



**IRVINE NETWORK ON
INTERVENTIONS IN DEVELOPMENT**

**Head Start at ages 3 and 4 versus Head Start followed by state pre-k:
Which is more effective?**

Jade Marcus Jenkins, University of California, Irvine
George Farkas, University of California, Irvine
Greg J. Duncan, University of California, Irvine
Margaret Burchinal, University of North Carolina at Chapel Hill
Deborah Lowe Vandell, University of California, Irvine

February 2014

Abstract

As policy-makers contemplate expanding preschool opportunities for low-income children, one possibility is to fund two, rather than one year of Head Start for children at ages 3 and 4. Another option is to offer one year of Head Start followed by one year of pre-k. We ask which of these options is more effective. We use data from the Oklahoma pre-k study to examine these two ‘pathways’ into kindergarten using regression discontinuity to estimate the effects of each age-4 program, and propensity score weighting to address selection. We find that children attending Head Start at age 3 develop stronger pre-reading skills in a high quality pre-kindergarten at age 4 compared with attending Head Start at age 4. Pre-k and Head Start were not differentially linked to improvements in children’s pre-writing skills or pre-math skills. This suggests that some impacts of early learning programs may be related to the sequencing of learning experiences to more academic programming.

Introduction

In light of the evidence that high quality early learning experiences can improve children's school readiness and future academic success (Duncan & Magnuson, 2013; Yoshikawa et al., 2013), a number of recent proposals at the federal and state levels would expand public early childhood education (ECE) programs. These initiatives aim to serve not just *more* children, but to also serve *younger* children, and to address the detrimental effects of poverty during early childhood on children's wellbeing in the short- and long-term (Duncan, Magnuson, Kalil, & Ziol-Guest, 2012). This expansion includes the federal Head Start program, a comprehensive child development program that provides children with preschool education and other services, which children can enter as early as age 3. Indeed, 3-year-olds are also the largest growing group of Head Start participants, increasing from 24 percent in 1980 to 40 percent in 2007, and comprising 63 percent of first-time Head Start children in 2010 (Aikens, Klein, Tarullo, & West, 2013; Tarullo, Aikens, Moiduddin, & West, 2010).

Expanding ECE programs means that children will have more opportunities to participate in programs for multiple years. In fact, over half of all 3-year-old entrants go on to complete two years of Head Start (Aikens, et al., 2013). Others transition from Head Start at age 3 to pre-kindergarten (pre-k) programs at age 4, which are state-created, academically-focused ECE programs. In fact, the latter combination of programs is precisely what President Obama proposed in his 2013 early learning agenda—expand Head Start to serve 3-year-olds, while helping states to increase their educational investments in 4-year-olds.

Unclear in the Head Start literature is whether the program is designed to provide two years' worth of developmental benefits for children. In K-12 education, cross-grade curricula can be designed so that material taught in each grade builds on the skills and knowledge learned previously, and incremental benefits from each year of schooling for learning and labor market outcomes are well established (Card, 1999). However, we know little about whether ECE programs are designed to do the same. Furthermore, unlike primary education where children are separated by grade or state pre-k programs that serve only 4-year-olds, the Head Start model combines 3- and 4-year-olds in most classrooms – 75% by one recent estimate (Hulsey et al., 2011). If children in their second year of Head Start continue to receive more of the same activities rather than increasingly complex, differentiated learning experiences, they may gain relatively little from a second year in the program and may gain more by switching to a more academic pre-k program at age 4.

The objective of this study is to answer one key question: If children participate in Head Start at age 3, is it more beneficial for them to remain in the program at age 4 or to participate in a pre-k program at age-4? We use data from the study of the Oklahoma Pre-kindergarten program (OK pre-k) to compare outcomes for two different preschool 'pathways' to kindergarten (Gormley et al., 2005, 2008, 2010). One of these involves Head Start at both ages 3 and 4. The other involves Head Start at age 3 followed by OK pre-k at age 4. We use a regression discontinuity design with a strict age eligibility cutoff for program participation to estimate the effect of these pathways on children's pre-academic skills at kindergarten. We apply propensity score weighting to the analyses to address selection into pathways and compare their effects on child outcomes.

It is important to note how our study differs from the prior studies using these data. The objective of Gormley and colleagues' work was to estimate treatment effects for OK pre-k and

for Head Start on a range of child outcomes at kindergarten entry for different child development outcomes. For academic outcomes they estimated two separate regression discontinuity specifications—one for pre-k and one for Head Start—calculated treatment effect sizes, and compared effect sizes descriptively (Gormley, 2008; Gormley & Gayer, 2005; Gormley, Gayer, Phillips, & Dawson, 2005). Since their goal was not to compare the effectiveness of attending OK pre-k and Head Start at age 4 amongst age 3 Head Start graduates, they do not need to pool both pre-k and Head Start children into the same RD model and address differential selection into the programs. Accordingly, they compare two separately generated RD effect sizes using only a basic significance test (a difference in z-scores) outlined in Pasternoster and colleagues (Gormley, Phillips, Adelstein, & Shaw, 2010; 1998). Our study is designed to make a rigorous statistical comparison between these two programs in a sample of children who attended Head Start at age 3.

We find that children attending Head Start at age 3 followed by OK pre-k at age 4 have stronger pre-reading outcomes at kindergarten compared with children who attend Head Start at ages 3 *and* 4. This suggests that the impacts of early learning programs may be related to the sequencing of ECE programs to a more academic curriculum at age 4 and the extent to which the Head Start curriculum offers differential learning experiences to 4-year-olds who were, and were not, in the program at age 3.

Background

The effects of different types of early learning programs

Head Start. Head Start is a comprehensive child development program that provides children with preschool education, health screenings and examinations, nutritious meals, and opportunities to develop social-emotional skills. This federal program targets very low-income families, and children who are at risk of entering school unprepared. Many studies have examined the benefits and long-term effects of Head Start, and there are several comprehensive and critical reviews of this literature, primarily using data for 4-year-old program participants (see Gibbs, Ludwig & Miller, 2011 and Ludwig and Phillips, 2008 for reviews).

Because of its use of random assignment, the experimental Head Start Impact Study is the best evidence on the short-term impacts of Head Start on children’s language, literacy and pre-writing skills at ages 3 and 4. The end-of-program-year effect sizes average 0.2 SD, with stronger differential impacts for children from high risk households relative to children from modest or low-risk households (ES for subgroup =0.3 SD)(Puma, Bell, Cook, & Heid, 2010). Even though short-term gains appear to ‘fade-out’, Ludwig and Phillips show that the short-term intent-to-treat effects are large enough for Head Start to pass a cost-benefit test (2008). They calculate larger treatment-on-the-treated estimates for some key outcomes (e.g. letter-word identification effect sizes, where the intent-to-treat impact was 0.24 SD and the corresponding treatment on the treated estimate was 0.35 SD). There is also strong quasi-experimental evidence on the effects of Head Start, which show long-term benefits of Head Start on academic outcomes, with effect sizes of 0.2-0.3 standard deviations (Currie & Thomas, 1995; Deming, 2009; Garces, Thomas, & Currie, 2002). These studies looked at single-year impacts of Head Start only, whereas our study compares a 2-year Head Start experience to a 1-year Head Start-1-year pre-k experience.

Pre-kindergarten. Pre-k programs are locally funded programs that provide a year or two of education prior to kindergarten for children ages 3 or 4. Nationally, 28 percent of all 4-year-

olds were enrolled in state-funded pre-k across 40 states in 2010 compared with 11 percent of 4-year-olds enrolled in Head Start (Barnett, Carolan, Fitzgerald, & Squires, 2011). However, “pre-k” does not have a standardized meaning with respect to children’s ECE experience because states created their pre-k programs independently, and with varying characteristics (Gilliam & Ripple, 2004; Jenkins, 2014; Lombardi, 2003; Pianta & Howes, 2009). Some pre-k programs—such as Oklahoma’s—are recognized as very high quality and offer features such as frequent instructional interactions in subject-matter learning, teachers who are emotionally supportive of children and who are credentialed, and classroom environments that are well-organized, efficient with time management, and include developmentally appropriate learning materials (Burchinal, 1999; Harms, Clifford, & Cryer, 1998; Mashburn et al., 2008; Phillips, Gormley, & Lowenstein, 2009; Pianta, Barnett, Burchinal, & Thornburg, 2009).

A randomized study of the state pre-k program serving socioeconomically disadvantaged children in Tennessee found short-term gains in language, literacy and math outcomes for pre-k participants compared with children who did not participate (Lipse, Farran, Bilbrey, Hofer, & Dong, 2011). The evaluations of Oklahoma and Boston’s pre-k programs use regression discontinuity designs based on a strict age eligibility cutoff and found large short-term improvements in pre-reading, pre-writing, pre-math skills, and executive function (ES range = .99-.36)(Gormley, 2008; Gormley & Gayer, 2005; Gormley, et al., 2005; Weiland & Yoshikawa, 2013). Using a similar regression discontinuity design, studies of pre-k programs in Arkansas (Hustedt, Barnett, & Jung, 2008) and a five-state pre-k comparison find positive effects for pre-reading, literacy, and math skills (ES range = .23-.96)(Wong, Cook, Barnett, & Jung, 2008).

Other studies of the effects of pre-k programs have used propensity score (PS) methods and find positive effects for programs in Chicago (Reynolds, Temple, Ou, Arteaga, & White, 2011; Reynolds, Temple, Robertson, & Mann, 2001), Georgia (Henry, Gordon, & Rickman, 2006) and in national samples (Magnuson, Ruhm, & Waldfogel, 2007), with lasting cognitive gains for the most disadvantaged children. Results from meta-analysis (Camilli, Vargas, Ryan, & Barnett, 2010), and correlational studies (Howes et al., 2008; Huang, Invernizzi, & Drake, 2012) also show that children benefit from state pre-k programs. Our analyses examine whether children benefit more from attending a state pre-k program after attending Head Start at age 3 relative to attending Head Start at ages 3 and 4.

Comparing the effects of two types: Head Start and pre-k. While there is a large body of research on the effectiveness of individual types of ECE programs, relatively few studies have directly compared the effectiveness of Head Start and pre-k. In one study, Henry and colleagues (2006) use propensity score matching to address selection and compare Head Start to state pre-k, finding that state pre-k participants had statistically significant but modestly higher scores at kindergarten entry relative to similar Head Start participants. Gormley and colleagues (2010) calculate separate RD estimates for each age-4 program in Tulsa, OK, and find larger effects for OK pre-k participants. The effects of Head Start and pre-k vary depending on the comparison treatment condition (Ludwig & Phillips, 2008). Zhai, Brooks-Gunn, and Waldfogel use PS to match Head Start children to children in different ECE programs and find that Head Start was associated with improved cognitive and social outcomes when compared with children who received parental care or other non-center-based care (2011). However, when compared to children who attended pre-k programs (across different states) and center-based care, Head Start children had better social but not academic outcomes. In this study, we compare the outcomes of

age 4 Head Start and age 4 pre-k participants at kindergarten entry for a sample of children who attended Head Start at age 3.

