriers to using a traditional two-group, control-versus-treatment design for this study. First, a naturally occurring local population of children who are not enrolled in preschool is hard to locate, since these children are not grouped together or listed on a roster as preschool students are; second, a naturally occurring group of children whose parents do not enroll them in preschool is likely to systematically differ from a group of children whose parents do enroll them, creating a selection bias which could jeopardize results; third, it would be impractical and unethical to eliminate this selection bias by randomly assigning children to a control group in which they would not receive early education, despite their parents' wishes.

The current study makes use of a regression discontinuity design, which allows LAUP to accurately calculate the effect size of one year of high-quality preschool, while avoiding the pitfalls associated with control groups. Despite the usefulness of regression discontinuity as a method of analysis, studies using this method are less common in social science. Before proceeding to describe the current study, we will provide a brief description of the method and its advantages.

One of the issues that consistently arises in research on social interventions is the need for an equivalent control control group that does not receive the intervention. By comparing the outcomes of this control group with the outcomes of the intervention group, or "treatment" group, researchers can draw conclusions about the efficacy of the intervention. Importantly, these types of conclusions can then be used to make causal statements (for example, "Intervention X led to a 5% increase in participants' vocabulary"). Without a control group, change may be observed in a treatment group when the period before the intervention is compared to the period after the intervention, but this does not show that the intervention caused the change; it is possible that this change or growth would have happened in the population regardless of the intervention. Therefore, experiments using randomly assigned control groups are favored because they allow researchers to make statements about causality. Random control trials may also trigger ethical concerns, however. In the case of preschool, a true random control trial would require researchers to randomly assign half the children in the study to receive preschool, while the other half would be assigned to receive no preschool. Many people would consider it unethical to forbid parents to enroll their children in preschool, especially given the existing research showing that preschool is frequently beneficial and may have long-term academic consequences for children who participate.

Regression discontinuity avoids these ethical issues by eliminating the need to treat one group while excluding another; in fact, it is "the only explicitly recognized quasi-experimental approach identified by the Institute of Education Statistics to meet the prerequisites of a causal relationship" (Smith, 2014; IES, 2013). The basis of the regression discontinuity, or RD, design is that membership in programs is often determined based on a strict cutoff which is more or less arbitrary. This cutoff can be conceptualized as a fixed point on a horizontal x-axis representing birthdate, income, or another continuous demographic factor; assignment to the program or treatment is determined by whether the participant falls to the left or the right of this fixed point. Although the x-axis (assignment) variable may itself be associated with the y-axis (outcome) variable, the relationship between x and y is assumed to be continuous (Figure 1a). If a treatment or intervention changes this existing relationship between x and y, a discontinuity or "jump" will appear in the regression line at the given cutoff point for treatment, giving this method the name "regression discontinuity" (Figure 1b). Assuming that proper statistical procedures have been followed to model the regression and measure the assignment variable, a discontinuity at the cutoff point can be interpreted as unbiased evidence of the treatment's effect (Imbens & Lemieux, 2008; Schochet et al., 2010). Because the existence of a cutoff point and its consistent enforcement across participants are crucial prerequisites for the RD analysis, this method is a good fit for situations in which groups are assigned using firm and transparent rules.

Figure 1a. Neither Group Receives Treatment

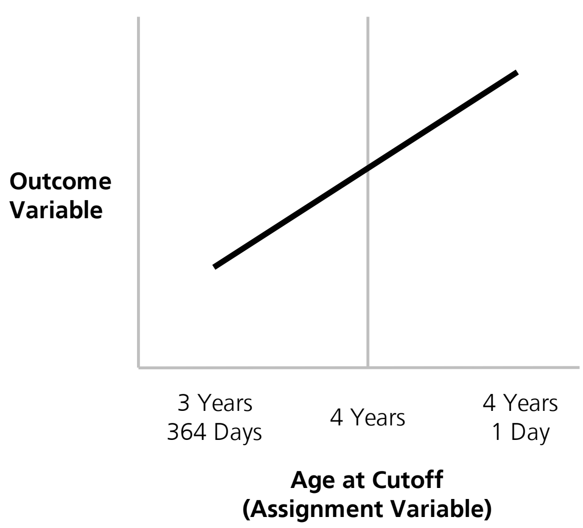
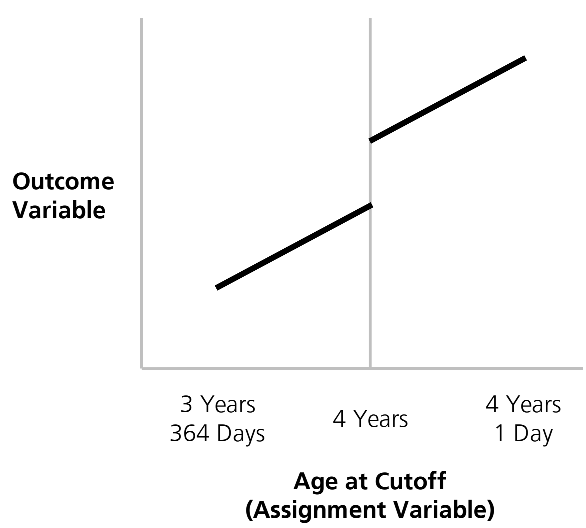


Figure 1b. Group Meeting Cutoff Receives Treatment



To provide a concrete example, preschools often maintain a cutoff birthdate for enrollment each year. This means that in a preschool program with a requirement that enrollees for the current year must turn four years old on or before December 2nd, children who had birthdays on December 1st and 2nd would be eligible for preschool in the current year, while children born on December 3rd and 4th would miss the cutoff for the current year, but would be eligible for preschool in the following year. A small, consistent amount of development takes place on each day of a child's life, and when this development is plotted on a chart where the x-axis represents age and the y-axis represents ability, the expected result is a smooth regression line with no jumps or discontinuities between one day and the next day. In other words, typical development occurs smoothly and continuously, and no noticeable difference in development is expected between a child born on December 2nd and a child born on December 3rd. Using the principles of regression discontinuity, then, researchers can exploit the firm cutoff date to look for effects of treatment. If a comparison of the outcomes for children born right before the cutoff (those who "barely received" treatment) to the outcomes for children born right after the cutoff (those who "barely missed" treatment) reveals a sharp discontinuity centered on the cutoff date of December 2nd, it is safe to conclude that this represents the effect of one group having received preschool. Importantly for the RD method, the children who "barely missed" treatment (children with birthdays after December 2nd) will go on to attend preschool in the coming year, making this method more ethically acceptable than a random control trial in which one group is assigned not to receive treatment at all.