Duration and dosage effects of ECE

The influence of program duration on children's outcomes is essential for understanding whether two years of Head Start would be more beneficial for children than one year of Head Start and then one year of pre-k. More than half of the children who enter Head Start at age 3 will stay for an additional year (Tarullo, et al., 2010), yet the research on duration in Head Start, and ECE more generally, is limited. The body of evidence from experimental and non-experimental studies suggests that on balance, more participation in center-based ECE is associated with stronger cognitive outcomes, especially for low-income children (Behrman, Cheng, & Todd, 2004; Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Dearing, McCartney, & Taylor, 2009; Hill, Brooks-Gunn, & Waldfogel, 2003; Loeb, Fuller, Kagan, & Carrol, 2004). However, the marginal effect of attending a first year of preschool is generally greater in magnitude than that of a second year for children's short and long-term outcomes (Arteaga, Humpage, Reynolds, & Temple, 2013; Reynolds, et al., 2011; Tarullo, Xue, & Burchinal, 2013). In addition, some research indicates potentially adverse consequences of long hours of care on social and behavioral outcomes in conjunction with positive academic and achievement effects (Belsky et al., 2007; Datta Gupta & Simonsen, 2010; Loeb, Bridges, Bassok, Fuller, & Rumberger, 2007; Magnuson, et al., 2007; Vandell et al., 2010). And while intensive early learning interventions such as Abecedarian and Perry Preschool provided 2 to 5 years of program services and produced significant effects (Campbell, et al., 2001; Schweinhart, 2005), other preschool programs produce substantial effects in only 1 year (Gormley, et al., 2005).

The Head Start duration research is equivocal, with some indication that two years are more advantageous than one, but not 'twice' as advantageous.¹ A number of studies in this area use PS methods to address possible bias due to selection into dosage. Burchinal and colleagues use the 2006 and 2009 FACES data and find that children who entered Head Start at age 3 and also participated at age 4 had modestly higher vocabulary scores relative to children who participated in Head Start at age 4 only, with the gains from the second year being much smaller than the first (ES of second year=0.10-0.17)(2013). Another PS study uses the 2003 FACES data, finding larger effects of 2-year participation (ES=27-.80)(Wen, Leow, Hahs-Vaughn, Korfmacher, & Marcus, 2012). Other PS (Domitrovich et al., 2013; Skibbe, Connor, Morrison, & Jewkes, 2011) and correlational studies of Head Start (Lee, 2011) also find slightly larger gains for 2 years of over 1.

On the other hand, PS analyses of the Chicago Parent Child ECE program did not show significant additional benefits for 2 years of participation versus 1 year (Reynolds, 1995; Reynolds, et al., 2011). The authors suggest that the program model may have provided redundant instruction for two-year participants. Barnett and Lamy also find no influence of duration in a pre-k program on print awareness and math, with some small effects for vocabulary (2006). Nores and Barnett conduct a meta-analysis of dosage effects across an international sample of ECE programs and find that programs lasting 1 to 3 years had average effect sizes of 0.3 standard deviations, as compared with 0.2 for programs lasting less than 1 year, with a maximum effect size of 0.3 at 3 years or more (2010).

¹ The Head Start Impact Study did not include an experimental analysis of participating in one year versus two because children were able to select into receiving Head Start at age 4 after being randomly assigned to treatment at age 3.

If longer exposure produces better outcomes, then 2 years of Head Start may be money well spent. But the literature does not provide consistent support for the notion that 2 years is better than 1, or that individual ECE programs are designed to provide multiple years of unique, developmentally appropriate, incremental learning. Thus, it may be that children continue to gain skills in a second year of Head Start, but they could gain even more by switching to a more academic age 4 program—state pre-k. Testing this is the goal of our study.

Possible Curricular and Peer Effects

Curricula. The extent to which the Head Start curriculum differentiates children's age 3 and age 4 learning experiences would influence the Head Start dosage effect and the comparative effect of Head Start to OK pre-k at age 4 (Yoshikawa, et al., 2013). A key difference between the two programs is that pre-k classrooms typically serve only 4-year-olds, whereas a majority of Head Start classrooms combine 3- and 4-year-olds. Consequently, age 3 Head Start graduates are very likely staying in the same classroom, with the same teacher, books, and other materials during their second year. If Head Start instruction is also the same during children's second year, Head Start children may not receive increasingly complex, differentiated learning experiences on a regular basis, which are critical for intellectual development (Bronfenbrenner, 1989).

We know relatively little about whether Head Start curricula are hierarchical, where learning activities evolve as children age, because of the variation in curricula and limited support of their efficacy (Clifford & Crawford, 2009). The Head Start program mandates that program curricula focus on the whole child, where learning occurs through participating in activities, whereas domain-specific curricula used in some the more effective pre-kindergarten programs (i.e. Boston) focus on presenting lessons that become increasingly complex and build on the inherent hierarchy of skills within that domain (Weiland & Yoshikawa, 2013). According to FACES data from 2000 to 2009, the most common curriculum used in Head Start classrooms is the Creative Curriculum (46% of teachers report using), followed by High/Scope (19%), a number of other widely available curricula (e.g. Scholastic, High Reach, Montessori)(13%), and other less commonly used curricula (e.g. Galileo, Houghton Mifflin, Links to Literacy)(20%). A study of pre-k programs also found that Creative Curriculum and High/Scope are the most frequently used curricula in pre-k programs (Clifford et al., 2005), which were also the modal responses from a sample of OK pre-k teachers (Phillips, et al., 2009). These curricula follow a whole-child approach to children's learning, and are not domain specific. While there is evidence of High/Scope's effectiveness on children's early learning (Belfield, Nores, Barnett, & Schweinhart, 2006; Schweinhart, 2005), there is little support for the Creative Curriculum (U.S. Department of Education, 2013). Curricula effectiveness also depends on the extent that teachers implement them with fidelity, which is largely unmeasured in ECE studies.

This variation in curricula, their limited efficacy, and the unknown degree to which learning activities change as children age highlight the ambiguity of the second-year Head Start experience. As explained by Reynolds in his study of dosage in the Chicago Parent Child program, "an additional year that simply repeats learning activities of the first year would not be expected to make much difference" (1995; p, 23). In contrast, the OK pre-k program may be an opportunity for age 3 Head Start participants to receive a novel age 4-*specific* learning experience and avoid any redundancy in the Head Start curriculum. While we lack information on the classroom characteristics in our Tulsa Head Start and pre-k data, we simply wish to highlight the important role that curricular differences may play in accounting for differential effects of the two pathways.

Peer effects. Classroom composition and peer effects may also play a role in creating differential effects of the two pathways. Children’s skill development could be substantially affected by the skills of their classroom peers in ECE because teacher-directed activities are often kept to a minimum. Henry and Rickman study peer effects in preschool children and find an effect size of 0.36 standard deviations for cognitive skills (2007). Cascio and Schanzenbach find positive peer effects of older, more mature students in kindergarten for other children in the classroom (2012). Studies also suggest positive peer effects on math and reading achievement for school-aged children (Elder & Lubotsky, 2009; Hanushek, Kain, Markman, & Rivkin, 2003; Zimmer & Toma, 2000).

In our study, it is possible that the classroom compositions in both age-4 preschool environments could have different and opposing peer effects on the age-4 learning experiences of Head Start graduates. If second-year Head Start children have more advanced skills than their new classmates that they acquired during the first year of Head Start, this could benefit the other first-time Head Start children through peer learning, increasing the rate at which age 4-only children can catch-up to their second-year peers (Winsler et al., 2002). Simultaneously, younger age 3 peers in mixed-age Head Start classrooms could slow additional progress for second-year students either from behavioral disruption, from an absence of positive academic peer effects, or related to the curriculum issue, the level of content teachers present based on the group’s overall ability (Betts & Shkolnik, 2000; Lavy, Paserman, & Schlosser, 2012; Moller, Forbes-Jones, & Hightower, 2008). In this situation, age 3 Head Start graduates are the *benefactors* of peer effects and are not likely the beneficiaries. Both mechanisms would reduce the additive effect of children’s second year in Head Start.

On the other hand, the age 3 Head Start graduates attending OK pre-k at age 4 may be the *beneficiaries* of positive peer effects because the OK pre-k program is universal, and classroom compositions are more mixed in terms of children’s socioeconomic backgrounds. Because poor and low-income children have substantially lower school-readiness skills than their higher income peers, peer effects in mixed socioeconomic classrooms are particularly valuable for the most disadvantaged children (Barnett & Belfield, 2006; Hart & Risley, 1995; Henry, et al., 2006; Rouse, Brooks-Gunn, & McLanahan, 2005; Zimmer & Toma, 2000). These two opposing peer effects—second-year Head Start children as benefactors and OK pre-k-Head Start graduates as beneficiaries—would attenuate the overall effect of Head Start. With our dataset, we are not able to estimate the effects of peers in an empirical model, but we can describe some of the conditions likely determining peer effects. We present this descriptive information in the results section below.

On balance, we believe that prior findings and the likely direction of curricular and peer effects suggest that age-3 Head Start graduates will have stronger pre-academic skills if they participate in the OK pre-k program at age 4 relative to children who stay in Head Start for a second year at age 4. It is important to know whether children would be better off in one age 4 preschool experience over another especially since this particular pathway – Head Start at age 3 followed by State Pre-K at age 4 – is the plan promoted by the Obama administration, and appears to be the direction in which national policy is evolving.

Methods

Research design and analysis

Our research question is as follows: If children participate in Head Start at age 3, do they have better pre-academic skills at kindergarten entry if they stay in Head Start for an additional year at age 4 or if they participate in a high-quality state pre-k program at age 4? Answering this question involved two analytic processes: estimating treatment effects for each pathway and addressing selection into age 4 treatment. We estimated treatment effects using a regression discontinuity model. We applied propensity score weighting to the regression discontinuity model make the groups as comparable as possible.

We used a dummy variable approach to deal with missing data.² All analyses were conducted using Stata 12 (StataCorp., 2011). We briefly describe the intuition of these procedures here and present the methodological details in Appendix 1, and supplemental figures and calculations in Appendix 2.

Data

Participants. The evaluation focused on the children enrolled in the Tulsa pre-K programs in 2006-7, using the data from the Tulsa Preschool Study 2006-07 Public Use Data File. This evaluation of the Oklahoma's state-funded universal pre-k program administered in Tulsa Public Schools, and the Tulsa County Head Start program administered by local Community Action Project sites was conducted by a team from Georgetown University who made the data public (Gormley, 2011). The data come from four sources: direct cognitive assessments of children at the beginning of the school year; parent surveys collected at their child's cognitive assessment; social-emotional assessments conducted by each child's teacher; and administrative data from Tulsa Public Schools and Head Start.

Our research questions focused on the children eligible for free or reduced-price lunch that attended Head Start at age 3 (n=540). Among these children, the analysis data set includes students who were entering the OK pre-k, age-4 Head Start, or OK public school kindergarten in the 2006-07 school year. The two preschool pathways we created and their sample sizes are: 1) participants in OK pre-k at age 4 who participated in Head Start at age 3 (211 total; 88 kindergarten entrants and 123 pre-k entrants), and 2) participants in Head Start at age 4 *and* age 3 (329 total; 119 kindergarten entrants, 210 HS entrants). Child and family characteristics for the OK pre-k and HS groups are presented in columns 1 and 2 of Table 1.