The use of a regression discontinuity design not only avoids the issues of random control trials, but also avoids selection bias, another problem which may arise in studies of interventions. If a group who has chosen not to receive the intervention (e.g., children whose parents have not enrolled them in preschool) is compared to a group who has chosen to receive the intervention (e.g., children whose parents did enroll them in preschool), both groups have self-selected into their categories, and there may be underlying differences between the two groups. In the case of preschool, these underlying differences could be related to cultural beliefs, level of disposable income, parents' education history, or other factors. When the two groups are compared, these differences may act as confounding variables. Regression discontinuity designs eliminate selection bias because the two groups are pulled from the same population grouped around the cutoff point. Any differences aside from variation in the "x" or predictor variable (i.e., age) should be random, and as such, these differences will not act as confounds or threaten the interpretation of results (Barnett, Lamy, & Jung, 2005). The only systematic variation between the two groups should be due to the effects of treatment – in this case, differences in ability due to having received or not having received a year of preschool.

Regression discontinuity has been used as a method of analysis in studies across a number of different fields, including economics (Imbens & Lemieux, 2008), education (Smith, 2014), program evaluation (Jacob, Zhu, Somers, & Bloom, 2012), and political elections (Lee, 2008). Researchers studying the effects of preschool have also begun to implement regression discontinuity, because it is unbiased and convenient, and provides a "control group" without requiring random assignment. For example, Barnett, Lamy, and Jung (2005) used a regression discontinuity design to sample children from five preschool programs (in Michigan, New Jersey, Oklahoma, South Carolina, and West Virginia), and found significant impacts of preschool on children's vocabulary, math, and print awareness skills (see also Wong, Cook, Barnett, & Jung, 2008). A separate study of the Oklahoma universal preschool program, using a regression discontinuity design based on a birthdate cutoff, also demonstrated a significant effect of preschool on children's letter-word identification, spelling, and applied problems skills (Gormley, Gayer, Phillips, & Dawson, 2005). Thus, regression discontinuity has been established as a valuable method of analysis in the field of early education research, as well as in the broader arena of social science research.

The current study was designed to answer two research questions: 1) Do children who have completed one year of LAUP preschool show an advantage in math and language performance, as compared to children of the same age who are beginning their year of preschool?, and 2) What is the size of this effect; that is, how much learning occurs as the result of one year of LAUP preschool?

Methods

Participants in the study were 138 preschool students and 76 kindergarten students, sampled from 9 LAUP preschools and 9 public elementary schools in the Los Angeles area. Data collection for each school was conducted as shortly as possible after the beginning of the 2015-2016 school year, in an attempt to capture only the effects of one year of preschool (and to avoid potential effects related to kindergarten participation). All data collection took place during September of 2015.

To have been admitted to an LAUP preschool, the participants in the kindergarten cohort of this study must have been born between October 2nd, 2009, and December 2nd, 2010; participants in the preschool cohort must have been born between October 2nd, 2010, and December 2nd, 2011. At the time of testing, preschool students were between the ages of 3 years, 9 months, 1 day and 4 years, 9 months, 22 days; 75 were male and 63 were female. Kindergarten students were between the ages of 4 years, 10 months, 9 days and 5 years, 9 months, 20 days; 35 were male and 41 were female. The kindergarten students had all attended an LAUP preschool during the prior school year, and their parents had consented to their participation in the study before the current school year began (while the students were still in preschool). For current preschoolers, consent was obtained from parents at the beginning of the school day, during morning drop-off. No incentives for participation were provided to either parents or students, and parents were assured that their children could choose to stop participating at any time. Assessment of students was conducted one-on-one, and took place either in a corner of the classroom away from the rest of the class, or occasionally, in a separate classroom provided for testing by school staff. Assessments were scheduled to avoid naptimes and meal times.

Classroom Selection
Preschool classrooms were selected for participation from the 2014-15 and 2015-16 LAUP cohorts. Consent forms were distributed to classrooms from the 2014-15 cohort at the end of the school year, but students did not participate until approximately 2-3 months later, once they were in kindergarten. Because it was important to collect all data as soon as possible after the beginning of the school year, sampled preschool classrooms were drawn from the subset of LAUP classrooms that were co-located on the same site as an elementary school, in an attempt to reduce the number of individual sites required for the study. Selecting classrooms from the same schools also helped to ensure that participants from the 2014-15 and 2015-16 cohorts were as comparable as possible. These co-located sites were sampled from within two large Los Angeles County school districts; these districts serve approximately 40% of all LAUP graduates. In total, 21 classrooms were selected for participation: 12 from the 2014-15 cohort, and 9 from the 2015-16 cohort.

In order to obtain permission for the study, LAUP personnel visited each site before or after each preschool session and explained the study to parents. Initially, teachers distributed consent forms to parents, but this approach did not reach all parents and resulted in a consent rate below the goal rate of 70%. Therefore, the recruitment strategy was modified, and three of the original classrooms were dropped. The final sample consisted of 9 classrooms from both the 2014-15 cohort, i.e., the kindergarten cohort, and the 2015-16 cohort, i.e., the preschool cohort. Consent and completion rates for these classrooms are displayed in Table 1. As shown, consent was obtained for over 85% of all enrolled students. The completion rate was considerably lower, particularly among the kindergarten cohort. This lower completion rate was due to student mobility between preschool and kindergarten, and the availability of numerous options for students in addition to their default or "home" public school. For one of the districts, the group of students who had attended three LAUP preschools went on to enroll in a total of 30 different schools for kindergarten; researchers were unable to visit all of these schools within the acceptable data collection period of the first several weeks of the school year.

Table 1. Consent and Completion Rates by Classroom for the LAUP Preschool and Kindergarten Cohorts

Site and Classroom	Enrolled	Consented	Completed	% Consented of Enrolled	% Completed of Consented	% Completed of Enrolled
Preschool Cohort						
1C	20	15	14	75.0%	93.3%	70.0%
2C	24	24	19	100.0%	79.2%	79.2%
2D	24	23	19	95.8%	82.6%	79.2%
3C	20	18	15	90.0%	83.3%	75.0%
3D	23	21	16	91.3%	76.2%	69.6%
4C	22	14	10	63.6%	71.4%	45.5%
4D	22	12	11	54.5%	91.7%	50.0%
5C	22	15	15	68.2%	100.0%	68.2%
5D	22	19	19	86.4%	100.0%	86.4%
Cohort Total	**199**	**161**	**138**	**80.9%**	**85.7%**	**69.4%**
Kindergarten Cohort						
1A	20	16	6	80.0%	37.5%	30.0%
1B	19	13	7	68.4%	53.8%	36.8%
2A	24	22	14	91.7%	63.6%	58.3%
2B	24	19	10	79.2%	52.6%	41.7%
3A	24	17	9	70.8%	52.9%	37.5%
4A	22	18	11	81.8%	61.1%	50.0%
4B	22	19	10	86.4%	52.6%	45.5%
5A	22	12	5	54.5%	41.7%	22.7%
5B	22	11	4	50.0%	36.4%	18.2%
Cohort Total	**199**	**147**	**76**	**73.9%**	**51.7%**	**38.2%**
Grand Total	**398**	**341**	**214**	**85.7%**	**62.8%**	**53.8%**

Participants

A total of 214 students participated in the study (76 from the kindergarten cohort, 138 from the preschool cohort). The participant characteristics are displayed in Table 2. Students in the kindergarten cohort were more likely to be female, and were also more likely to be Latino. Educational levels for both mothers and fathers were higher in the preschool sample, and the proportion of families who earned over $75,000 a year was also higher. Therefore, the subsequent analyses control for gender, ethnicity, parental levels of education, and family income.

Table 2. Demographic Characteristics as a Percentage of LAUP Group

Characteristic	Kindergarten Sample (n=76)	Preschool Sample (n=138)	LAUP Total 2015-16 (N=9,100)
Gender			
Male	46.1	54.3	50.5
Female	53.9	45.7	49.5
Ethnicity			
Asian	3.0	8.1	7.8
White	1.5	6.5	7.7
Latino	90.9	72.4	65.7
Black	1.5	3.3	7.2
Other	8.1	12.7	11.6
Father's educational attainment			
Less than high school	40.0	31.5	25.2
High school	26.7	26.9	30.1
Some college/Less than BA	26.6	25.0	28.7
Bachelor’s degree or higher	6.7	16.7	16.0
Mother's educational attainment			
Less than high school	30.2	22.0	20.0
High school	23.8	27.1	26.2
Less than a Bachelor's degree	41.3	27.1	35.3
Bachelor’s degree or higher	4.8	23.7	18.6
Household income			
<$10,000	14.3	15.0	15.5
$10,000-$24,999	34.7	29.0	30.6
$25,000-$49,999	32.6	31.0	30.9
$50,000-$74,999	16.3	13.0	11.1
$75,000 or over	2.0	12.0	11.9
Previous preschool experience			
Yes	21.0	25.6	35.0
No	79.0	74.4	65.0
Child primary language			
English	39.1	57.6	63.1
Spanish	59.4	40.0	32.0
Other	1.5	2.4	4.9

Note: Percentages are based on the number of parents who answered each question on LAUP’s Child Enrollment Form, completed at preschool entry.

Materials

Students’ progress was evaluated in the domains of language and mathematics, because these areas are frequently used as markers of academic achievement in education studies, allowing for direct comparison of the current results to those of other preschool program evaluations.

Demographic variables. The Child Enrollment Form is an internal form completed by all parents of new LAUP students. It includes demographic information such as ethnicity, parental income, parental level of education, and primary home language. The demographic data for children in this study was collected at the time of preschool enrollment, and does not reflect changes that may have occurred during the preschool year for kindergarteners.

Desired Results Developmental Profile – School Readiness (DRDP-SR). The DRDP is a formative assessment tool for young children designed by the California Department of Education (California Department of Education, 2010). Seven domains are assessed by the DRDP: self and social development; language and literacy development; English language development; cognitive development; mathematical development; physical development; and health. For each measure, teachers rate the developmental level that the child has mastered at the time of the assessment, using a 4-point continuum from Exploring, to Developing, to Building, to Integrating. Teachers provide written evidence for each of their ratings. The teacher also indicates whether the child is "emerging" to the next level. In LAUP preschools, teachers complete the DRDP at the beginning and end of the academic year.

The DRDP was developed using item-response theory. The intra-correlations for the domains of the DRDP-SR range from .64 to .83. The scale reliabilities of the scale are acceptable, and range from .83 to .90. For the purposes of the current study, DRDP scores were not analyzed as outcomes, but were used as a control variable in analyses.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002). The Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) were used to assess students' language ability. Students were assessed on three of the DIBELS modules: Letter Naming Fluency, Phoneme Segmentation Fluency, and Initial Sound Fluency.

The three DIBELS modules used in the current study have a one-month alternate-form reliability of .72-.89 in the kindergarten year (Good et al., 2004). Concurrent validity of these modules with the Woodcock-Johnson Broad Reading Cluster is .36-.75, and 12-month predictive validity with the Woodcock-Johnson Total Reading Cluster is .33-.65. The somewhat broad range of values for concurrent and predictive validity may reflect the fact that each module tests a different subset of early literacy skills. The DIBELS Initial Sound Fluency module asks students to recognize and produce the initial sounds in a set of words presented orally by the assessor. The assessor presents four pictures to the student and names them each individually while pointing to the picture (for example, "rooster," "mule," "fly," and "soap"). The student is then asked, for example, "Which picture begins with /r/?" or "What sound does 'soap' begin with?" The DIBELS Phoneme Segmentation Fluency module asks students to orally segment words into their individual phonemes. For example, the assessor says "punch," and the student replies, "/p/u/n/ch/" to receive full credit for that word. Each student has one minute to correctly segment as many words as possible, based on the assessor's oral prompts. The DIBELS Letter Naming Fluency module asks students to name non-sequential, upper- and lowercase letters in rows from left to right, following the assessor's finger as a guide. Each student is asked to produce as many letter names as possible within the time frame of one minute. If a student becomes "stuck" on a given letter, the assessor tells him or her the correct answer and proceeds to the next letter.