Measures. Child academic assessments occurred in August 2006 and included three academic subtests from the Woodcock-Johnson Achievement Tests (Woodcock & Johnson, 1989). The Letter-Word Identification subtest measures pre-reading skills, whereby children are asked to identify letters and pronounce words. The Spelling subtest requires children to trace letters, write letters in upper and lowercase, and to spell words, measuring pre-writing and spelling skills. The Applied Problems test has children perform simple calculations to solve math problems, which measures their early math reasoning. The reliability coefficient for the 3- to 5-year-old age group ranges from .97 to .99 (Woodcock, McGrew, & Mather, 2001). The same subtests of a comparable Spanish test, the Woodcock-Muñoz Bateria, were given to Hispanic students capable of being tested in Spanish. The assessment values are in raw scores

² To our knowledge, the literature is unclear as to how one should handle missing data in a propensity score analysis. Because multiple imputation models the relationship between the outcomes, exposure and covariates simultaneously, this violates the analytic feature of PS whereby the relationship between the covariates and exposure and covariates and outcome are separated. We attempted to implement Full Information Maximum Likelihood methods, but our pathway sample sizes were not adequate to achieve convergence in these models.

and are not nationally normed. Further detail regarding the sample, procedures, measurement, and assessments are available in Gormley et al. (2005).

[Table 1 about here]

1. Estimating treatment effects: Regression discontinuity design

Our study implements the regression discontinuity (RD) design, a rigorous method for estimating unbiased treatment effects under certain conditions. The RD technique exploits the fact that the OK preschool programs enforced a strict age cutoff for participation based on child's birth date, so that children who turned 4 before the cutoff (September 1 of 2005-06 school year) were eligible to participate in the OK pre-k and age-4 Head Start programs, and children who turned 4 after the cutoff were not. The primary condition for conducting an RD analysis is the use of a quantitative assignment variable with a designated cutoff score that determines exposure to treatment (Imbens & Lemieux, 2008; Shadish, Cook, & Campbell, 2002). In our analysis, child age—measured as distance between their birthdate and the cutoff birthdate in days—is the assignment variable for the RD specification.

Using RD to compare the mean outcomes of children who made the cutoff to those who did not provides 'pseudo' pre- and post-test measures for OK pre-k and Head Start because all children in the study—those who made the cutoff and those who missed the cutoff—were assessed at the same time (August 2006). The RD sample includes two cohorts of children; cohort 1 children are 5-6 years old and are entering kindergarten at the outcome assessment date, and cohort 2 children are 4-5 years old and are entering a preschool program at the outcome assessment date. Therefore at the time of testing, cohort 1 was treated by Head Start or OK pre-k during the 2005-06 school year (i.e. born *before* the cutoff), and cohort 2 had not yet participated in either age-4 program (i.e. born *after* the cutoff). Because the children in cohort 2 had *selected into* either age 4 Head Start or OK pre-k at the testing date, the members of cohort 2 entering pre-k or Head Start in 2006-07 can serve as the pre-test comparison group for cohort 1 children who completed the same program. The intuition here is that our RD estimates within-pathway changes in children's outcomes by comparing the mean outcomes of the two cohorts.

The important feature to this between-cohort, within-pathway comparison using RD is that the pathway treatment effects are identified by comparing the average outcomes for children with birthdays just above and below the cutoff date. This difference in mean outcomes at the cutoff point is captured by a dichotomous indicator variable (i.e. making the treatment cutoff=1) shown in the model below. Therefore, a key assumption of this RD model is that the children on either side of the cutoff differ only in age, and are otherwise comparable (with respect to potential outcomes). Because age—measured as distance from the birthdate cutoff—is included in the analysis model, this removes any age-related contributions to differences in outcomes so that, conditional on other covariates, all that remains is the effect of the age-4 program. That is, regression adjustment removes the effects of age for those in each cohort, so their outcome is adjusted to what it would have been as follows: The older students within cohort 1 (who have completed the preschool program) have their scores adjusted back to what they would have been at their 5th birthday, and since these adjusted scores include the effect of the preschool program, they can be used as post-test measures. The younger students within cohort 2 have their scores adjusted forward to what they are expected to be at their 5th birthday, and since these adjusted scores do not include the effect of the preschool program they are just entering, they can be used as pretest measures.

Model specification. We estimated the RD models using Ordinary Least Squares regression with PS weights (described below) to generate treatment effects of each pathway and to test for pathway differential effects on outcomes at kindergarten entry. Comparing two different exposures with RD involved a nuanced RD specification. We include an interaction term between the treatment indicator (birthdate occurs before the cutoff=1) and an indicator for one of the two pathways (cutoff*age 3 and age 4 Head Start) to test for differential effects between the two exposures. The model also controls for parent’s education, child race, sex, reduced-price lunch status, exposure to other non-parental care (yes=1), and missing data indicators, presented below:

$$Y_{ij} = \alpha + \beta_1 Cutoff_i + \beta_2 (Cutoff_i * HS_i) + \beta_3 HS_i + \beta_4 (Age_i - Q) + \beta_5 (Age_i - Q)^2 + Z_i + e$$

Where Y is one of three pre-academic skill outcome measures (*j*), indexed by child (*i*). *Cutoff* is a dichotomous indicator of whether the child’s birthdate occurs before the eligibility cutoff for OK pre-k or Head Start and equals 1 if the child was treated. OK pre-k is the reference group and only the indicator for Head Start (at age 4) is included (β_3). Therefore, the differential treatment effect for age 4 Head Start is indicated by β_2 , which is an interaction between the cutoff indicator (treated) and the Head Start indicator. A linear combination of $\beta_1 + \beta_2$ represents the (local) average treatment effect for Head Start, whereas β_1 represents the treatment effect for OK pre-k, the reference group.³ β_4 is the effect of the quantitative assignment variable, age, which is measured in days and is centered at the birthdate cutoff *Q* (September 1). β_5 is a quadratic version of age and Z is a vector of control variables.

Because the treatment effect comes from this discontinuity in outcomes at the birthdate cutoff for treatment, it is critical to check for an appropriate ‘bandwidth’, which involves an analysis of restricted samples of observations clustered around the cutoff within a range of the assignment variable (e.g. +/- 90 days, 180 days) (Schochet et al., 2010; Van Der Klaauw, 2008). The intuition behind this procedure is that the units close to the cutoff are likely to differ only in their exposure to the treatment, but those further from the cutoff might differ in additional ways. In our RD models we used a modest bandwidth restriction of 270 days (3/4 year) to ensure exchangeability in observations on either side of the treatment cutoff while also preserving power and precision in our relatively small treatment groups (Schochet, et al., 2010). See Appendix 1 for further detail.

2. Addressing selection: Propensity score methodology

The information in Table 1 shows that children’s characteristics differ between pathways. We use PS weighting methods to adjust for these observable differences. Propensity score weights induce comparability between Head Start and OK pre-k children, allowing us make a statistical comparison of the two treatment effects in the same RD model.

The PS is the predicted probability of a given exposure conditioned on a rich set of covariates. This score is then applied in analyses to reduce confounding between the exposure of interest and outcomes from observable factors (Heckman, Ichimura, & Todd, 1998; Rosenbaum & Rubin, 1983). A critical feature of PS methods is the assumption that there is no confounding

³ We were unable to estimate the RD models using instrumental variables estimation because this specification includes the treatment variable twice, creating two endogenous variables relative to our one instrument, age. In this situation, the RD specification would not be identified in an instrumentals variables estimation (Angrist & Pischke, 2008).

due to unobserved variables. Because this assumption is untestable, we cannot be confident that our results represent causal estimates of the impact and differential effects of the preschool pathways. They are merely the best possible correlational estimates of our effects of interest. This is especially true in our study since we do not know why age 3 Head Start participants would choose pre-k over Head Start at age 4.

One can implement PS methods in a number of ways, and are often implemented with matching (Caliendo & Kopeinig, 2008). In this study, we use a method based on Inverse Probability of Treatment Weights (IPTW) a form of the Thompson-Horvitz survey sampling weight (Foster, 2011). Weights are calculated as the inverse of the predicted probability of receiving the exposure a person actually received (i.e. Treated group weights = $1/PS$ Comparison group weights = $1/1-PS$). Because the PS is a summary of the observed covariates used in the specification to predict an individual's treatment status, this technique then inflates the importance of cases that are underrepresented in a given exposure to create comparable groups (i.e. by having a smaller value in the denominator of their IPTW). In this way, IPTWs create a pseudo-population in which selection bias from observed factors is removed and observations (children) are exchangeable between exposures (pathways). Our analyses use these IPTWs in the RD models described above.

After implementing PS methods, it is critical to assess comparability in covariate means across exposure groups, referred to as balance checking. Our balance checking involved regressing each covariate on the exposure using the propensity score weights. The results are reported in columns 3 and 4 of Table 1, which shows the IPT-weighted group means for both pathways compared with the unweighted group means. The two groups become very similar with respect to observed covariates after weighting, and there are no remaining significant relationships between Head Start or pre-k and the covariates. See Appendix 1 for further detail.

Results

Pathway effects

Full model results are presented in Table 2, and the main findings are illustrated in Figure 1. The coefficients in Table 2 represent changes in raw scores after participation in an age 4 preschool program, estimated from PS-weighted RD models. Our key coefficients of interest are in the grey box at the top of the table that includes the calculated effect sizes shown below the standard error of the estimate.

[Table 2 about here]

We find that both age 4 programs improved children's pre-reading and pre-writing skills and neither program significantly improved children's pre-math scores. The primary difference in effects between the two preschool pathways was in children's letter-word recognition, with a significant difference in effects size of .46 indicating that the OK pre-k group show treatment effects twice as large as the age 4 Head Start group. Both preschool pathways improved children's pre-spelling scores equally well.

The effect sizes for the WJ-Letter-word subtest at kindergarten entry are 0.92 for age 3 Head Start graduates who attended OK pre-k at age 4, and 0.46 for children who stayed in Head Start at age 4. The effect sizes for the WJ-Spelling subtest are 0.68 for children who attended OK pre-k at age 4, and 0.53 for those who attended Head Start at age 4. The difference in effect sizes for spelling is not significant.

[Figure 1 about here]

Descriptive comparison of classroom peers

In Appendix 2.1 we present the average assessment scores for all age 3 Head Start graduates measured at the beginning of their age 4 programs in 2006-07 (using the younger cohort in the sample) as a proxy for a post-age 3 Head Start assessment.⁴ Comparing age 3 Head Start graduates between age-4 programs shows that the two groups have insignificantly different letter-word and applied problems scores ($p=0.45, 0.50$), although second-year Head Start entrants have higher spelling scores (Standardized mean difference (SMD)=0.27, $p=0.00$). Comparing the ability and characteristics of the *peers* of age 3 Head Start graduates in their age-4 programs indicate—at least descriptively—potentially different peer effects for both the OK pre-k entrants and age 4 Head Start entrants (further detail in Appendix).

Discussion

Motivated by the increasing number of children entering Head Start at age 3 and the expansion of public preschool programs for children at age 4, the objective of this study was to answer the question: If children participate in Head Start at age 3, is it more beneficial for them to stay in the Head Start program at age 4 or to participate in a high quality state pre-kindergarten program at age 4? There was limited prior research on whether the Head Start program is effective at providing a second year of instruction and care that builds upon what children learned at age 3, or whether Head Start is best thought of as a 1-year program that children can enter at age 3 or age 4, with minimal incremental benefits from the second year of the program. To examine this issue, we compared two sets of age 3 and age 4 preschool exposure sequences that we called pathways into kindergarten: 1) age 3 Head Start and age 4 OK pre-k, and 2) age 3 Head Start and age 4 Head Start. We employed a unique combination of strong quasi-experimental methods, using regression discontinuity to estimate the effects of both age-4 programs, and propensity score weighting to address selection into these two ‘pathways’ into kindergarten.