Although the DIBELS modules have not been evaluated for reliability or validity in the preschool age range, there were several considerations that led to the use of the DIBELS as an assessment for both age groups in this study. The most important consideration was the need to directly compare the kindergarten group to the preschool group in order to calculate effect size. Using the same assessment for both age groups made this direct comparison possible. Additionally, the DIBELS assesses phoneme segmentation and letter identification skills as well as vocabulary knowledge, unlike many other language assessments which focus only on vocabulary. These skills are important indicators of pre-literacy, and are less affected by background cultural knowledge and previous exposure to specific words. Other considerations included the ease of conducting the DIBELS and the time required to assess each student. The DIBELS does not require specialized equipment other than the testing booklets, and the three modules took approximately five minutes, in total, for each student to complete. This allowed for assessment of all students from the same classroom over the period of one or two days, which reduced disruption to the classroom and ensured that all data collection could be completed shortly after the beginning of the school year. These reasons contributed to the identification of the DIBELS as a useful method of early language assessment for both preschool and kindergarten students, for the purposes of this study.

Individual Growth and Development Indicators of Early Numeracy (IGDIs-EN; Hojnoski & Floyd, 2004). For assessment of students' mathematics ability, the chosen measure was the Individual Growth and Development Indicators of Early Numeracy (IGDIs-EN; Hojnoski & Floyd, 2004). This measure was selected for its ease and rapidity of administration; the IGDIs-EN took approximately five minutes to complete. Administration of both the DIBELS and the IGDIs-EN in the same session required no more than 10 minutes per student, an important consideration when working with small children whose attention spans are limited.

The IGDIs-EN was also selected for its established reliability and validity. The IGDIs-EN has an overall test-retest reliability of .71-.88, and its concurrent validity is .60-.75 (Hojnoski & Floyd, 2004). The assessment is comprised of four modules: Oral Counting, Number Naming, Quantity Comparison, and One-To-One Correspondence Counting. In Oral Counting, students are asked by the assessor to count as high as they can, starting at the number 1. The assessor records how many numbers the student produces before either skipping a number or failing to produce the next consecutive number. In Number Naming, students are asked to name a series of non-consecutive numbers, one at a time, each printed on a separate display page. The student receives credit for every number he or she correctly names within the allotted time. In Quantity Comparison, students are shown a display page with two large square outlines, each filled with a different number of smaller circles. The assessor asks the student to identify which side has more, without counting the circles. This module tests the student's understanding of the concepts of "less" and "more," and also tests the student's ability to subitize (that is, to estimate quantities without explicitly counting individual items). In One-To-One Correspondence Counting, students are shown a display page with 20 small circles, evenly distributed, and are asked to count all the circles. Counting proceeds from the number 1 until the student either does not know the next number, skips a number, counts the same circle twice, or finishes counting all the circles.

Procedure

Parental consent was obtained for all students participating in the study. Students in the kindergarten cohort were consented at the end of their preschool year, between May and June of 2015. Each parent was asked for their permission to allow their child to be tested in 2 to 3 months, once the school year started. Students in the preschool cohort were typically consented the same day that they were tested. Trained interviewers and preschool teachers verbally explained the study to parents as they came to drop off their students, and they were asked for their written consent.

To find the kindergarten students, researchers provided the roster of consented students to the two school districts within the first two weeks of school, and administrative staff helped to identify the assigned schools of enrolled students. Due to the methodological considerations described above, researchers tested only those students who were enrolled in the same school at which they had attended preschool.

For the preschool cohort, assessors visited the participating classroom and found an appropriate area in which to conduct the IGDIs-EN and DIBELS assessments. For the kindergarten cohort, researchers first asked administrative staff in the school office to identify the assigned classrooms of the consented students. Trained assessors then visited each classroom and conducted the IGDIs-EN and DIBELS assessments, again in a space separate from classroom activities. Assessments for both the kindergarten and preschool cohorts lasted about 10 minutes, and all testing was completed by the end of September 2016.

Scoring of the IGDIs-EN assessments was conducted according to the instructions of the tool. Because the DIBELS assessments were normed and validated for kindergarteners, but not for preschoolers, adjusted scores based on age could not be calculated. Thus, raw scores on the DIBELS were calculated for all participants, and these raw scores were used in analyses.

Analytic Methods

Variables. Due to demographic differences between the two cohorts, the following demographic variables were used as control variables in the analysis: (1) gender is female (0/1); (2) ethnicity is Latino or African American (0/1); (3) household income is $25,000 or less, which is roughly the poverty line for a family of four (0/1); (4) father is not a high school graduate (0/1); (5) mother is not a high school graduate (0/1); and (6) English is not the primary home language (0/1). In addition, the results of the subscales for the DRDP were used to control for children's incoming level of school readiness. Missing values for the student demographic variables were imputed under the assumption that the data were missing at random, and the variables had a multivariate normal distribution. The regression coefficients and p-values were obtained via bootstrapping.