Our findings suggest that after children attend Head Start at age 3, they will have stronger pre-reading skills if they attend a high quality state pre-k at age 4 rather than a second year of Head Start. We find that among Tulsa children attending Head Start at age 3, those attending the OK pre-k program at age 4 have stronger letter-word recognition at kindergarten entry when compared with attending Head Start again at age 4. The comparative effect of the two age 4 programs was striking, with a differential that was two times the effect size of the Head Start program itself on letter and word identification skills (ES=0.98, 0.46, OK pre-k and Head Start, respectively). OK pre-k and Head Start were both equally as effective at improving children’s pre-spelling skills (ES= 0.68, 0.53; no significant difference) and neither program significantly improved children’s pre-math skills. Note that the effect sizes for pre-k are similar to those found in other studies, particularly those of Gormley and colleagues on the OK pre-k program (0.2-0.9), and that the effect sizes for Head Start are larger than those found in the Head Start Impact Study experiments (0.2-0.3).

These findings are consistent with other studies of dosage in early education that show little to no marginal effect of a second year of an ECE program on child outcomes in the short and long term (Arteaga, et al., 2013; Reynolds, 1995; Reynolds, et al., 2011; Schweinhart &

⁴ We assume that the selection mechanisms into OK pre-k or Head Start at age four do not vary between cohorts.

Weikart, 1981; Tarullo, et al., 2013). We identified several possible explanations for why age 3 Head Start graduates in OK pre-k at age 4 outperform children who remain in Head Start at age 4, or why we did not identify a strong dosage effect of Head Start in our study. It may be that the Head Start curriculum does not adequately differentiate children's age 3 and age 4 learning experiences. Because a majority of Head Start classrooms combine 3- and 4-year-olds, it is likely that age 3 Head Start graduates remain in the same classroom, with the same teacher and other materials during their second year. This may not provide Head Start children with the increasingly complex, differentiated learning experiences that are essential to children's intellectual development (Bronfenbrenner, 1989). Because the OK pre-k advantage was concentrated to pre-reading outcomes, the instructional repetition may be specifically related to Head Start children's exposure to new books or literacy activities in their second year. In contrast, the OK pre-k program may have provided novel age 4-specific learning experiences and materials, avoiding curriculum redundancy in a more academically focused environment.

Furthermore, if programs are not designed to build on gains, they may show lower incremental impacts when measured towards the end of the program relative to children's outcomes measured mid-program. Some ECE programs appear to have larger effects when assessments occur during implementation with effect sizes decreasing at the end of treatment, which occurred in the Abecedarian Project and Project CARE (Ramey, Bryant, Sparling, & Wasik, 1985; Ramey et al., 2000). Children were assessed at the end of their age 4 program in the OK preschool study, but for our research question, we ideally would have measured outcomes at the end of the age 3 program year. In this vein, the outcome measurement for the 1-year OK pre-k exposure would be timed to catch the maximal benefit of pre-k, but we would not know the contribution of age 3 Head Start without a post-age 3 Head Start measure. Measuring this 'value-added' from age 3 Head Start in both pathways could be particularly important if Head Start is not actually designed to be a 2-year program, and we may have underestimated the effects of Head Start for second-year students.

It is also possible that peer effects in each of the age 4 preschool environments could have different and opposing effects on the age 4 learning experiences of age 3 Head Start graduates. If second-year Head Start children have more advanced skills than their new classmates that they acquired during the first year of HS, this could benefit the other first-time age 4 Head Start children through peer learning. In this situation, age 3 Head Start graduates are *benefactors* of peer effects, while the age 3 Head Start graduates who attend OK pre-k at age 4 may become *beneficiaries* of positive peer effects because the OK pre-k program brings in children from higher income families with stronger school readiness skills. These two opposing effects could have reduced the identified impact of Head Start. While we could not empirically estimate the effects of peers, we conducted some descriptive analyses of the ability and characteristics of the peers of age 3 Head Start graduates. This suggested that the opposing peer effects hypotheses are plausible for both age-4 programs.

Another way to test for dosage effects of a second year in Head Start would be to compare the outcomes of children who attended two years of Head Start to those that only attended one year. We tested this using the OK study data, comparing children who attended Head Start at age 4 to those who attended at ages 3 and 4. We employed the same methodology as above, combining regression discontinuity and propensity score weighting. The results are shown in Appendix 2. Both the 1 and 2-year participants showed significant improvements in applied problems (ES= .39, .46, respectively), but the improvements made by second-year Head Start

children were not significantly larger than those of first-year children. There were no other significant effects of either pathway.⁵ These additional results support our main findings of limited marginal benefits of staying in Head Start for a second year at age 4.

The most substantial limitation of our study is that propensity score methods assume there is no unobserved confounding, which is not testable, and therefore our estimates do not represent causal effects. The other study limitations are as follows: 1) the OK pre-k program may not be representative of most state pre-k programs because of its very high quality standards; 2) children living in Tulsa, OK are not representative of the broader population of children in the U.S.; 3) we cannot identify benefits from age 3 treatments beyond what is summarized into the scores of the age 4 assessment of the younger cohort in our sample; 4) our sample sizes may not provide sufficient power to detect effects, 5) we cannot know why some parents took their children out of Head Start in the second year, and; 6) Head Start and pre-k have different goals and may often serve different populations. While Head Start supports child cognitive, emotional, and physical development for very low income children, pre-k programs focus solely on academic activities to prepare children for school entry, and also may be offered to any child who is age-eligible regardless of income or need.

Acknowledgements

We are grateful to the Institute of Education Sciences (IES) for supporting this work through grant [redacted] awarded to [redacted]. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Research reported in this publication was also supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number [redacted]. The content is solely the responsibility of the authors and does not necessarily represent the official views of IES, the U.S. Department of Education, or the National Institutes of Health. We would also like to thank Ana Auger, Marianne Bitler, Robert Crosnoe, Thad Domina, Dale Farran, and Sean Reardon for helpful comments on prior drafts.

⁵ The differences in propensity score weights constructed for the 1 vs. 2 years of Head Start analyses and the age 4 Head Start vs. OK pre-k analyses (for age 3 Head Start graduates) account for the differences in pathway effect sizes and significance across comparisons.

References

- Aikens, N., Klein, A. K., Tarullo, L. B., & West, J. (2013). Getting ready for Kindergarten: Children's progress during Head Start. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricists companion*. Princeton, NJ: Princeton University Press.
- Arteaga, I., Humpage, S., Reynolds, A. J., & Temple, J. A. (2013). One Year of Preschool or Two-Is It Important for Adult Outcomes? Results from the Chicago Longitudinal Study of the Child-Parent Centers. *Economics of Education Review*(0). doi: <http://dx.doi.org/10.1016/j.econedurev.2013.07.009>
- Barnett, W. S., & Belfield, C. R. (2006). Early Childhood Development and Social Mobility. *The Future of Children*, 16(2, Opportunity in America), 73-98.
- Barnett, W. S., Carolan, M. E., Fitzgerald, J., & Squires, J. H. (2011). The State of Preschool 2011. New Brunswick, N.J.: National Institute for Early Education Research.
- Barnett, W. S., & Lamy, C. E. (2006). Estimated impacts of number of years of preschool attendance on vocabulary, literacy and math skills at kindergarten entry. New Brunswick, NJ: National Institute for Early Education Research.
- Behrman, J. R., Cheng, Y., & Todd, P. E. (2004). Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach. *Review of Economics and Statistics*, 86(1), 108-132. doi: 10.1162/003465304323023714
- Belfield, C. R., Nores, M., Barnett, W. S., & Schweinhart, L. J. (2006). The High/Scope Perry Preschool Program. *Journal of Human Resources*, XLI(1), 162-190. doi: 10.3368/jhr.XLI.1.162
- Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., & The, N. E. C. C. R. N. (2007). Are There Long-Term Effects of Early Child Care? *Child Development*, 78(2), 681-701. doi: 10.1111/j.1467-8624.2007.01021.x
- Betts, J. R., & Shkolnik, J. L. (2000). Key difficulties in identifying the effects of ability grouping on student achievement. *Economics of Education Review*, 19(1), 21-26. doi: [http://dx.doi.org/10.1016/S0272-7757\(99\)00022-9](http://dx.doi.org/10.1016/S0272-7757(99)00022-9)
- Bronfenbrenner, U. (1989). Ecological systems theory. In R. Vasta (Ed.), *Annals of child development* (Vol. 6, pp. 187-249). Greenwich, CT: JAI Press.
- Burchinal, M. R. (1999). Child care experiences and developmental outcomes. *The Annals of the American Academy of Political and Social Science*, 563(1), 73-97.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31-72. doi: 10.1111/j.1467-6419.2007.00527.x
- 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., Miller-Johnson, S., Burchinal, M. R., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental psychology*, 37(2), 231-242.

- Card, D. (1999). The causal effect of education on earnings. In C. A. Orley & C. David (Eds.), *Handbook of Labor Economics* (Vol. Volume 3, Part A, pp. 1801-1863): Elsevier.
- Cascio, E., & Schanzenbach, D. W. (2012). First in the Class? Age and the Education Production Function. *National Bureau of Economic Research Working Paper Series, No. 13663*.
- Clifford, R. M., Barbarin, O., Chang, F., Early, D., Bryant, D., Howes, C., . . . Pianta, R. (2005). What is Pre-Kindergarten? Characteristics of Public Pre-Kindergarten Programs. *Applied Developmental Science, 9*(3), 126-143.
- Clifford, R. M., & Crawford, G. M. (Eds.). (2009). *Beginning School: U.S. policies in international perspective*. New York: Teachers College Press.
- Currie, J., & Thomas, D. (1995). Does Head Start Make a Difference? *The American Economic Review, 85*(3), 341-364. doi: 10.2307/2118178
- Datta Gupta, N., & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics, 94*(1,Äi2), 30-43. doi: <http://dx.doi.org/10.1016/j.jpubeco.2009.10.001>
- Dearing, E., McCartney, K., & Taylor, B. A. (2009). Does higher quality early child care promote low-income children's math and reading achievement in middle childhood? *Child Development, 80*, 1329-1349.
- Deming, D. (2009). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics, 1*(3), 111-134. doi: 10.2307/25760174
- Domitrovich, C. E., Morgan, N. R., Moore, J. E., Cooper, B. R., Shah, H. K., Jacobson, L., & Greenberg, M. T. (2013). One versus two years: Does length of exposure to an enhanced preschool program impact the academic functioning of disadvantaged children in kindergarten? *Early Childhood Research Quarterly, 28*(4), 704-713. doi: <http://dx.doi.org/10.1016/j.ecresq.2013.04.004>
- Duncan, G. J., & Magnuson, K. (2013). Investing in Preschool Programs. *The Journal of Economic Perspectives, 27*(2), 109-132. doi: 10.1257/jep.27.2.109
- Duncan, G. J., Magnuson, K., Kalil, A., & Ziol-Guest, K. (2012). The Importance of Early Childhood Poverty. *Social Indicators Research, 108*(1), 87-98. doi: 10.1007/s11205-011-9867-9
- Elder, T. E., & Lubotsky, D. H. (2009). Kindergarten Entrance Age and Children's Achievement. *Journal of Human Resources, 44*(3), 641-683.
- Foster, E. M. (2011). Deployment and the Citizen Soldier: Need and Resilience. *Medical Care, 49*(3), 301-312 310.1097/MLR.1090b1013e318202abfc.
- Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. *The American Economic Review, 92*, 999-1012.
- Gilliam, W. S., & Ripple, C. H. (2004). What can be learned from state-funded preschool initiatives? A data-based approach to the Head Start devolution debate. In E. F. Zigler & S. J. Styfco (Eds.), *The Head Start debates* (pp. 477-497). Baltimore, MD: Brookes Publishing.
- Gormley, W. T. (2008). The Effects of Oklahoma's Pre-K Program on Hispanic Children*. *Social Science Quarterly, 89*(4), 916-936. doi: 10.1111/j.1540-6237.2008.00591.x
- Gormley, W. T. (2011). Tulsa 2006-07 Public Use Data Set. In G. University (Ed.). Washington, D.C. .