Explicit model. The design for this study was sharp, in that there were no exceptions to the age cut-off. All students in the kindergarten cohort were born on or before 12/2/2010, and all students in the preschool cohort were born on or before 12/2/2011. The explicit model is given below. For each student *i*:

$$Y_i = \beta_0 + \beta_1 * T_i + \beta_2 * (Z_i - Z_c) + \beta_3 * T_i * (Z_i - Z_c) + \mathrm{B}_x * \mathrm{X_i} + e_i$$

In this equation, the outcome Y_i represents the score on the reading or math sub-test for each student *i*. A separate model was created for each of the reading and math sub-scores. T_i represents the dummy variable of whether student *i* was in the kindergarten cohort, that is, whether the student had received the treatment of a year of preschool. The quantity $Z_i - Z_c$ represents the number of days from the student's birthday to the birthday cut-off for eligibility to attend preschool in 2014-15. This quantity ranged from -426 days for the youngest children in the preschool cohort to 301 days for the oldest children in the kindergarten cohort, a span that was just shy of two years. The term $\mathrm{X_i}$ is a matrix which includes a term for each of the student demographic variables. Overall, this equation contains an intercept, an effect for treatment, an effect for age, an interaction term for treatment and age, and effects for the student demographic characteristics included in the model. Due to the relatively small sample size, no tests were conducted for higher-order interaction terms, according to Lee and Lemieux's recommendation (2010).

Results

Graphical Displays

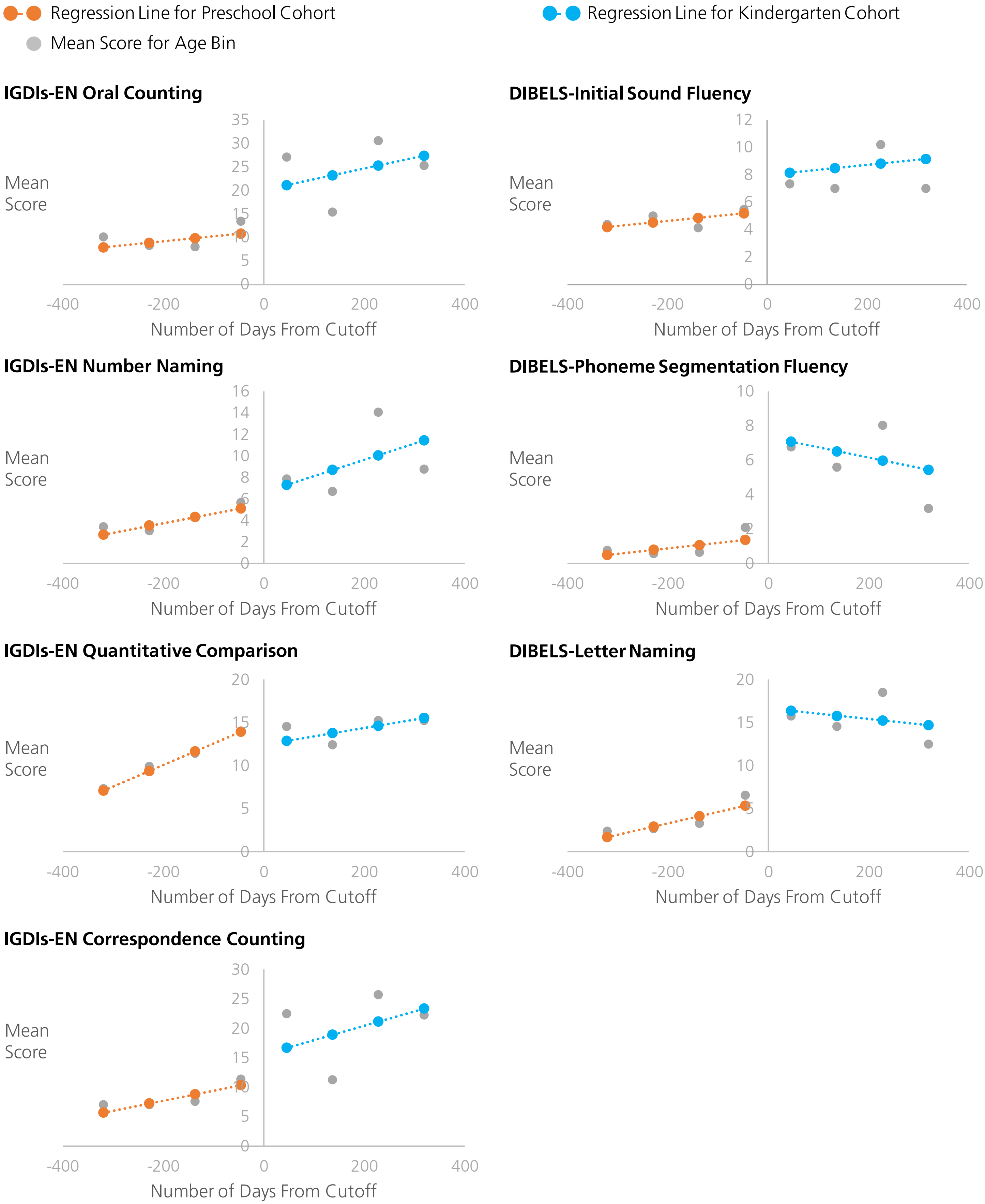
First, researchers created graphical displays of the seven outcome variables versus age, to identify whether there were visually apparent differences in performance between the two groups. In order to determine the appropriate number of bins, the two F-tests outlined in Jacob, Zhu, Somers, & Bloom (2012) were performed. The two tests gave slightly different results: the first F-test indicated that four bins, with two bins on either side of the age cut-off, were needed to capture the relationship with age for three of the seven outcome variables. The second F-test indicated that eight bins were needed for five of the seven outcome variables. In order to facilitate comparisons of the graphs, eight bins, the maximum number of bins indicated, were utilized for all seven outcome variables. Sample sizes for each of these age bins are displayed in Table 3. Sample sizes for the preschool cohort were larger than those for the kindergarten cohort, due to the higher difficulty of locating the kindergarten population.

Table 3. Sample Sizes by Age Bins for the LAUP Preschool and Kindergarten Cohorts

Bin #	Number of Days from Age Cutoff	Students
Preschool Cohort		
1	Less than -273	32
2	-273 to -183	42
3	-182 to -92	38
4	-91 to -1	26
Cohort Total		**138**
		Kindergarten Cohort
5	0 to 91	13
6	92 to 182	20
7	183 to 273	24
8	Greater than 273	19
Cohort Total		**76**

The regression line for each cohort was plotted, using the treatment effect and slope from a simple regression of the outcome variable on three other variables: (1) the treatment group (a dummy variable); (2) the number of days from cutoff; and (3) the interaction of these two effects. These plots are displayed in Figure 2.

Figure 2. Mean Scores By Number of Days From Preschool Age Cutoff



There were discontinuities at the age cutoff for some, but not all, of the outcome variables. As shown in the plots, there were large discontinuities at the age cutoff for two of the language outcome variables: DIBELS Letter Naming Fluency and DIBELS Phoneme Segmentation Fluency. There were smaller discontinuities at the age cut-off for the DIBELS Initial Sound Fluency and the IGDIs-EN Oral Counting modules. No clear discontinuities were displayed in the graphs for the remaining math modules.

Regression Results

To quantify the treatment effects, researchers first utilized the procedure recommended by Jacob et al. (2012) to estimate the functional form for each outcome variable. Testing began with a linear model, and then additional models were tested in the following order: a linear interaction model, a quadratic model, a quadratic interaction model, a cubic model, and a cubic interaction model. The quadratic model included the squared value of age, and the cubic model included both the squared and cubed value of age. The interaction models included additional interaction terms for all of the age variables and the treatment variable, to determine whether the slopes varied for the two groups.

Results varied across the seven outcomes. For the three DIBELS measures and the IGDIs-EN Quantitative Comparisons measure, the F-test revealed that the simplest model, i.e., the linear model, was adequate to estimate the outcome variable. Two of the math measures, the IGDIs-EN Correspondence Counting and IGDIs-EN Oral Counting modules, required additional terms. Specifically, the Oral Counting module required a quadratic interaction model, and the Correspondence Counting module required a cubic interaction model. For the final math outcome (the IGDIs-EN Number Naming module), none of the models were adequate to describe the relationship between outcome and age. Upon inspection of the graph, this result is not surprising; the graph shows virtually no discontinuity at the age cut-off. For this reason, we did not estimate a model for the IGDIs-EN Number Naming module. After the appropriate functional form for each outcome had been determined, the final set of regression analyses was conducted. Researchers ran the appropriate model for each outcome variable after adding terms for the six student demographic variables.

As shown in Table 4, the effect of LAUP participation was significant for two of the language measures: the DIBELS Letter Naming Fluency module and the DIBELS Phoneme Segmentation Fluency module. The graphs for these two outcome measures showed the largest jumps at the age cut-off. Both measures also had large effect sizes; an effect size of .83 was found for Letter Naming, and an effect size of .76 was found for Phoneme Segmentation. These effects translate into a gain of 9.71 points on the DIBELS Letter Naming Fluency module, and a gain of 6.09 points on the Phoneme Segmentation model, both of which can be directly attributed to the effect of preschool.

Results on the DIBELS Letter Naming Fluency module also demonstrated a significant negative effect of minority and low-income status; children in these groups were able to name fewer letters in the allotted time. The effect of participation in preschool was not significant for the remaining three measures. In summary, students' performance on two of the three language measures demonstrated a significant and positive effect of participation in an LAUP preschool, but this effect was not found for any of the three mathematics outcome measures.

Table 4. Regression Results for ELA and Math Outcomes

Predictor	IGDIs-EN Oral Counting		IGDIs-EN Quant. Comparison		IGDIs-EN Correspond. Counting		DIBELS-Initial Sound Fluency		DIBELS-Phoneme Segment. Fluency		DIBELS-Letter Naming Fluency	
	β	*(SE)*	β	*(SE)*	β	*(SE)*	β	*(SE)*	β	*(SE)*	β	*(SE)*
Intercept	27.87	6.74**	14.09	1.38**	24.04	6.22**	4.71	3.87	2.14	1.19	10.35	2.20**
Participation in LAUP preschool	1.59	11.68	-2.75	1.80	25.99	14.96	2.66	4.58	6.09	2.01**	9.71	3.00**
Time from cutoff (days)	0.11	0.05*	0.02	0.00**	0.16	0.09	0.00	0.01	0.00	0.00	0.01	0.01
Female	-3.80	2.12	-1.27	0.79	-2.80	1.50	-0.14	1.89	-0.95	0.66	0.10	1.23
Minority (African American or Latino)	-5.46	2.85	2.03	1.18	-3.54	2.40	1.01	2.80	-0.28	1.38	-4.88	1.89**
Low household income (<$25,000)	-4.85	2.68	-0.04	1.03	-1.96	2.07	0.42	2.67	-0.08	0.90	-3.57	1.52**
Father not a high school graduate	0.07	3.37	-0.52	1.48	0.71	2.36	0.74	3.24	-0.39	1.08	0.34	2.01
Mother not a high school graduate	0.78	3.81	0.85	1.50	-2.65	2.72	-1.19	3.85	-0.22	1.09	0.26	2.20
Non-English primary home language	-3.01	2.42	-2.88	1.10**	-1.46	1.66	-1.15	2.46	-1.49	1.01	-0.42	1.49
Time from cutoff squared	0.00	0.00*			0.00	0.00						
Time from cutoff cubed					0.00	0.00						
Time from cutoff x Treatment	-0.07	0.14			-0.71	0.28**						
Time from cutoff squared x Treatment	0.00	0.00			0.00	0.00						
Time from cutoff cubed x Treatment					0.00	0.00**						
R^2	0.33		0.25		0.36		0.12		0.21		0.39	
Effect size	0.08		-0.50		1.83		0.43		0.76		0.83	

* *p*<.05; ** *p*<.01

Discussion

The current study used regression discontinuity methods to demonstrate significant effects of LAUP preschool participation on students' early language and literacy knowledge. Previous LAUP students received higher scores than entering LAUP students on the DIBELS modules of Phoneme Segmentation and Letter Naming. Results on both of these modules displayed large effect sizes (0.76 and 0.83, respectively), representing benefits for students who had attended LAUP. Many of LAUP's students are bilingual or are learning English as a second language; these results are especially encouraging because they demonstrate that LAUP helps these students to enter kindergarten with the skills needed to succeed in primarily English-speaking classrooms. Children who start kindergarten with weaker language skills are likely to struggle in elementary school and beyond (Hoff, Laursen, & Bridges, 2012); in contrast, stronger early literacy and language skills are predictive of later success, including a smooth transition to kindergarten, success in learning to read, and advanced literacy skills in elementary school (Wasik, Bond, & Hindman, 2006; Whitehurst & Lonigan, 2001; Snow, Burns, & Griffin, 1998). Thus, these students' exposure to preschool and early literacy support may conceivably benefit them in the long term, as well as in the kindergarten classroom.