- Gormley, W. T., & Gayer, T. (2005). Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program. *Journal of Human Resources*, *XL*(3), 533-558. doi: 10.3368/jhr.XL.3.533
- 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. doi: 10.1037/0012-1649.41.6.872 pmid:
- Gormley, W. T., Phillips, D., Adelstein, S., & Shaw, C. (2010). Head Start's Comparative Advantage: Myth or Reality? *Policy Studies Journal*, *38*(3), 397-418. doi: 10.1111/j.1541-0072.2010.00367.x
- Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*(5, Empirical Analysis of Social Interactions), 527-544.
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale*. New York: Teachers College Press.
- Hart, B., & Risley, T. (1995). *Meaningful Differences in the Everyday Experiences of Young American Children*. Baltimore, M.D.: Brookes.
- Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, *65*(2), 261-294.
- Henry, G. T., Gordon, C. S., & Rickman, D. K. (2006). Early Education Policy Alternatives: Comparing Quality and Outcomes of Head Start and State Prekindergarten. *Educational Evaluation and Policy Analysis*, *28*(1), 77-99.
- Henry, G. T., & Rickman, D. K. (2007). Do peers influence children's skill development in preschool? *Economics of Education Review*, *78*, 100-112.
- Hill, J. L., Brooks-Gunn, J., & Waldfogel, J. (2003). Sustained effects of high participation in an early intervention for low-birth-weight premature infants. *Developmental Psychology*, *39*(4), 730-744. doi: <http://dx.doi.org/10.1037/0012-1649.39.4.730>
- Howes, C., Burchinal, M. R., Pianta, R., Bryant, D., Early, D., Clifford, R. M., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-Kindergarten programs. *Early Childhood Research Quarterly*, *23*(1), 27-50.
- Huang, F. L., Invernizzi, M. A., & Drake, E. A. (2012). The differential effects of preschool: Evidence from Virginia. *Early Childhood Research Quarterly*, *27*(1), 33-45. doi: <http://dx.doi.org/10.1016/j.ecresq.2011.03.006>
- Hustedt, J. T., Barnett, W. S., & Jung, K. (2008). Longitudinal effects of the Arkansas Better Chance program: Findings from kindergarten and first grade. New Brunswick, NJ: Rutgers, The State University of New Jersey, National Institute for Early Education Research.
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615-635. doi: 10.1016/j.jeconom.2007.05.001
- Jenkins, J. M. (2014). Early childhood development as economic development: Considerations for state-level policy innovation and experimentation. *Economic Development Quarterly*, *28*(1).
- Lavy, V., Paserman, M. D., & Schlosser, A. (2012). Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom*. *The Economic Journal*, *122*(559), 208-237. doi: 10.1111/j.1468-0297.2011.02463.x

- Lee, K. (2011). Impacts of the duration of Head Start enrollment on children's academic outcomes: moderation effects of family risk factors and earlier outcomes. *Journal of Community Psychology, 39*(6), 698-716. doi: 10.1002/jcop.20462
- Lipsey, M. W., Farran, D. C., Billbrey, C., Hofer, K. G., & Dong, N. (2011). Initial results of the evaluation of the Tennessee Voluntary Pre-K Program. Nashville, TN: Peabody Research Institute, Vanderbilt University.
- Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review, 26*(1), 52-66. doi: <http://dx.doi.org/10.1016/j.econedurev.2005.11.005>
- Loeb, S., Fuller, B., Kagan, S. L., & Carrol, B. (2004). Child Care in Poor Communities: Early Learning Effects of Type, Quality, and Stability. *Child development, 75*(1), 47-65.
- Lombardi, J. (2003). *Time to care: Redesigning child care to promote education, support families, and build communities*. Philadelphia, PA: Temple University Press.
- Ludwig, J., & Phillips, D. A. (2008). Long-Term Effects of Head Start on Low-Income Children. *Annals of the New York Academy of Sciences, 1136*(Reducing the Impact of Poverty on Health and Human Development: Scientific Approaches), 257-268.
- Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *The Economics of Early Childhood Education, 26*(1), 33-51.
- Mashburn, A. J., Pianta, R., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., . . . Howes, C. (2008). Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills. *Child Development, 79*(3), 732-749. doi: 10.1111/j.1467-8624.2008.01154.x
- Moller, A. C., Forbes-Jones, E., & Hightower, A. D. (2008). Classroom age composition and developmental change in 70 urban preschool classrooms. *Journal of Educational Psychology, 100*(4), 741-753. doi: <http://dx.doi.org/10.1037/a0013099>
- Nores, M., & Barnett, W. S. (2010). Benefits of early childhood interventions across the world: (Under) Investing in the very young. *Economics of Education Review, 29*(2), 271-282. doi: <http://dx.doi.org/10.1016/j.econedurev.2009.09.001>
- Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology, 36*(4), 859-866. doi: 10.1111/j.1745-9125.1998.tb01268.x
- Phillips, D. A., Gormley, W. T., & Lowenstein, A. E. (2009). Inside the pre-kindergarten door: Classroom climate and instructional time allocation in Tulsa's pre-K programs. *Early Childhood Research Quarterly, 24*(3), 213-228. doi: <http://dx.doi.org/10.1016/j.ecresq.2009.05.002>
- Pianta, R., Barnett, W. S., Burchinal, M. R., & Thornburg, K. R. (2009). The Effects of Preschool Education: What We Know, How Public Policy Is or Is Not Aligned With the Evidence Base, and What We Need to Know. *Psychological Science in the Public Interest, 10*(2), 49-88. doi: 10.1177/1529100610381908
- Pianta, R., & Howes, C. (Eds.). (2009). *The promise of pre-k*. Baltimore, MD: Paul H. Brookes Publishing Co.
- Puma, M., Bell, S., Cook, R., & Heid, C. (2010). Head Start Impact Study. Final Report. Washington, DC.: U.S. Department of Health and Human Services, Administration for Children and Families.

- Ramey, C. T., Bryant, D. M., Sparling, J. J., & Wasik, B. H. (1985). Project CARE: A Comparison of Two Early Intervention Strategies to Prevent Retarded Development. *Topics in Early Childhood Special Education, 5*(2), 12-25. doi: 10.1177/027112148500500203
- Ramey, C. T., Campbell, F. A., Burchinal, M., Skinner, M. L., Gardner, D. M., & Ramey, S. L. (2000). Persistent Effects of Early Childhood Education on High-Risk Children and Their Mothers. *Applied Developmental Science, 4*(1), 2-14. doi: 10.1207/s1532480xads0401_1
- Reynolds, A. J. (1995). One year of preschool intervention or two: Does it matter? *Early Childhood Research Quarterly, 10*, 1-33.
- Reynolds, A. J., Temple, J. A., Ou, S.-R., Arteaga, I. A., & White, B. A. B. (2011). School-Based Early Childhood Education and Age-28 Well-Being: Effects by Timing, Dosage, and Subgroups. *Science, 333*(6040), 360-364. doi: 10.1126/science.1203618
- Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term Effects of an Early Childhood Intervention on Educational Achievement and Juvenile Arrest: A 15-Year Follow-up of Low-Income Children in Public Schools. *JAMA : the journal of the American Medical Association, 285*(18), 2339-2346. doi: 10.1001/jama.285.18.2339
pmid:
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.
- Rouse, C., Brooks-Gunn, J., & McLanahan, S. (2005). School readiness: Closing racial and ethnic gaps: Introducing the issue. *The Future of Children, 15*(1), 5-13.
- Schochet, P., Cook, T. D., Deke, J., Imbens, G. W., Lockwood, J. R., Porter, J., & Smith, J. (2010). Standards for regression discontinuity designs: What Works Clearinghouse, Institute for Education Sciences, Department of Education.
- Schweinhart, L. J. (2005). Lifetime Effects: The High/Scope Perry Preschool Study through Age 40 (Vol. 14). Ypsilanti, MI: High/Scope Educational Research Foundation.
- Schweinhart, L. J., & Weikart, D. P. (1981). Effects of the Perry Preschool Program on Youths Through Age 15. *Journal of Early Intervention, 4*(1), 29-39. doi: 10.1177/105381518100400105
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.
- Skibbe, L. E., Connor, C. M., Morrison, F. J., & Jewkes, A. M. (2011). Schooling effects on preschoolers: self-regulation, early literacy, and language growth. *Early Childhood Research Quarterly, 26*(1), 42-49. doi: <http://dx.doi.org/10.1016/j.ecresq.2010.05.001>
- StataCorp. (2011). Stata statistical software: Release 12. College Station, TX.
- Tarullo, L. B., Aikens, N., Moiduddin, E., & West, J. (2010). A second year in Head Start: Characteristics and outcomes of children who entered the program at age three. Washington, D.C.: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.
- Tarullo, L. B., Xue, Y., & Burchinal, M. R. (2013, April). *Are two years better than one? Examining dosage of Head Start attendees using propensity score matching methodology*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Seattle, WA.
- U.S. Department of Education. (2013). Early childhood education intervention report: The Creative Curriculum for Preschool, Fourth Edition.: Institute of Education Sciences, What Works Clearinghouse.

- Van Der Klaauw, W. (2008). Regression–Discontinuity Analysis: A Survey of Recent Developments in Economics. *LABOUR*, 22(2), 219-245. doi: 10.1111/j.1467-9914.2008.00419.x
- Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & Network, N. E. C. C. R. (2010). Do Effects of Early Child Care Extend to Age 15 Years? Results From the NICHD Study of Early Child Care and Youth Development. *Child Development*, 81(3), 737-756. doi: 10.1111/j.1467-8624.2010.01431.x
- Weiland, C., & Yoshikawa, H. (2013). Impacts of a Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills. *Child Development*, 84(6), 2112-2130. doi: 10.1111/cdev.12099
- Wen, X., Leow, C., Hahs-Vaughn, D. L., Korfmacher, J., & Marcus, S. M. (2012). Are two years better than one year? A propensity score analysis of the impact of Head Start program duration on children's school performance in kindergarten. *Early Childhood Research Quarterly*, 27(4), 684-694. doi: <http://dx.doi.org/10.1016/j.ecresq.2011.07.006>
- Winsler, A., Caverly, S. L., Willson-Quayle, A., Carlton, M. P., Howell, C., & Long, G. N. (2002). The social and behavioral ecology of mixed-age and same-age preschool classrooms: A natural experiment. *Journal of Applied Developmental Psychology*, 23(3), 305-330. doi: [http://dx.doi.org/10.1016/S0193-3973\(02\)00111-9](http://dx.doi.org/10.1016/S0193-3973(02)00111-9)
- Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, 27(1), 122-154.
- Woodcock, R. W., & Johnson, M. B. (1989). *Tests of achievement, standard battery*. Chicago, IL: Riverside Publishing.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson-III Tests of Achievement*. Itasca, IL.
- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W., . . . Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. New York, NY: Foundation for Child Development, Society for Research in Child Development.
- Zhai, F., Brooks-Gunn, J., & Waldfogel, J. (2011). Head Start and urban children's school readiness: A birth cohort study in 18 cities. *Developmental Psychology*, 47(1), 134-152. doi: <http://dx.doi.org/10.1037/a0020784>
- Zimmer, R. W., & Toma, E. F. (2000). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management*, 19(1), 75-92. doi: 10.1002/(sici)1520-6688(200024)19:1<75::aid-pam5>3.0.co;2-w

Table 1.

Covariate balance between children who attended Age 3 Head Start + Age 4 OK Pre-k and Age 3 Head Start + Age 4 Head Start in observed data and in Propensity Score weighted data

	(1)	(2)	(3)	(4)
	Observed group means		PS weighted group means	
	HS Age 3; OK Pre-k Age 4	HS Age 3; HS Age 4	HS Age 3; OK Pre-k Age 4	HS Age 3; HS Age 4
Covariates				
Reduced-price lunch	0.10	0.03	0.06	0.06
White	0.10	0.09	0.10	0.10
Black	0.64	0.44	0.54	0.52
Hispanic	0.17	0.39	0.27	0.30
Asian/Native/Other	0.08	0.07	0.08	0.07
Female	0.50	0.55	0.52	0.52
Below High School	0.08	0.19	0.12	0.15
High School	0.30	0.28	0.31	0.29
Some college	0.32	0.32	0.31	0.32
College +	0.09	0.05	0.07	0.07
Child had some non-parental care at age 3	0.55	0.46	0.52	0.50
Internet access in home	0.33	0.29	0.30	0.32
Number of books in home (1-5 scale)	1.86	1.93	1.93	1.94
Parent is foreign-born	0.28	0.43	0.36	0.36
English is home language	0.71	0.59	0.65	0.65
Child has health insurance	0.79	0.78	0.78	0.80
Married	0.26	0.36	0.31	0.33
Child tested in both English and Spanish	0.11	0.31	0.20	0.24
Father lives in home	0.35	0.44	0.40	0.41
Outcomes				
Assessment at Kindergarten entry				
WJ Letter-Word raw score – Cohort 1	10.51 (4.06)	7.98 (4.06)	10.35 (4.14)	8.08 (4.01)
WJ Applied Problems raw score – Cohort 1	13.15 (3.97)	12.95 (3.94)	13.03 (3.90)	12.62 (4.08)
WJ Spelling raw score – Cohort 1	9.06 (2.90)	8.53 (2.41)	9.05 (2.96)	8.46 (2.40)
Assessment at Age 4 program entry				
WJ Letter-Word raw score – Cohort 2	4.55 (3.14)	4.81 (3.14)	4.53 (3.12)	4.82 (4.03)
WJ Applied Problems raw score – Cohort 2	8.39 (4.76)	8.00 (4.66)	8.42 (4.55)	7.86 (4.63)
WJ Spelling raw score – Cohort 2	4.25 (2.14)	5.04 (3.03)	4.54 (2.55)	4.86 (3.10)
<i>Observations</i>	211	329	211	329

Notes: HS-Head Start. Sample restricted to children who are free and reduced-price lunch eligible. Cohort 1 refers to the group of children who participated in OK pre-k or Head Start during the 2005-06 school year and are entering kindergarten at the time of the assessment, the start of the 2006-07 school year. Cohort 2 refers to the group of children who are entering OK pre-k or Head Start in the 2006-07 school year.

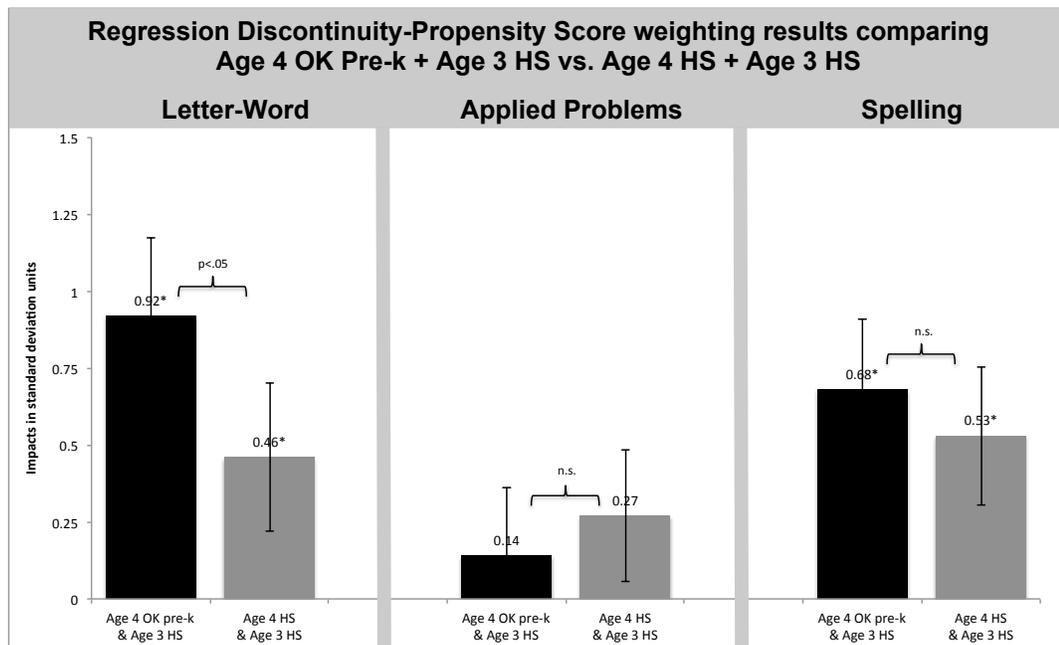
Table 2.

Propensity Score weighted Regression Discontinuity results for the effects of
Age 3 Head Start + Age 4 OK Pre-k vs. Age 3 Head Start + Age 4 Head Start

	Letter-Word	Applied Problems	Spelling
	<i>b/se</i>	<i>b/se</i>	<i>b/se</i>
Age 4 OK Pre-k & Age 3 HS effect (cutoff)	3.77*** (1.03)	0.69 (1.10)	2.17*** (0.73)
<i>Effect size</i>	0.92	0.14	0.68
Age 4 HS & Age 3 HS effect (Pathway *cutoff + cutoff)	1.88* (0.98)	1.36 (1.06)	1.72** (0.72)
<i>Effect size</i>	0.46	0.27	0.53
Age 4 HS & Age 3 HS differential effect (Pathway *cutoff)	-1.89** (0.88)	0.66 (1.09)	-0.45 (0.69)
<i>Effect size and direction of difference</i>	-0.46	+0.13	-0.14
<i>p-value of difference</i>	0.02	0.54	0.51
Age 4 HS & Age 3 HS	0.23 (0.53)	-0.40 (0.76)	0.35 (0.45)
Age as distance from treatment cutoff	0.0048* (0.0029)	0.010*** (0.0029)	0.0064*** (0.0020)
Age squared	0.0000059 (0.0000097)	0.0000080 (0.000010)	0.0000063 (0.0000069)
Female	0.68* (0.39)	0.45 (0.44)	0.82*** (0.29)
Child had some non-parental care at age 3	0.0014 (0.57)	0.56 (0.73)	0.40 (0.39)
Reduced-price lunch	-0.47 (0.77)	-1.86* (0.93)	-0.87 (0.72)
Below High School	-0.67 (0.52)	0.63 (0.60)	-0.077 (0.39)
Some college	0.71 (0.57)	1.42** (0.58)	0.90** (0.45)
College +	1.38 (0.88)	1.99** (0.98)	0.66 (0.57)
Black	1.28** (0.61)	1.02 (0.81)	0.92* (0.47)
Hispanic	0.41 (0.66)	0.74 (0.82)	1.94*** (0.58)
Asian/Native/Other	0.53 (0.78)	2.10** (0.97)	0.44 (0.59)
Missing parent education	-0.22 (0.68)	-0.46 (0.73)	0.17 (0.53)
Missing non-parental care	1.19 (0.62)	1.45** (0.66)	0.60 (0.42)
Constant	3.60*** (0.82)	7.70*** (1.10)	3.45*** (0.69)
Observations	407	404	391

***significant at .01 level, ** significant at .05 level, * significant at .10 level. Reference group for effect of exposure is age 4 OK Pre-k + age 3 Head Start . Observations that fall within the 270 day bandwidth from the treatment cutoff are included (Age-birthdate cutoff <= 270 in absolute value). Outcome variable is a raw score. All models use clustered SEs by teacher.

Figure 1:



Caption: Bars represent preschool exposure effect sizes for each outcome. Brackets indicate the significance of the difference in effect sizes between the two preschool exposures.

Appendix 1: Details of study methodology

A1.1: Propensity score methodology

We predicted treatment status using a logit model using the following covariates: reduced price lunch eligibility, race, sex, parent education level, child's exposure to any non-parental care at age 3, number of books in the home (1-5 scale), and indicators for internet access in the home, parent foreign-born status, English is child's home language, whether the child took both English and Spanish assessments, marital status, father living with child, child health insurance coverage, and missing data. Prior research shows that race and parent education are primary predictors of Head Start enrollment, whereby children who are black and whose mothers have at least a high school degree are more likely than others to be enrolled for two years (Arteaga, et al., 2013; Hofferth, 1994; Lee, 2011). We also included a variety of interaction terms between covariates to adequately model selection processes. We omitted age from the propensity score model because it is central to the RD identification of within-pathway treatment effects, and we did not want to incorporate this variation to predict the selection into pathways. The choice of pathway (Head Start or OK pre-k at age 4) serving as the outcome in the propensity score model was arbitrary and is inconsequential for the results.

After calculating the propensity scores for each age 3 Head Start graduate, we assessed whether there was common support across the age 4 pre-k and Head Start groups using the histograms shown in Appendix 2.2. This indicated that there was adequate overlap in propensity scores, meaning that individuals in both treatment states were comparable with respect to their propensity for treatment (i.e. exchangeable), allowing us to use PS methods in the outcome analysis.

Next, we used the propensity scores to create the Inverse probability of treatment weights (IPTW). Weights are the inverse of the predicted probability of receiving the exposure a person actually received. Our logit model predicted whether children attended Head Start at age 4, so the propensity score represents this probability. Therefore the weights for children who attended Head Start at age 4 are calculated as 1 divided by the propensity score ($1/PS$), and the weights for children who attended OK pre-k at age 4 are 1 over 1 minus the propensity score ($1/1-PS$).

We assessed balance in covariate means across exposure groups by regressing each covariate on the exposure using the propensity score weights. If the relationships between the exposure and the covariates are not significant, the sample is adequately balanced across the treated and untreated groups. The results of this balance checking are reported in columns 3 and 4 of Table 1, which shows the IPT-weighted group means for both pathways compared to the unweighted group means. The two groups become very similar with respect to observed covariates after weighting, and there are no remaining significant relationships between Head Start or pre-k and the covariates. We also used our propensity scores to match observations between groups to check the robustness of our balance and outcome models using nearest-neighbor matching with replacement and a 0.01 caliper. The results were similar but the IPTW results were more efficient and allowed us to sustain our sample size (even though some observations may have a small weight).