No significant effect of LAUP participation appeared on the IGDIs-EN math modules. There are several potential explanations for this finding. One potential explanation is that preschool teachers tend to feel less confident about their teaching skills in math, as compared to other domains (e.g., Clements & Sarama, 2011; Ryan, Whitebook, & Cassidy, 2014). It is possible that a lack of confidence in teaching math led teachers to focus less on math and more on other subjects, such as language and literacy. Research also suggests that teachers' beliefs about how math skills and knowledge are generated (internally, within the child, or externally, from the teacher) can create a persistent challenge to math-focused teaching in early childhood by affecting teachers' confidence in the usefulness of mathematics instruction (Copley & Padron, 1998; Takunyaci & Takunyaci, 2014).

It is also possible that teachers focused more heavily on English proficiency than on mathematics for other reasons; as mentioned above, LAUP serves many students who are bilingual from birth ("true" or "crib" bilinguals), or who have been exposed primarily to Spanish at home, and who must become proficient in English when they begin to attend school. Although many LAUP teachers are also Spanish-English bilingual, it would be reasonable to think that teachers might focus on improving students' English proficiency in order to increase their levels of kindergarten readiness and their ability to communicate with teachers and peers. To discover whether teachers do in fact tend to focus on language-building at the expense of mathematical skills, researchers would need to survey or consistently observe teachers across preschool sites.

A third possible explanation is that LAUP participation did have a positive effect on students' math knowledge, but one which our measures did not reveal. A previous study of LAUP outcomes discovered a link between LAUP participation and higher scores on English and math standardized tests in second grade, and in some cases, in third grade (Barnhart & Kyger, 2015). Because the sample size for the kindergarten cohort was small, chances of detecting any but the largest effects were limited. Any existing effect for mathematics may have failed to appear because of the small sample size, or because the IGDIs-EN were not sufficiently sensitive to detect it; effects could also have been hidden by the testing context and constraints. Although assessors did their best to work with students in a quiet and comfortable atmosphere, time and space constraints meant that testing sometimes occurred in a location with elevated noise levels (i.e., close to a group of active preschoolers), and there was little time available for building rapport between the assessor and the student.

One limitation of the current research is the fact that some participants in the study reported other preschool exposure in addition to their participation in LAUP. Therefore, it is possible that effects of preschool beyond those effects specific to LAUP may have been captured. However, this limitation does not pose a large concern. Approximately the same percentage of each cohort (21% of the kindergarten cohort and 25.6% of the preschool cohort) reported preschool experience in addition to LAUP attendance. Thus, it is not likely that this added exposure caused a difference to appear between the cohorts that would not have otherwise appeared.

Another limitation is the sample of schools from which the participant sample was drawn; all of the preschools included in the current study were co-located on elementary school campuses. None of the participating preschools were family child care centers (FCCs), although the LAUP network includes many FCCs. LAUP's network includes a wide variety of preschool settings, and the use of co-located sites for the purposes of the current study may limit the results' applicability to the experiences of children in other preschool settings. However, measures of preschool quality used by LAUP to rate classrooms (the Early Childhood Environment Rating Scale–Revised (ECERS; Harms, Clifford, & Cryer, 2005) and the Classroom Assessment Scoring System PreK (CLASS PreK; Pianta, La Paro, & Hamre, 2008)) indicated that the levels of quality in the classrooms selected for participation were not significantly different from the levels of quality in classrooms across the rest of the network. This suggests that the results reported here are likely to apply throughout the LAUP network.

Finally, potential limitations caused by overlap in cohort birthdates were considered. The cut-off birthdates for preschool attendance overlapped by two months. Specifically, the eligible birthdates for the kindergarten cohort were between October 2nd, 2009, and December 2nd, 2010; the eligible birthdates for the preschool cohort were between October 2nd, 2010, and December 2nd, 2011. This two-month overlap occurred because, in 2009-10, the eligibility dates were expanded to allow students to participate in a year of preschool before attending transitional kindergarten (which many Los Angeles County school districts began offering to students who were born close to the cut-off for kindergarten). If parents in the current study had chosen to send their children born in the overlapping months to preschool in different years, it could have biased the estimate of the treatment effect. However, all children born in the overlapping months attended preschool in Cohort 1, presumably so that they could attend transitional kindergarten the following year. Because there were no children born between October 2nd and December 2nd of 2010 who attended preschool in Cohort 2, this overlap in eligibility dates did not ultimately affect the study design or results.

Results of this study carry positive implications for children participating in LAUP's high-quality preschool programs. These children have demonstrated significant benefits to their early literacy and language skills, and the statistical characteristics of the regression discontinuity method make it possible to state that participation in LAUP led to these benefits. Future research in this area should attempt to use regression discontinuity methods, when possible, in order to make statements about cause and effect in cases where random control trials are infeasible or undesirable. The detection of effects directly attributable to high-quality preschool programs will provide preschool researchers and advocates with strong evidence for the usefulness of early childhood education, an area of study in which comparison across programs is often difficult. By using statistically robust, easily replicated methods, researchers will begin to disentangle the complicated web of factors influencing cognitive development; in the long term, this research will allow advocates and policy-makers to directly assess and compare benefits across early childhood education programs.

References

Aikens, N., Moiduddin, E., Xue, Y., Chen, J., & Cannon, J. (2015). *A portrait of Los Angeles Universal Preschool children: Outcomes in early elementary school*. Washington, DC: Mathematica Policy Research.