We were also concerned that there may be differential selection into age 4 treatments based on children's outcomes at the start of the preschool year (i.e. after their age 3 Head Start experience). Therefore, we wanted to assess whether there were differences between our two pathway groups at the start of their age 4 preschool exposure. We tested for differences in means in the age 4 program entrants (excluding kindergarten entrants) using t-tests, and for differences in the distributions of the three outcomes (Letter-word, Applied Problems, Spelling) using a Kolmogorov-Smirnov test and a Mann-Whitney rank-sum test. Results and histograms are shown in Appendix 2.3, which indicate that there were no significant differences in the means or distributions of the outcomes between the Head Start and OK pre-k children at the start of their age 4 program.

A1.2: Regression discontinuity methodology

A graphical analysis of the discontinuity in the dependent variables near the treatment cutoff date is shown for each outcome in Appendix 2.4. We used these bar graphs, as well as the histogram of children's age relative to the treatment cutoff date in Appendix 2.5 to check for clustering of children near the cutoff. In combining the PS weighting with an RD model, we wanted to check that the PS weights were evenly distributed on both sides of the age cutoff to ensure that we did not compromise the RD identification strategy. In Appendix 2.6, we present a scatterplot of PS weights by age for both pathways, which shows an even distribution of PS weights on both sides of the cutoff in each treatment condition.

A key assumption of the RD model is that the individuals who are closest to the cutoff—just above and below—are comparable (i.e. have similar potential outcomes) because the value of their assignment variable is very similar (Van Der Klaauw, 2008). All other characteristics of these individuals can be considered independent of treatment status. This is an important assumption because the treatment effect identified through the discontinuity at the cutoff compares the average outcomes for those with values just above and below the cutoff. Therefore the RD estimate must be interpreted conditionally; rather than being an average treatment effect, it is a *local* average treatment effect, applying primarily to cases that fall within a close range around the cutoff.

For this reason, it is also important to check for an appropriate 'bandwidth', which involves an analysis of restricted samples of observations clustered around the cutoff within a range of the assignment variable (e.g. +/- 90 days, 180 days) (Schochet et al., 2010; Van Der Klaauw, 2008). We display our optimal bandwidth of 270 days in Table 2, but we also tested two other bandwidth restrictions to gauge the robustness of our RD estimates. These results are available in the Appendix 2.7. The pattern of results are similar to what we present in main text, where a wider bandwidth produces larger coefficients, and a narrower bandwidth produces smaller coefficients, some of which lose significance due to decreased power and efficiency.

Appendix 2: Supplemental tables and figures

A2.1: Descriptive analysis of peers

A. Outcomes for Age 3 Head Start graduates at the start of their age 4 preschool program, by age 4 preschool program

	OK Pre-k at age 4	Head Start at age 4	Standardized Mean Difference
Letter-Word	4.55	4.81	-.07
Applied Problems	8.38	7.99	.08
Spelling	4.25	5.04	-.27*
Parent Ed – College +	.05	.03	
Parent Ed – Some College	.27	.26	
Parent Ed – High School	.21	.24	X ² =0.01
Parent Ed – Below High Sch.	.06	.16	

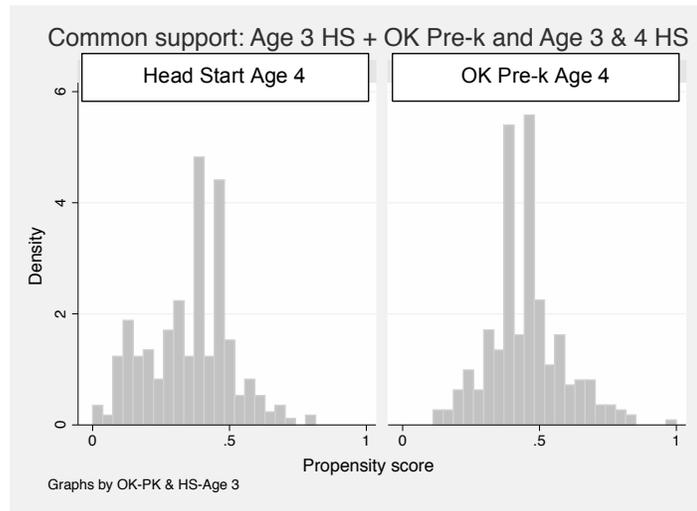
B. Outcomes and covariates for children who attended age 3 Head Start compared to their cohort peers who did not attend age 3 Head Start, measured at the start of the age 4 preschool program

	OK Pre-k at age 4			Head Start at age 4		
	Age 3 HS Graduates	No Age 3 HS	Standardized Mean Difference	Age 3 HS Graduates	No Age 3 HS	Standardized Mean Difference
Letter-Word	4.47	4.49	.01	4.89	3.52	.38*
Applied Problems	8.14	8.70	-.12	8.02	6.61	.31*
Spelling	4.03	4.78	-.27*	5.12	4.38	.27*
Parent Ed – College +	.07	.13		.03	.02	
Parent Ed – Some College	.27	.25		.26	.13	
Parent Ed – High School	.21	.18	X ² =0.07	.24	.20	X ² =0.00
Parent Ed – Below High Sch.	.06	.12		.16	.19	
Free or reduced-price lunch eligible	.92	.74	.18* [^]	.98	.94	.03* [^]

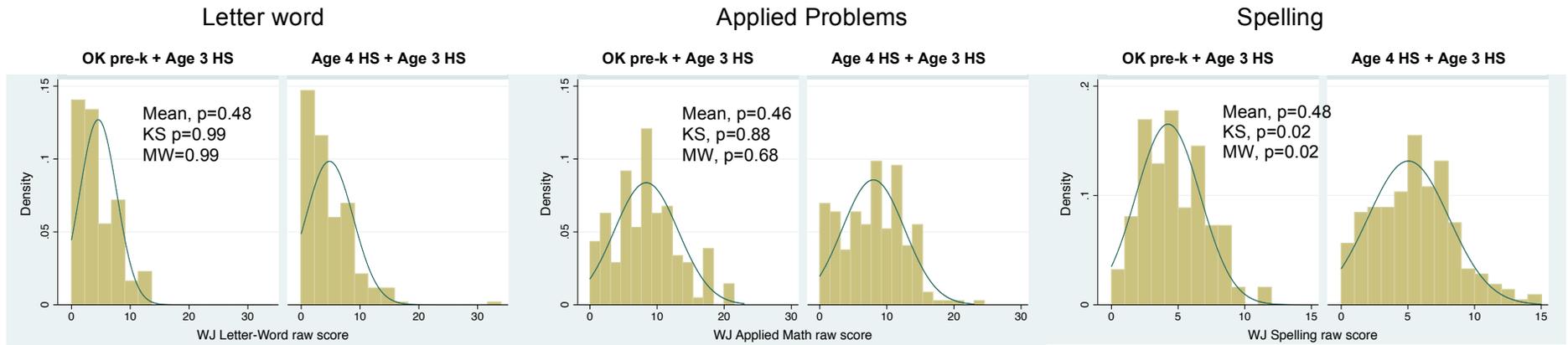
Notes: HS- Head Start. [^]Value is not standardized mean difference for ordinal and dichotomous variables; difference in proportion determined by X² or z-test. * indicates significance at the .05 level. All children are free or reduced-price lunch eligible in panel A.

We present the average assessment scores for all age 3 Head Start graduates measured at the beginning of their age 4 programs in 2006-07 (using the younger cohort in the sample) as a proxy for a post-age 3 Head Start assessment. Comparing age 3 Head Start graduates between age-4 programs shows that the two groups have insignificantly different letter-word and applied problems scores ($p=0.45, 0.50$), although second-year Head Start entrants have higher spelling scores (Standardized mean difference (SMD)=0.27, $p=0.00$). A comparison of age 3 Head Start graduates with their peers who *did not* attend age 3 Head Start show more consistent differences; second year Head Start students appear to have an advantage over their peers with respect to pre-academic skills, while those attending OK pre-k are fairly similar to their peers. In OK pre-k, children who attended age 3 Head Start have similar applied problems and spelling scores relative to their pre-k peers ($p=0.35, 0.50$), but lower letter-word scores (SMD=.27, $p=0.01$). Within age 4 Head Start, age 3 Head Start graduates have significantly higher letter-word, applied problems, and spelling scores compared with their peers who did not attend Head Start at age 3 (SMD=.38, .31, .27, respectively). OK pre-k also has a lower percentage of children qualifying for free and reduced price lunch, indicating that higher-income children join the age 3 Head Start graduates in OK pre-k. At least descriptively, comparing the ability and characteristics of the *peers* of age 3 Head Start graduates in their age-4 programs indicate potentially different peer effects for both the OK pre-k entrants and age 4 Head Start entrants.

A2.2: Histogram of propensity scores to assess common support between age 4 treatment states



A2.3: Distributional plots of raw scores for children entering age 4 programs (i.e., younger cohort only)



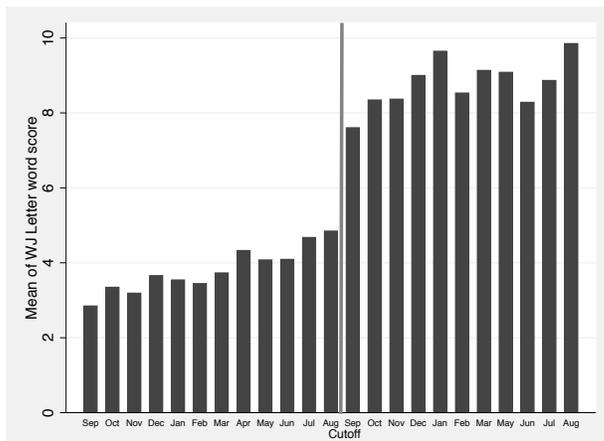
Legend

Mean (weighted)= Z-test from logit of outcome on treatment indicator using PS weights
KS= Kolmogorov-Smirnov test for equality of distribution functions, without PS weights*
MW= Mann-Whitney rank-sum test, without PS weights*

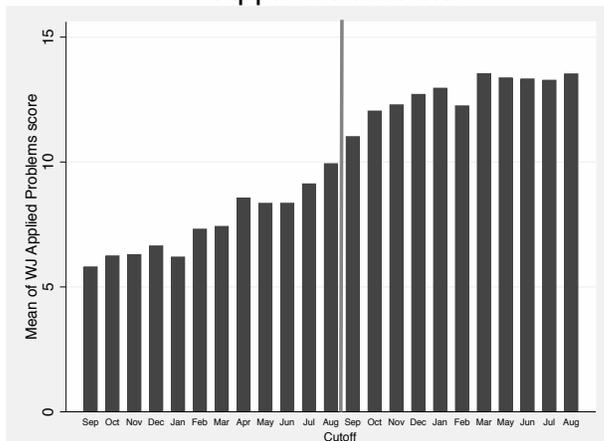
* Weights cannot be applied with this statistical test

A2.4: Graphs of outcome variables by child month of birth on both sides of birthdate cutoff