Atkins-Burnett, S., Xue, Y., Moiduddin, E., Aikens, N., & Cannon, J. (2013). *Informing the performance-based contract between First 5 LA and LAUP, volume 1: Child progress in the 2012-2013 program year.* Washington, DC: Mathematica Policy Research.

Barnett, W. S., Lamy, C., & Jung, K. (2005*). The effects of state prekindergarten programs on young children's school readiness in five states.* Rutgers University: The National Institute for Early Education Research.

Barnhart, M., & Kyger, M. (2015). *Longitudinal outcomes of LAUP alumni.* Los Angeles: Los Angeles Universal Preschool. Available at: http://laup.net/documents/resources/research/re_docs_071916/cdsr/laup_cdsr_brief_alumoutcomes_rev20160202.pdf.

Bartik, T. J. (2013). *Effects of the pre-k program of Kalamazoo County Ready 4s on kindergarten entry test scores: Estimates based on data from the fall of 2011 and the fall of 2012* (Upjohn Institute Working Paper 13-198). Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

California Department of Education. (2010). *The Desired Results Developmental Profile – Preschool© 2010*. CDE Press: Sacramento, CA.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*(3), 579-620.

Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., … Ramey, C. T. (2012). Adult outcomes as a function of an early childhood educational program: An Abecedarian Project follow-up. *Developmental Psychology, 48*, 1033-1043.

Clements, D. H., & Sarama, J. (2011). Early childhood mathematics intervention. Science, 333(6045), 968-970.

Copley, J. V., and Padron, Y. (1998). *Preparing teachers of young learners: professional development of early childhood teachers in mathematics and science. Forum on Early Childhood Science, Mathematics, and Technology Education.* Washington, DC: American Association for the Advancement of Science. Available at: http://www.project2061.org/publications/earlychild/online/fostering/copleyp.htm.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills: Sixth edition*. Eugene, OR: Institute for the Development of Educational Achievement. Available at: http://dibels.uoregon.edu.

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

Harms, T., Clifford, R. M., & Cryer, D. (2005). *Early Childhood Environment Rating Scale (Rev. ed.).* New York: Teachers College Press.

Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics, 54,* 3-56.

Heckman, J. J. (2008). The case for investing in disadvantaged young children. In *Big Ideas for Children: Investing in Our Nation's Future*, 49-58. Washington DC: First Focus.

Hoff, E., Laursen, B., & Bridges, K. (2012). Measurement and model building in studying the influence of socioeconomic status on child development. In M. Lewis and L. Mayes (Eds.), *A Developmental Environment Measurement Handbook* (pp. 590-606). Cambridge University Press.

Hojnoski, R., & Floyd, R. (2004*). Individual Growth and Development Indicators of Early Numeracy: First edition*. Saint Paul, MN: Early Learning Labs, Inc.

Institute of Education Sciences. (2013). *Review of individual study protocol*. IES: What Works Clearinghouse. Available at: http://ies.ed.gov/ncee/wwc/Document/262.

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

Jacob, R., Zhu, P., Somers, M. A., & Bloom, H. (2012). *A practical guide to regression discontinuity.* New York, NY: MDRC. Available at: http://www.mdrc.org/sites/default/files/regression_discontinuity_full.pdf.

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics, 142*(2), 655-674.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature, 48*, 281–355.

Moiduddin, E., Xue, Y., & Atkins-Burnett, S. (2011). *Informing the performance-based contract between First 5 LA and LAUP: Child progress in the 2010-2011 program year: Final report*. Washington, DC: Mathematica Policy Research.

Palfrey, J. S., Hauser-Cram, P., Bronson, B. M., Warfield, M. E., Sirin, W., & Chan, E. (2005). The Brookline Early Educational Project: A 25-year follow-up study of a family-centered early health and development intervention. *Pediatrics, 116,* 144–152.

Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™ Manual, Pre-K*. Baltimore, MD: Brookes Publishing.

Ryan, S., Whitebook, M., & Cassidy, D. (2014). *Strengthening the math-related teaching practices of the early care and education workforce: Insights from experts*. Berkeley, CA: Center for the Study of Child Care Employment, University of California, Berkeley. Available at: http://cscce.berkeley.edu/files/2014/Math-Expert-Paper-Report.pdf.

Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J. R., Porter, J., & Smith, J. (2010). *Standards for regression discontinuity designs*. Available at: https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_rd.pdf.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). The High/Scope Perry Preschool Study through age 40: Summary, conclusions, and frequently asked questions. Excerpted from *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40* (Monographs of the HighScope Educational Research Foundation, 14). Ypsilanti, MI: HighScope Press.

Slaby, R., Loucks, S., & Stelwagon, P. (2005). Why is preschool essential in closing the achievement gap? *Educational Leadership and Administration, 17*, 47-57.

Smith, W. C. (2014). Estimating unbiased treatment effects in education using a regression discontinuity design. Practical Assessment, *Research & Evaluation, 19*(9). Available at: http://pareonline.net/getvn.asp?v=19&n=9.

Snow, C., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Takunyaci, M., & Takunyaci, M. (2014). Preschool teachers' mathematics teaching efficacy belief. Procedia – *Social and Behavioral Sciences, 152*, 673-678.

Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology, 98*(1), 63-74.

Whitehurst, G. J., & Lonigan, C. J. (2001). Emergent literacy: Development from prereaders to readers. In S. B. Neuman and D. K. Dickinson (Eds.), *Handbook of Early Literacy Research* (pp. 11-29). New York: Guildford.

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*, 122–154.

Xue, Y., Atkins-Burnett, S., Caronongan, P., & Moiduddin, E. (2010). *Informing the performance-based contract between First 5 LA and LAUP: Assessing child progress*. Washington, DC: Mathematica Policy Research.

Xue, Y., Atkins-Burnett, S., & Moiduddin, E. (2012). *Informing the performance-based contract between First 5 LA and LAUP: Child progress in the 2011-2012 program year*. Washington, DC: Mathematica Policy Research.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M., Espinosa, L., Gormley, W. T., Ludwig, J., Magnuson, K., Phillips, D., & Zaslow, M. (2013). *Investing in our future: The evidence base on preschool*. Society for Research in Child Development: Washington, D.C.