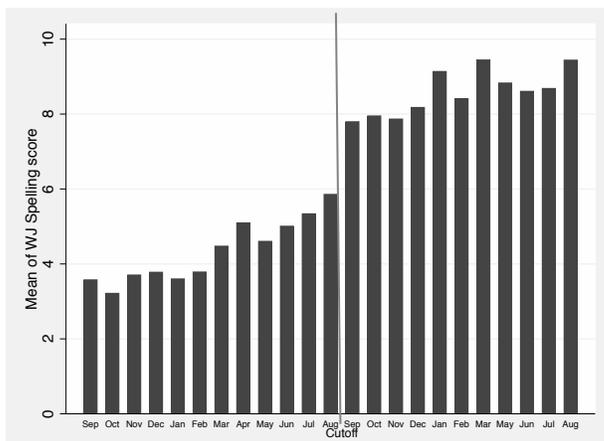
Letter Word



Applied Problems

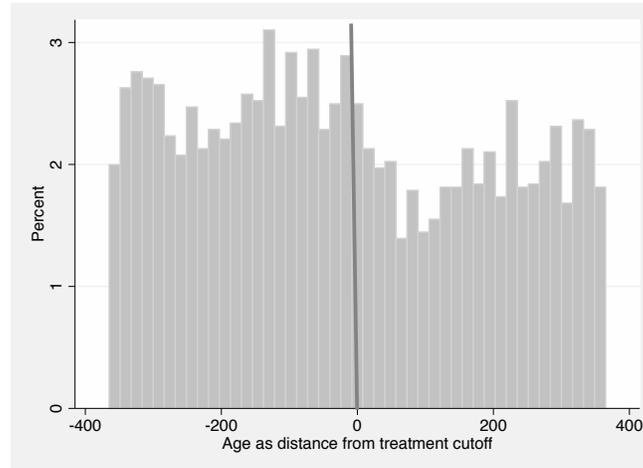


Spelling



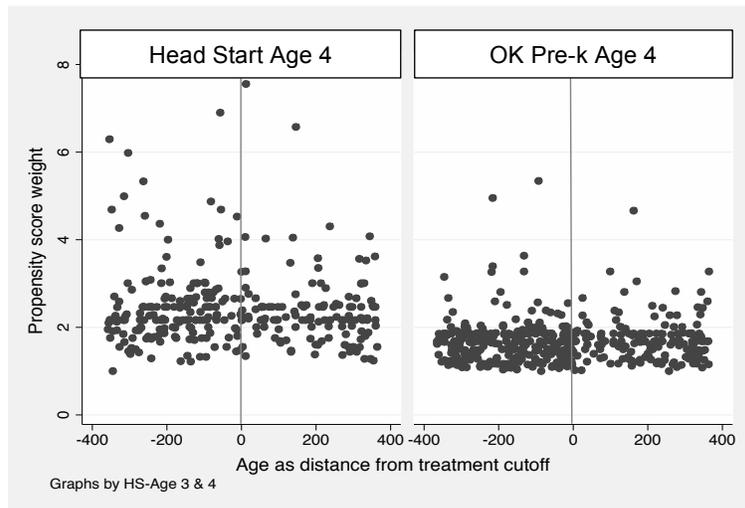
Caption: Graphs illustrate presence or lack of potential discontinuity in raw outcome scores at the birthdate cutoff for treatment at age 4, and no evidence of sorting near the cutoff.

A2.5: Histogram of age of OK Pre-k Study sample relative to the treatment cutoff date



Caption: Y-axis indicates the percent of children within a birthdate range (i.e., bar) in the entire study sample. This figure shows that the distributions of children's ages are similar on both sides of the cutoff, and that there are more children in the sample who did not make the birthdate cutoff relative to the number of children who made the cutoff (n=2143, 1661, respectively).

A2.6: Scatterplot of propensity score weights by age and preschool pathway for analysis sample (i.e. all age 3 Head Start participants)



Caption: Y-axis indicates the sampling weight applied to an observation in the propensity score analysis. The similar range of weights across both graphs indicates common support between the two age four treatment conditions (OK pre-k and Head Start). The similar ranges of weights on both sides of the cutoff in each treatment condition indicates that the propensity score weighting does not compromise the identification strategy of the regression discontinuity design.

A2.7: Propensity Score weighted Regression Discontinuity results using alternative bandwidth restrictions for Letter Word, Applied Problems, and Spelling scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	LW - all obs.	LW with BW 270	LW with BW 180	AP - all obs.	AP with BW 270	AP with BW 180	Spelling - all obs.	Spelling with BW 270	Spelling with BW 180
Born Before 0506 Pre-k Cut-off	4.41*** (0.96)	3.77*** (1.03)	2.45* (1.10)	1.35 (1.07)	0.70 (1.10)	-0.36 (1.30)	2.55*** (0.68)	2.18** (0.73)	1.71* (0.86)
Age + & Age 4 HS differential effect (Age 3 + 4 HS * cutoff)	-2.62*** (0.75)	-1.89* (0.85)	-1.70 (1.04)	-0.25 (0.97)	0.66 (1.09)	0.89 (1.28)	-1.25* (0.54)	-0.45 (0.69)	-0.26 (0.78)
HS-Age 3 + 4	0.34 (0.48)	0.23 (0.53)	0.68 (0.67)	-0.51 (0.72)	-0.40 (0.76)	-0.10 (0.90)	0.38 (0.36)	0.35 (0.45)	0.98 (0.53)
Age as distance from treatment cutoff	0.0035 (0.0018)	0.0048 (0.0029)	0.013** (0.0047)	0.0093*** (0.0017)	0.010*** (0.0029)	0.017*** (0.0051)	0.0058*** (0.0012)	0.0064** (0.0020)	0.0088* (0.0037)
Age squared	0.0000016 (0.0000048)	0.0000059 (0.0000097)	0.000079** (0.000026)	-0.0000020 (0.0000047)	0.0000080 (0.000010)	0.000036 (0.000028)	0.000000070 (0.0000031)	0.00000063 (0.0000069)	0.000047** (0.000017)
Female	0.66* (0.33)	0.68 (0.39)	0.60 (0.48)	0.55 (0.36)	0.45 (0.44)	0.27 (0.52)	0.91*** (0.25)	0.82** (0.29)	1.00** (0.36)
Child had some non-parental care at age 3	0.41 (0.47)	0.0014 (0.57)	0.14 (0.70)	0.57 (0.58)	0.56 (0.73)	0.66 (1.02)	0.36 (0.32)	0.40 (0.39)	0.75 (0.50)
Reduced-price lunch	0.19 (0.63)	-0.47 (0.77)	-0.49 (0.87)	-0.79 (0.74)	-1.86* (0.93)	-2.14 (1.15)	-0.41 (0.59)	-0.87 (0.72)	-0.37 (0.79)
Below High School	-0.63 (0.47)	-0.67 (0.52)	0.076 (0.67)	0.48 (0.46)	0.63 (0.60)	1.16 (0.78)	-0.12 (0.33)	-0.077 (0.39)	-0.014 (0.54)
Some college	0.81 (0.43)	0.71 (0.57)	0.81 (0.61)	1.27** (0.47)	1.42* (0.58)	1.56* (0.75)	0.80* (0.34)	0.90* (0.45)	0.60 (0.51)
College +	1.03 (0.71)	1.38 (0.88)	0.89 (1.04)	0.82 (0.87)	1.99* (0.98)	1.29 (1.24)	0.57 (0.53)	0.66 (0.57)	0.34 (0.69)
Black	0.54 (0.57)	1.28* (0.61)	1.20 (0.69)	0.52 (0.71)	1.02 (0.81)	0.60 (0.96)	0.65 (0.39)	0.92 (0.47)	1.16 (0.62)
Hispanic	-0.018 (0.60)	0.41 (0.66)	0.22 (0.76)	0.48 (0.71)	0.74 (0.82)	0.59 (0.93)	1.74*** (0.48)	1.94** (0.58)	2.56*** (0.70)
Asian/Native/Other	-0.070 (0.67)	0.53 (0.78)	-0.068 (0.94)	1.36 (0.87)	2.10* (0.97)	1.63 (1.26)	0.95 (0.49)	0.44 (0.59)	0.71 (0.77)
Missing parent education	0.14 (0.60)	-0.22 (0.68)	0.62 (0.92)	-0.41 (0.55)	-0.46 (0.73)	0.15 (1.05)	-0.0019 (0.41)	0.17 (0.53)	0.12 (0.72)
Missing non-parental care	0.70 (0.45)	1.19 (0.62)	1.16 (0.90)	1.11 (0.56)	1.45* (0.66)	1.30 (1.09)	0.43 (0.34)	0.60 (0.42)	1.03* (0.52)
Constant	3.82*** (0.65)	3.60*** (0.82)	3.27** (0.99)	8.38*** (0.92)	7.70*** (1.10)	7.98*** (1.41)	3.61*** (0.57)	3.45*** (0.69)	2.27* (0.89)
Observations	570	407	263	565	404	262	544	391	251

Standard errors in parentheses. Reference group for effect of exposure is OK pre-k & age 3 Head Start. Outcome variable is a raw score. * p<0.05; ** p<0.01; *** p<0.001. HS-Head Start; LW- WJ Letter-Word score; AP- WJ Applied Problems score. BW 270 results are presented in main tables and results section.

**A2.8: Propensity Score Weighted Regression Discontinuity results for the effects of
Age 3 Head Start + Age 4 Head Start vs.
Age 4 Head Start with no Age 3 Head Start**

	Letter-Word	Applied Problems	Spelling
	<i>b/se</i>	<i>b/se</i>	<i>b/se</i>
Age 4 HS; no Age 3 HS effect (cutoff)	0.74 (1.03)	1.94** (0.79)	0.42 (0.62)
<i>Effect size</i>	0.18	0.39	0.14
Age 3 HS + Age 4 HS effect (Pathway *cutoff + cutoff)	1.26 (1.17)	2.28** (0.97)	0.90 (0.73)
<i>Effect size</i>	0.31	0.46	0.29
Age 3 HS + Age 4 HS differential effect (Pathway *cutoff)	0.52 (0.64)	0.33 (0.85)	0.48 (0.51)
<i>Effect size and direction of difference</i>	+0.13	+0.07	+0.15
<i>p-value of difference</i>	0.41	0.69	0.36
Age 4 HS + Age 3 HS	0.59 (0.48)	0.88 (0.54)	0.49 (0.43)
Age as distance from treatment cutoff	0.0056* (0.0029)	0.0081** (0.0025)	0.0077*** (0.0019)
Age squared	-0.000013 (0.0000085)	-0.000021** (0.0000074)	-0.000016** (0.0000061)
Female	0.43 (0.38)	0.29 (0.39)	0.56 (0.30)
Child had some non-parental care at age 3	-0.42 (0.37)	0.69 (0.51)	-0.057 (0.35)
Reduced-price lunch	-0.27 (0.70)	-0.97 (0.67)	-0.59 (0.58)
Below High School	-0.78* (0.39)	-0.42 (0.49)	-0.11 (0.35)
Some college	1.17** (0.48)	-0.50 (0.42)	0.24 (0.40)
College +	2.36* (1.31)	-0.28 (1.13)	1.87 (1.28)
Black	0.13 (0.63)	-0.72 (0.80)	0.60 (0.53)
Hispanic	-0.91 (0.76)	-1.59* (0.83)	0.76 (0.66)
Asian/Native/Other	-0.70 (0.86)	-0.77 (0.95)	-0.13 (0.60)
Missing parent education	0.076 (0.87)	-1.44 (0.73)	-0.25 (0.41)
Missing non-parental care	0.38 (0.51)	0.97 (0.77)	-0.16 (0.42)
Constant	5.35*** (0.81)	9.91*** (0.80)	5.60*** (0.61)
Observations	571	567	558

***significant at .01 level, ** significant at .05 level, * significant at .10 level. Reference group for effect of exposure is age 4 Head Start with no age 3 Head Start. Observations that fall within the 270 bandwidth from the treatment cutoff are included (Age-birthdate cutoff <= 270 in absolute value). Outcome variable is a raw score. All models use clustered SEs by teacher.